Article

Operational Automatic Remote Sensing Image Understanding Systems: Beyond Geographic Object-Based and Object-Oriented Image Analysis (GEOBIA/GEOOIA). Part 1: Introduction

Andrea Baraldi * and Luigi Boschetti

Department of Geography, University of Maryland, 4321 Hartwick Rd, Suite 209, College Park, MD 20740, USA; E-Mail: luigi@hermes.geog.umd.edu

* Author to whom correspondence should be addressed;
E-Mail: andrea.baraldi@hermes.geog.umd.edu; Tel.: +1-301-314-1467; Fax: +1-301-405-6806.

Abstract: According to existing literature and despite their commercial success, state-of-the-art two-stage non-iterative geographic object-based image analysis (GEOBIA) systems and three-stage iterative geographic object-oriented image analysis (GEOOIA) systems, where GEOOIA $\supset$ GEOBIA, remain affected by a lack of productivity, general consensus and research. To outperform the degree of automation, accuracy, efficiency, robustness, scalability and timeliness of existing GEOBIA/GEOOIA systems in compliance with the Quality Assurance Framework for Earth Observation (QA4EO) guidelines, this methodological work is split into two parts. The first part of this work provides a multi-disciplinary Strengths, Weaknesses, Opportunities and Threats (SWOT) analysis of the GEOBIA/GEOOIA approaches that augments similar analyses proposed in recent years. In line with constraints stemming from human vision, this SWOT analysis promotes a shift of learning paradigm in the pre-attentive vision first stage of a remote sensing (RS) image understanding system (RS-IUS), from sub-symbolic statistical model-based (inductive) image segmentation to symbolic physical model-based (deductive) image preliminary classification. Hence, a symbolic deductive pre-attentive vision first stage accomplishes image sub-symbolic segmentation and image symbolic pre-classification simultaneously. In the second part of this work a novel hybrid (combined deductive and inductive) RS-IUS architecture featuring a symbolic deductive pre-attentive vision first stage is proposed and discussed in terms of: (a) computational theory (system design); (b) information/knowledge representation; (c) algorithm design; and

(d) implementation. As proof-of-concept of symbolic physical model-based pre-attentive vision first stage, the spectral knowledge-based, operational, near real-time Satellite Image Automatic Mapper™ (SIAM™) is selected from existing literature. To the best of these authors' knowledge, this is the first time a symbolic syntactic inference system, like SIAM™, is made available to the RS community for operational use in a RS-IUS pre-attentive vision first stage, to accomplish multi-scale image segmentation and multi-granularity image pre-classification simultaneously, automatically and in near real-time.

Acronyms and Abbreviations

AI:	Artificial Intelligence
ATCOR:	Atmospheric/Topographic Correction
Cal/Val:	Calibration and Validation
CEOS:	Committee on Earth Observation Satellites
CS:	Computer Science
CV:	Computer Vision
DN:	Digital Number
EO:	Earth Observation
ESA:	European Space Agency
GEO:	Group on Earth Observations
GEOBIA:	Geographic Object-Based Image Analysis
GEOOIA:	Geographic Object-Observation Image Analysis
GEOSS:	Global EO System of Systems
GIS:	Geographic Information System
GIScience:	Geographic Information Science
GMES:	Global Monitoring for the Environment and Security
IT:	Information Technology
LAI:	Leaf Area Index
LC:	Land Cover
LCC:	Land Cover Change
LCLUC:	Land Cover and Land Use Change program
MAI:	Machine Intelligence

MAL: Machine Learning
MAT: Machine Teaching
MS: Multi-Spectral
NASA: National Aeronautics and Space Administration
OO: Object-Oriented
OQI: Quality Index of Operativeness
QA: Quality Assurance
QA4EO: Quality Accuracy Framework for Earth Observation
QI: Quality Index
RS: Remote Sensing
RS-IUS: Remote Sensing Image Understanding System
SR: Spatial Resolution
SIAM™:   Satellite Image Automatic Mapper™
SURF: Surface Reflectance
SVM: Support Vector Machine
SWOT: Strengths, Weaknesses, Opportunities and Threats analysis
TM: Trademark
TOA: Top-Of-Atmosphere
TOARF: TOA Reflectance
TOC: Topographic Correction
USGS: US Geological Survey
VHR: Very High Resolution
WELD: Web-Enabled Landsat Data set project
WGCV: Working Group on Calibration and Validation

1. Introduction

This methodological work aims at one traditional, albeit visionary goal of the remote sensing (RS) community: the development of operational (good-to-go, press-and-go, turnkey) satellite-based information/knowledge processing systems capable of automating the quantitative analysis of large-scale spaceborne multi-source multi-resolution image databases ([1]; p. 451), in compliance with the guidelines of the Quality Assurance Framework for Earth Observation (QA4EO) delivered by the Working Group on Calibration and Validation (WGCV) of the Committee on Earth Observation Satellites (CEOS), the space arm of the Group on Earth Observations (GEO) [2].

According to the terminology adopted in this work, satellite-based information/knowledge processing systems include satellite-based measurement systems as a special case. To further investigate the concepts of (numerical, sensory) '*data*' (observables, true facts), (sub-symbolic, quantitative, unequivocal) '*information-as-thing*' according to the Shannon theory of communication [3], (symbolic, qualitative, equivocal) '*information-as-(an intepretation)process*', *i.e.*, information as interpreted data, and '*knowledge*', refer to [4,5].

For publication purposes this theoretical contribution is split into two parts. The present first part identifies possible causes of the lack of productivity affecting existing academic and commercial RS

image understanding systems (RS-IUSs) outpaced by the ever-increasing rate of collection of spaceborne and airborne sensory data. To reach its objective, this contribution adopts a holistic convergence-of-evidence approach to provide an inter-disciplinary analysis of biological vision, computer vision (CV), artificial intelligence (AI), machine learning (MAL) and RS-IUS design and implementation, with special emphasis on state-of-the-art two-stage non-iterative geographic (2-D) object-based image analysis (GEOBIA) systems [6–11] and three-stage iterative geographic (2-D) object-oriented image analysis (GEOOIA) systems [6], where GEOBIA is a special case of GEOOIA, *i.e.*, GEOOIA ⊃ GEOBIA.

In compliance with the QA4EO guidelines together with constraints stemming from human vision, the second part of this work proposes an original *hybrid* (combined deductive and inductive) RS-IUS design and implementation as a viable alternative to the current state-of-the-art GEOBIA/GEOOIA systems [12]. Quality indexes (QIs) of operativeness (OQIs) of the new class of hybrid RS-IUSs are required to score high in real-world applications, including RS image classification at large (e.g., continental, global) spatial scale and fine semantic granularity. The degree of novelty of the proposed hybrid RS-IUS is investigated at the four levels of understanding of an information processing system [13,14], namely: (A) computational theory (system architecture); (B) information/knowledge representation; (C) algorithm design; and (D) implementation. It is important to mention that, according to existing literature, "the linchpin of success (of an information processing system) is addressing the (computational) theory rather than algorithms or implementation" ([14]; p. 376) (which is in line with holism—the whole is greater than the sum of its parts) [13].

With regard to the terminology adopted in this work in compliance with philosophical hermeneutics [4,5], the following considerations hold (refer to Section 3 below).

- Synonyms of (sub-symbolic or symbolic) deductive inference are: (sub-symbolic or symbolic) deductive learning, top-down inference, coarse-to-fine inference, driven-by-knowledge inference, learning-by-rules, physical model, prior knowledge-based decision system, rule-based system, expert system, syntactic inference, syntactic pattern recognition.
- Synonyms of (sub-symbolic or symbolic) inductive inference are: (sub-symbolic or symbolic) inductive learning, bottom-up inference, fine-to-coarse inference, driven-without-knowledge (knowledge-free) inference, learning-from-examples, statistical model.
- Terms sub-symbolic, sensory, numerical, non-semantic, quantitative, objective, unequivocal are synonyms.
- Terms symbolic, semantic, cognitive, categorical, ordinal, nominal, qualitative, subjective, equivocal are synonyms.

The main thesis of this work is that, to outperform OQIs featured by existing state-of-the-art GEOBIA/GEOOIA systems, an alternative *hybrid* RS-IUS design is required to accomplish a shift of learning paradigm in the pre-attentive vision first stage, from sub-symbolic statistical model-based image segmentation to symbolic physical model-based image preliminary classification (pre-classification). Hence, a symbolic deductive pre-attentive vision first stage accomplishes image sub-symbolic segmentation and image symbolic pre-classification simultaneously. In fact, the generation of a segmentation map from a binary mask or multi-level image (e.g., a thematic map) is a well-posed segmentation problem (*i.e.*, the problem solution exists and is unique), typically solved by

a computationally efficient two-pass connected-component image labeling algorithm [14]. In practice, a unique (sub-symbolic) segmentation map can be generated from a (symbolic) thematic map, but the contrary does not hold, *i.e.*, varying thematic maps can generate the same segmentation map [15].

As proof-of-concept of symbolic deductive pre-attentive vision first stage, the spectral knowledge-based, operational, near real-time, multi-sensor, multi-resolution, application-independent (general-purpose) Satellite Image Automatic Mapper™ (SIAM™) is selected from existing literature [16–24]. SIAM™ is termed '*fully automatic*' because it requires neither user-defined parameters nor training data samples to run. As output SIAM™ automatically generates RS image segmentation maps at multiple spatial scales together with RS image pre-classification maps at multiple semantic granularities.

In the RS literature expert systems have been (almost) exclusively proposed in the attentive vision second-stage classification [25–31]. To the best of these authors' knowledge, this is the first time a symbolic syntactic inference system like SIAM™ is made available to the RS community for operational use in a RS-IUS pre-attentive vision first stage to accomplish multi-scale image segmentation and multi-granularity image pre-classification simultaneously, automatically and in near real-time.

The proposed shift of learning paradigm from sub-symbolic inductive to symbolic deductive inference at the pre-attentive vision first stage is in line with three important quotes from authors belonging to different scientific disciplines like MAL, CV and psychophysics.

- Mulier and Cherkassky: "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" ([32]; p. 39).
- Marr: "vision goes symbolic almost immediately, right at the level of zero-crossing (first-stage primal sketch)... without loss of information" ([13]; p. 343).
- Vecera and Farah: "we have demonstrated that image segmentation can be influenced by the familiarity of the shape being segmented", "these results are consistent with the hypothesis that image segmentation is an interactive (hybrid inference) process" "in which top-down knowledge partly guides lower level processing". "If an unambiguous, yet unfamiliar, shape is presented, top-down influences are unable to overcome powerful bottom-up cues. Some degree of ambiguity is required to overcome bottom-up cues in such situations. The main conclusion from these simulation studies is that while bottom-up cues are sometimes sufficient for processing, these cues do not act alone; top-down cues, on the basis of familiarity, also appear to influence perceptual organization" ([33]; p. 1294).

The rest of this paper is organized as follows. Section 2 identifies inadequacies of existing RS-IUSs and opportunities for improvement. The terminology adopted in this work is proposed in Section 3. Section 4 provides a critical analysis of deductive inference at the basis of AI and inductive inference at the basis of the MAL discipline. Section 5 reviews the basic principles of biological and artificial vision. Section 6 provides an introduction to the GEOBIA objectives, principles, architecture and implementation. Section 7 provides a sketch of the three-stage iterative GEOOIA design and implementation, where GEOOIA ⊃ GEOBIA. In Section 8, a Strengths, Weaknesses, Opportunities

and Threats (SWOT) analysis of the GEOBIA/GEOOIA paradigm is proposed to augment similar analyses proposed by Hay and Castilla in recent years [34,35]. Conclusions are reported in Section 9.

2. Problem Recognition and Opportunity Identification

Founded in 2003, the GEO is a voluntary partnership of governments and international organizations whose mandate is to provide a framework for the coordination of efforts and strategies capable of addressing common goals in Earth observation (EO) disciplines. In 2005 GEO launched a "ten-year implementation plan" to establish its visionary goal of a Global Earth Observation System of Systems (GEOSS). The GEOSS key objective is to *deliver operational, comprehensive and timely "knowledge/information products"* (refer to Section 1) generated (rather than extracted [12]) from a variety of satellite, airborne and *in situ* sensory data sources [2]. Interoperability in terms of synergistic use of multi-source multi-resolution data depends upon the successful implementation of two key principles–*Accessibility/Availability* and *Suitability/Reliability*, to allow the provision of and access to *the Right Information*, *in the Right Format*, *at the Right Time*, *to the Right People*, *to Make the Right Decisions*. This is tantamount to saying that the necessary and sufficient condition for the development of satellite-based information/knowledge processing systems to be used in operational mode in local- to global-scale monitoring programs [1] is the successful implementation of the GEOSS key objectives of: (a) *Accessibility/Availability* and (b) *Suitability/Reliability* of RS data and data-derived information/knowledge products.

To pursue the two GEOSS key principles, GEO identified the need to develop a GEO data quality assurance (QA) strategy where calibration and validation (Cal/Val) activities become critical to data QA and thus to data usability. According to the GEO-CEOS QA4EO guidelines (refer to Section 1) [2]:

- An appropriate coordinated program of Cal/Val activities throughout all stages of a spaceborne mission, from sensor building to end-of-life, is considered mandatory to ensure the harmonization and interoperability of multi-source multi-temporal observational data and derived products. By definition, radiometric calibration is the transformation of dimensionless digital numbers (DNs) into a community-agreed physical unit of radiometric measure.
- Sensory data and derived products generated at each step of a satellite-based information processing workflow must have associated with them a set of quantifiable metrological/ statistically-based mutually uncorrelated quality indicators (QIs) featuring a degree of uncertainty in measurement to provide a documented traceability of the propagation of errors through the information processing chain in comparison with established community-agreed reference standards.

In past years the development of operational RS-IUSs was pursued almost exclusively by international organizations, such as the GEO [2], in collaboration with scientific institutions involved in research programs on detection of land cover (LC) and land cover change (LCC) at continental or global scales ([1]; pp. 452–453). In the same years the large majority of the RS community seemed to be focused on LC and LCC applications at local or regional scales, where accuracy rather than automation, as automation can come on the expense of accuracy, was considered of main interest for assessment and comparison purposes.

In recent years the ambitious objective of developing operational RS-IUSs has become increasingly urgent due to multiple drivers. Firstly, 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 in line with the GEO vision [1,2,36–40]. Secondly, 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 [40]. Thirdly, an increasing number of ongoing international research projects aims at developing operational capabilities and services that require harmonization and interoperability of EO data and derived geo-spatial information products generated from a variety of spaceborne imaging sensors at global, regional and local scales [1]. Among these ongoing programs worth mentioning is the Global Monitoring for the Environment and Security (GMES), an initiative led by the European Union (EU) in partnership with the European Space Agency (ESA) [41,42], the National Aeronautics and Space Administration (NASA) Land Cover and Land Use Change (LCLUC) program ([1]; p. 3) and the US Geological Survey (USGS)-NASA Web-Enabled Landsat Data (WELD) project [43], in addition to the aforementioned GEO GEOSS [38,39].

Unfortunately, to date the automatic or semi-automatic transformation of huge amounts of multi-source multi-resolution EO images into information/knowledge can still be considered far more problematic than might be reasonably expected. In practice, the increasing rate of collection of EO data of enhanced spatial, spectral and temporal quality outpaces the ability of existing RS-IUSs to generate information (e.g., LC and LCC maps) from RS data. This means that productivity in terms of quality, quantity and value of RS data-derived products delivered by the RS community can still be considered low. This conjecture is made strong by the many kinds of supporting evidence collected from the literature, which include the following.

➢ Still now the percentage of data downloaded by stakeholders from the ESA EO databases is estimated at about 10% or less [44].

➢ In large portions of the RS literature: (i) The sole mapping accuracy is selected from the possible set of mutually independent OQIs eligible for parameterizing RS-IUSs for assessment and comparison purposes (refer to this section below), (ii) the statistical estimate of the mapping accuracy is not provided with any degree of uncertainty in measurement in compliance with the principles of statistics together with the QA4EO recommendations [2], and (iii) alternative RS data mapping solutions are tested exclusively in toy problems at a small spatial scale (e.g., local scale) or coarse semantic granularity. The practical consequences of these experimental drawbacks are that, firstly, the mapping accuracy of the proposed RS-IUSs remains unknown in statistical terms and, secondly, the robustness of these RS-IUSs to changes in the input data set together with their scalability to real-world RS applications at large (e.g., continental, global) spatial scale and fine semantic granularity remain unknown or appear questionable.

➢ In line with the QA4EO recommendations [2] the RS community regards as an indisputable fact that "the prerequisite for physically based, quantitative analysis of airborne and satellite sensor measurements in the optical domain is their calibration to spectral radiance" ([45]; p. 29). Irrespective of this common knowledge, radiometric calibration is often neglected in the RS literature and surprisingly ignored by scientists, practitioners and institutions in RS common

practice, including large-scale spaceborne image mosaicking and mapping, e.g., see [46,47]. For example, in conflict with the QA4EO guidelines, popular RS-IUS commercial software products, such as those listed in Table 1, do not consider radiometric calibration of RS imagery as a pre-requisite, with the sole exception of the physical model-based Atmospheric/Topographic Correction (ATCOR-2/3/4) commercial software [48,49]. The relaxation of the requirement of radiometric consistency of multi-source multi-temporal multi-spectral (MS) imagery brings, as an inevitable consequence, that these RS-IUS commercial software products, but ATCOR-2/3/4, are based on (inherently ill-posed) statistical rather than physical models, which means they are intrinsically semi-automatic and site-specific (refer to Section 4 below).

Table 1. Existing commercial RS-IUS software products and their degree of match with the international QA4EO guidelines [2].

Commercial RS-IUS Software Products	Sub-Symbolic (Asemantic) Versus Symbolic (Semantic) Information Primitives, Namely, Pixels/Polygons/ Multi-Part Polygons (Strata), as Output of the Pre-Attentive Vision First Stage	Radiometric Calibration (RAD. CAL.) Requirement in Compliance with the International QA4EO Guidelines [2]
PCI Geomatics GeomaticaX	Sub-symbolic pixels	NO RAD. CAL. $\Rightarrow$ statistical model-based: semi-automatic and site-specific
Definiens Developer	Unsupervised data learning sub-symbolic polygons	NO RAD. CAL. $\Rightarrow$ statistical model-based: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 polygons	NO RAD. CAL. $\Rightarrow$ statistical model-based: semi-automatic and site-specific
ERDAS IMAGING Objective	Supervised data learning symbolic polygons	NO RAD. CAL. $\Rightarrow$ semi-automatic and site-specific
Atmospheric/Topographic Correction-2/3/4 (ATCOR-2/3/4) [48,49]	Sub-symbolic pixels or symbolic pixels (where the semantic label is a spectral type provided by the physical model-based spectral decision-tree classifier (SPECL))	Consistent with the QA4EO recommendations: surface reflectance, SURF $\Rightarrow$ inherently ill-posed atmospheric correction first stage $\Rightarrow$ semi-automatic and site-specific.
Novel three-stage stratified hierarchical hybrid RS-IUS employing the Satellite Image Automatic Mapper (SIAM™) as its preliminary classification first stage	Physical model-based symbolic pixels $\in$ symbolic polygons $\in$ symbolic multi-part polygons	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.

➢ In academic and commercial GEOBIA and GEOOIA system implementations, sub-symbolic inductive inference (e.g., image segmentation, unlabeled data clustering) is adopted in the near totality of the pre-attentive vision first stage implementations. The sole exceptions these authors are aware of employ the physical model-based spectral decision-tree classifier (SPECL), implemented as a by-product in the ATCOR-2/3/4 commercial software product [48,49], for biophysical variable estimation from RS optical imagery [50,51]. For more details about SPECL, refer to Section 2 in [12]. In addition, supervised data learning classification is employed in a large majority of the attentive vision second stage implementations of the GEOBIA and GEOOIA systems proposed in literature. In practice, inductive inference is dominant in existing GEOBIA/GEOOIA systems. Hence, if a lack of productivity affects these RS-IUSs independently of their implementation, it may be due to an intrinsic insufficiency of inductive inference to accomplish OQIs superior to reference standards.

➢ There is an "enigmatic" lack of inter-dependence between machine and human vision, namely, between the CV discipline and the studies of biological vision conducted by neurophysiology and psychophysics [52]. For example, in the CV literature it is acknowledged that "many computer vision systems implicitly use some aspects of processing that can be directly related to the perceptual grouping processes of the human visual system. Frequently, however, no claim is made about the pertinence or adequacy of the digital models as embodied by computer algorithms to the proper model of human visual perception. Edge-linking and region-segmentation, which are used as structuring processes for object recognition, are seldom considered to be a part of an overall attempt to structure the image. *This enigmatic situation arises because research and development in computer vision is often considered quite separate from research into the functioning of human vision. A fact that is generally ignored is that biological vision is currently the only measure of the incompleteness of the current stage of computer vision, and illustrates that the problem is still open to solution*" [53].

➢ According to philosophical hermeneutics, the impact upon Computer Science (CS), Information Technology (IT), AI and MAL of existing different quantitative and qualitative concepts of information (respectively, '*information-as-thing*' and '*information-as-(an intepretation)process*'), embedded in more or less explicit information theories, appears largely underestimated (refer to Section 1) [4,5]. It means that fundamental questions—like: When do sub-symbolic data become symbolic information? When does vision go symbolic? *etc.*—appear largely underestimated and, as a consequence, far from being answered.

➢ There is an on-going multi-disciplinary debate about a claimed inadequacy of scientific disciplines such as CV, AI/Machine Intelligence (MAI) and Cybernetics/MAL, whose origins date back to the late 1950s, in the provision of operational solutions to their ambitious cognitive objectives [54,55]. Deductive inference is the main focus of interest of traditional AI. Inductive inference is the basis of the MAL discipline. It may mean that, if they are not combined, deductive and inductive inference show intrinsic weaknesses in operational use, irrespective of implementation.

To outperform existing deductive and inductive inference systems, a novel trend in recent literature aims at developing *hybrid inference systems* for retrieval of sub-symbolic (e.g., leaf area index, LAI) and symbolic variables (e.g., LC and LCC classes) from sensory data (e.g., optical imagery) [25,26,56,57]. By

definition, *hybrid inference systems* combine both statistical and physical models to take advantage of the unique features of each and overcome their shortcomings [25,56].

In line with this trend, new opportunities in the design and implementation of operational hybrid RS-IUSs have been proposed to the RS community in recent years [16–24]. Significant contributions of these related works include the following:

➢ A set of quantifiable metrological/statistically-based OQIs, to be community-agreed in compliance with the GEO-CEOS QA4EO guidelines [2], is proposed to parameterize RS-IUSs for assessment and comparison purposes. The proposed set of OQIs includes: (i) degree of automation (ease-of-use), monotonically decreasing with the number of system free-parameters to be user-defined, it is also affected by the physical meaning, if any, and the range of variation (e.g., bounded, unbounded, normalized) of the system free-parameters; (ii) accuracy, e.g., thematic and spatial accuracy of a classification map; (iii) efficiency, e.g., computation time and memory occupation; (iv) robustness to changes in input parameters; (v) robustness to changes in the input data set acquired across time, space and sensors; (vi) scalability, to cope with changes in input data specifications and user requirements; (vii) timeliness, defined as the time span between sensory data collection and data-derived product generation, it increases monotonically with computer power and manpower (e.g., the manpower required to collect reference samples for training an inductive data learning system); and (viii) costs, which increase monotonically with computer power and manpower.

➢ An RS-IUS is called '*operational*' if and only if all of its OQIs, to be community-agreed (refer to this section above), score high in real-world RS image understanding (classification, mapping, recognition) problems, including RS applications at large spatial (e.g., continental, global) scale and fine semantic granularity. The proposed definition for an RS-IUS to be considered operational is not trivial. In practice, it is in contrast with a large portion of existing RS literature where the sole mapping accuracy is estimated without degree of uncertainty in toy problems at a small spatial scale or coarse semantic granularity (refer to this section above).

➢ An original three-stage, stratified, hierarchical, hybrid RS-IUS architecture is proposed to comprise the following components (for further details, refer to Section 5 in [12]).

(i) An RS image pre-processing Stage 0 (zero), including the radiometric calibration of DNs into top-of-atmosphere reflectance (TOARF) or surface reflectance (SURF) values, where SURF is a special case of TOARF in very clear sky conditions [58], *i.e.*, TOARF $\supseteq$ SURF. This radiometric calibration is mandatory, in compliance with the QA4EO guidelines [2]. In addition to ensuring the harmonization and interoperability of multi-source observational data, *radiometric calibration is considered a necessary not sufficient condition for automatic* (*hybrid model-based*) *interpretation of EO imagery*.

(ii) A symbolic, physical model-based (refer to Section 4.1 below), per-pixel pre-attentive vision first stage for RS image preliminary classification (pre-classification) (refer to Section 5 below), identified as Stage 1. It is implemented as an original, operational, automatic, near-real-time SIAM™ preliminary classifier (for further details about SIAM™, refer to Section 5.4 in [12]).

(iii) A feedback loop feeds the pre-attentive vision Stage 1 (categorical) output back to the pre-processing Stage 0 input for stratified (driven-by-knowledge, symbolic mask-conditioned) automatic RS image enhancement (e.g., stratified topographic correction [20], stratified image-coregistration, stratified image mosaic enhancement [19], *etc*.). Thus, depending on the Stage 1 output, the hybrid (numerical and categorical) input of Stage 0 is adjusted so as to reach a system steady state. This means that the proposed hybrid RS-IUS is a feedback system. It is worth mentioning that the *principle of stratification* is popular in statistics (e.g., refer to the well-known stratified random sampling design [59]). Its advantage is that "stratification will always achieve greater precision provided that the strata have been chosen so that members of the same stratum are as similar as possible in respect of the characteristic of interest" [60]. In other words, (inherently ill-posed) statistical models become better posed (conditioned, constrained) by incorporating the "stratified" or "layered" approach to accomplish driven-by-knowledge regularization (simplification) of the solution space. In general, the problem of stratification is that the collection of appropriate strata may be difficult [60]. In the proposed RS-IUS implementation where SIAM™ is adopted as preliminary classifier, symbolic strata are generated automatically as output of the pre-attentive vision Stage 1.

(iv) An attentive vision second stage battery of stratified, hierarchical, context-sensitive, application-dependent modules for class-specific feature extraction and classification, identified as Stage 2 (refer to Section 5 below). This stratified classification second stage: (i) provides a possible instantiation of a focus-of-visual-attention mechanism to mimic that adopted by attentive vision in mammals [61–64], which increases the overall degree of biological plausibility of the proposed hybrid RS-IUS; and (ii) allows second-stage (inherently ill-posed) inductive data learning algorithms, if any, to be better posed (conditioned) by symbolic prior knowledge (namely, semantic strata) stemming from the preliminary classification first stage.

To recapitulate, almost ten years from the GEOSS launch, the GEO-CEOS QA4EO guidelines have been successful in gaining attention of the RS community on the GEOSS principle of *Accessibility/Availability* of sensory data and derived products. On the other hand, the second GEOSS principle of *Suitability/Reliability* of *operational, comprehensive and timely "knowledge/information products"* derived from RS data can still be considered far from being accomplished by the RS community.

According to philosophical hermeneutics, the cause of this dichotomy is well known. The first GEOSS key principle is quantitative (unequivocal), has nothing to do with meaning and is related to the Shannon concept of '*information-as-thing*' [3]. Therefore, it is easier to deal with than the second GEOSS principle, which is qualitative (equivocal), has to deal with the meaning (interpretation, understanding) of (quantitative) data and is related to the concept of '*information-as-*(*an interpretation*)*process*' [11,12]. In the words of philosophical hermeneutics, "there is no knowledge without both an object of knowledge and a knowing subject. The claim that there is absolute knowledge, or knowledge in itself, above and beyond concrete knowing subjects, is fantastic" [11,12].

To conclude this section, the goal of this work (refer to Section 1) can be reformulated as follows: In compliance with the GEO-CEOS QA4EO guidelines, provide an original methodological contribution to the successful implementation of the inherently difficult (ill-posed) GEOSS principle of *Suitability/Reliability* in *satellite-based hybrid information/knowledge processing systems* to be used: (i) in operational mode; (ii) at spatial scales ranging from local to global; and (iii) for the estimation of both continuous biophysical variables and categorical LC and LCC variables in MS imagery acquired across time, space and sensors.

To reach its goal, the rest of this work is focused on the comparison, encompassing the four levels of understanding of an information processing system (refer to Section 1), of state-of-the-art GEOBIA and GEOOIA systems with the novel hybrid RS-IUS design and operational implementation proposed in [16–24].

3. Adopted Terminology

There are two classical types of inference (learning) known as *induction*, progressing from particular cases (e.g., true facts, training data samples, *etc.*) to a general estimated dependency or model, and *deduction*, progressing from a general model to particular cases (e.g., output values) [32].

Both deductive and inductive inference are called differently depending on the application domain and the scientific discipline. The following terms are synonyms of deductive inference and become interchangeable in the rest of this work.

(Sub-symbolic or symbolic) *deductive inference, deductive learning, top-down inference system, coarse-to-fine inference, driven-by-knowledge inference, learning-by-rules, physical model, prior knowledge-based decision system, rule-based system, expert system, syntactic inference, syntactic pattern recognition.*

The following terms are synonyms of inductive inference and become interchangeable in the rest of this paper.

(Sub-symbolic or symbolic) *inductive inference, inductive learning, bottom-up inference, fine-to-coarse inference, driven-without-knowledge (knowledge-free) inference, learning-from-examples, statistical model.*

According to philosophical hermeneutics, the discipline that studies the theory and practice of interpretation (e.g., of written texts), there are two main concepts of information embedded in more or less explicit theoretical structures: (quantitative, unequivocal, sub-symbolic) '*information-as-thing*' [3] and (qualitative, equivocal, symbolic) '*information-as-(an interpretation)process*' [4,5] (refer to Section 2).

This is tantamount to saying that, in general, *variables can be either sub-symbolic* (e.g., continuous or discrete physical variables) *or symbolic, where symbolic variables are always discrete* (e.g., categorical variables belonging to a 4-D spatio-temporal ontology of the physical world [65]). In line with the nomenclature adopted in philosophical hermeneutics [4,5], the rest of this paper considers the following terms as synonyms.

- *Symbolic, semantic, cognitive, categorical, ordinal, nominal, qualitative, subjective, equivocal.* For example, (discrete) categorical variable.
- *Sub-symbolic, sensory, numerical, non-semantic, quantitative, objective, unequivocal.* For example, continuous or discrete sensory variable (data, observables, true facts).

Hence, in the rest of this work, expressions like *sub-symbolic (either discrete or continuous) variable, symbolic (necessarily discrete) variable and sub-symbolic/symbolic information* are adopted, where sub-symbolic information is a synonym of quantitative data or '*information-as-thing*' while symbolic information is a synonym of '*information-as-(an interpretation)process*' [4,5].

It is important to note that the proposed taxonomy of *sub-symbolic/symbolic variables* does not coincide with expressions like '*supervised* (*labeled*) *data*' and '*unsupervised* (*unlabeled*) *data*' typically adopted in the MAL discipline (e.g., supervised / unsupervised data learning system). In particular, unsupervised data is always sub-symbolic, supervised data can be symbolic (for data classification applications) or sub-symbolic (for function regression). In addition, discrete sub-symbolic labeled data-objects or data-clusters can exist (as output of unlabeled data learning algorithms, like unsupervised data clustering, say, cluster 1, cluster 2, or image segmentation algorithms, say, segment 1, segment 2, *etc.*).

For example, in the MAL discipline, inductive learning-from-examples methods are either unsupervised (unlabeled) data learning algorithms (e.g., unlabeled data clustering, data quantization, image segmentation, density function estimation) or supervised (labeled) data learning algorithms for classification or function regression [32] (refer to Section 4.2 below). The former generate as output a discrete set of sub-symbolic labeled data-objects or data-clusters provided with no semantic meaning. Hence, they belong to the category of sub-symbolic inductive inference systems. Also supervised data learning algorithms for function regression belong to the category of sub-symbolic inductive inference systems, while supervised data learning classification is a synonym of symbolic inductive inference.

Analogously, in a syntactic inference system (refer to Section 4.1 below), production rules can deal with sub-symbolic variables (e.g., Newton's law of universal gravitation applies to input and output continuous physical variables) as well as categorical variables (e.g., if an instance of class *roads* hits an instance of class *private houses* then that road is assigned to class *private roadways*).

Hence, expressions like *sub-symbolic/symbolic inductive/deductive/hybrid inference* are adopted in the rest of this work, depending on whether the inference system deals with, respectively, sub-symbolic continuous/discrete variables or (symbolic and discrete) categorical variables.

4. Critical Review of AI and MAL Principles

In Section 2 it was observed that syntactic inference systems are not particularly popular in the RS literature: Neither sub-symbolic nor symbolic syntactic inference systems have almost ever been employed in the RS-IUS pre-attentive vision first stage (with the sole exception of the SPECL applications [48,49], refer to Section 2), while symbolic expert systems have been set up only in a minority of the RS-IUS implementations at the attentive vision second stage [25–31]. It means that the objective of this work, which is the development of an operational hybrid RS-IUS (refer to Section 2), asks for more prior physical knowledge than that found in existing RS-IUSs.

To provide existing RS-IUSs with an ignition of deductive inference, this section starts from a critical analysis of the peculiar properties and limitations of deductive and inductive inference at the basis of, respectively, the AI and the MAL discipline (refer to Section 2).

4.1. Deductive Inference at the Basis of AI

In traditional AI, an *expert system*, also called *syntactic inference system* [14], is the integration of the following components [66].

(i) A knowledge base, comprising a set of *production rules* (such as IF premises, e.g., facts, THEN conclusion, e.g., action) and meta rules (*i.e.*, rules to select other rules). In general, the knowledge base encompasses a s*tructural knowledge* and a *procedural knowledge* [67]. *Structural knowledge* is a synonym of a 4-D spatio-temporal ontology of the world [65], also called *world model* [25], whose graphical notation can be a *semantic (concept) network* or *conceptual graph*. A *semantic network* consists of [30,31,67,68]: (i) a hierarchical taxonomy of classes (concepts) of 4-D objects-through-time represented as nodes featuring elementary properties (attributes), called *primitives*; and (ii) spatial relations (either topological, e.g., adjacency, inclusion, *etc.*, or non-topological, e.g., distance, in-between angle, *etc.*), non-spatial relations (e.g., is-a, part-of, subset-of) and temporal relations between classes represented as arcs between nodes. *Procedural knowledge* is concerned with specific computational functions and inference capabilities [67]. Typically, it has to deal with the order of presentation of decision rules in the knowledge base. For each class, a *class grammar* exists. It consists of a set of *substitution rules* that must be followed when *words* of the *class-specific description language*, which represents the set of all words that can be used to describe objects from one class, are constructed from letters of the alphabet, where each letter corresponds to one *primitive* in the *world model* [14].

(ii) A base of input facts, e.g., sensory data, narrative descriptions of spatial facts (e.g., Sicily is at the toe of Italy), *etc.*, and output facts as results of inference rules.

(iii) A knowledge engineering interface, to codify human knowledge of domain experts into the fact base and the knowledge base. This learning paradigm is also called deductive machine teaching (MAT)-by-rules [23,55], complementary to the inductive MAL-from-examples paradigm. In the words of Lang, "transferring existing experience effectively into procedural and structural knowledge remains a challenge of AI systems... we need to carefully feed the (information processing) system with (the interpreter's) experience in a usable form" [67]. For example, "the entire process of image analysis is characterized by the transformation of knowledge. Finally, a (3-D) scene description representing the (2-D) image content should meet the (equivocal!) conceptual reality of an interpreter" [67]. Hence, there is the need "to carefully feed the system with (intepreter's or operator's) experience in a usable form" through "pro-active engagement by the operator" [67]. This must take place within a "pro-active classification" framework based on "systemic" (top-down, syntactic) class modeling, alternative to the "mechanicistic" (bottom-up, inductive) learning-from-data approach typical of the MAL discipline. It is interesting to note that the data interpretation system conceived by Lang, where the inquirer (receiver, knower, cognitive agent) plays a pro-active role in the generation of information from data, is exactly

what philosophical hermeneutics calls "fusion of horizons" that always takes place between a speaker and the listener(s) according to the concept of '*information-as-*(*an interpretation*)*process*', complementary to the concept of '*information-as-thing*', refer to Section 3 [12]. To recapitulate, terms like MAT and knowledge engineering, adopted in AI, and "fusion of horizons", used in philosophical hermeneutics, are synonyms.

(iv)   An inference engine capable of:

   a.   applying class-specific grammars for syntactic pattern recognition, namely, to decide whether an input word, consisting of a combination of letters that identifies a combination of primitives, is or is not syntactically correct according to a particular class grammar.

   b.   Logical reasoning (inference) to generate higher-level information from the knowledge and fact bases based on *rules of inference* (*transformation rules*), refer to this section below.

The four *transformation rules* typically adopted in the inference engine of an expert system are summarized below [66].

1. Deduction (*modus ponens*) rule of inference or forward chaining: $(P; P \rightarrow R) \Rightarrow R$, meaning that if fact P is true and the rule if P then R is also true, then we derive by deduction that R is also true. It is the way to test the effects of some starting fact or cause.

2. Abduction (*modus tollens*) rule of inference or backward chaining: $(R; P \rightarrow R) \Rightarrow P$, meaning that if R is true and the rule if P then R is also true, then we obtain by abduction that P is also true. It is adopted for diagnosis to discover the potential causes generating the observed facts.

3. Induction rule of inference: $(P; R) \Rightarrow (P \leftrightarrow R)$, meaning that if two facts P and R are (always observed as) concomitant, then we can derive (induce) a correlation (!) rule $P \leftrightarrow R$ that when P is true, then R is also true and vice versa. It is important to stress that, in general, *correlation relationships* highlighted by inductive reasoning *are statistical relationships which may have little or nothing to do with cause-and-effect relationships in the physical (real) world*. For example, it is well known that, in Italy, a high-value correlation exists between tourist road traffic and the leaf phenology. Obviously, this high correlation value has nothing to do with the finding of a cause-and-effect relationship between a monotonically increasing growth of leaves with tourist road traffic or *vice versa*. To summarize, *it is important to remark that inductive inference, which deals with correlation between input facts, has nothing to do with inference by abduction* (backward chaining) *or deduction* (forward chaining) *between input and output variables*. Unfortunately, *in the MAL, CV and RS disciplines, statistical systems dealing with correlation relationships between input variables are sometimes adopted to infer (unknown) cause-and-effect relationships between input and output variables, either sub-symbolic or symbolic, belonging to a 4-D spatio-temporal ontology (model) of the world-through-time.*

4. Transitivity rule of inference: $(P \rightarrow Q; Q \rightarrow R) \Rightarrow (P \rightarrow R)$, where a new rule is produced by transitivity if two different rules, the first implying Q and the second starting from Q, hold true.

"Deduction, *i.e.*, progressing from general (e.g., model) to particular cases (e.g., output values)" based on decision rules provided by an external expert or supervisor prior to looking at true facts is the main focus of interest of traditional AI [32]. Its peculiar properties and limitations are highlighted below.

➢ In the words of Sonka *et al.* ([14]; p. 283), "*Syntactic object description should be used whenever* (*quantitative sub-symbolic*) *feature description is not able to represent the* (*semantic*) *complexity of the described object and/or when the* (*semantic*) *object can be represented as a hierarchical structure consisting of simpler parts* (*simpler semantic objects*)" or, in addition to the relation *part-of*, other relations (e.g., *subset-of*, spatial relations, *etc.*) exist between classes of objects to form a *semantic network* (refer to this section above). "*The main difference between statistical and syntactic recognition is in the learning process.* (*Class*) *grammar* (*as well as semantic network*) *construction can rarely be algorithmic using today's* (*automated grammar inference-from-examples*) *approaches, requiring significant human interaction*". *In practice, class grammar and semantic network construction is still left to a human analyst based on his/her own "heuristics, intuition, experience and prior information about the problem*".

➢ In the words of Shunlin Liang ([56]; p. 2), physical models (e.g., eligible for assessing categorical variables or continuous biophysical variables from EO sensory data) are provided by a human expert (supervisor) with prior knowledge concerning the physical laws of the (4-D) world-through-time based on his/her own intuition, expertise and evidence from data observations, before the physical model starts examining the objective sensory data at hand. Thus, physical models are human-driven (herein, equivocal, refer to Section 2 [4,5,23]) abstracts (simplified representations, approximations) of reality. Physical models try to establish cause-and-effect relationships, which have nothing to do with statistical correlation (refer to this section above).

➢ In the words of Lang, "transferring existing experience effectively into procedural and structural knowledge remains a challenge of AI systems" [67], where philosophical hermeneutics is also involved with the concept of "fusion of horizons" [12] (refer to this section above).

➢ Typical advantages of (static) syntactic inference systems are listed below:

  o They 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 [56].

  o In the words of Lang: "establishing a (physical model-based) rule set is often time-, labor- and cost-intensive. But once a (physical rule-based) system is set up and proved to be transferable, the effort pays off" [67].

➢ Typical limitations of (static) syntactic inference systems are listed below:

  o In general, it takes a long time for human experts to learn physical laws of the real world-through-time and tune physical models [14,32,56].

  o They suffer from an intrinsic lack of flexibility, *i.e.*, decision rules do not adapt to changes in the input data format and users' needs, hence the knowledge base may soon become obsolete.

  o They suffer from an intrinsic lack of scalability, in particular rule-based systems are impractical for complex problems.

4.2. Inductive Inference at the basis of MAL

"Induction, *i.e.*, progressing from particular cases (e.g., training data) to generalizations (e.g., estimated dependency or model)" [32] is the main focus of interest of MAL-from-examples. Its peculiar properties and limitations are highlighted below.

➢ A typical taxonomy of symbolic inference algorithms comprises [32]: (i) Supervised (labeled) data learning algorithms for function regression-from-examples, which deals with the estimation of an output continuous variable from an input discrete and finite training set of data samples with label, where each label is the target value of the output continuous variable for that data sample; and (ii) supervised data learning algorithms for classification-from-examples, which deals with the estimation of an output (discrete) categorical variable from an input discrete and finite training set of data samples with label, where each label is the target value for that data sample of the output categorical variable. A typical taxonomy of sub-symbolic inference algorithms comprises [32]: (i) Unsupervised (unlabeled) data learning algorithms for density function estimation; (ii) unsupervised data learning algorithms for data quantization; (iii) unsupervised data learning algorithms for data clustering (providing as output sub-symbolic labeled data clusters, say, Cluster 1, Cluster 2, *etc.*, provided with no meaning); and (iv) unsupervised (knowledge-free) image segmentation algorithms (providing as output sub-symbolic labeled image-polygons, say, Polygon 1, Polygon 2, *etc.*, provided with no meaning).

➢ In the words of Cherkassky and Mulier [32] (p. 39), inductive inference "is an inherently difficult (ill-posed) problem and its solution requires *a priori* knowledge in addition to data" (refer to Section 1). This is perfectly consistent with Jacques Hadamard's definition of the ill-posed problem [69,70]. According to Hadamard, mathematical models of physical phenomena are defined as *well-posed* when they satisfy the following requirements [70]: (1) A solution exists and (2) the solution is unique. Examples of archetypal well-posed problems include the heat equation with specified initial conditions. Problems that are not *well-posed* in the sense of Hadamard, *i.e.*, problems that admit multiple solutions, are termed *ill-posed*. Inverse problems are often ill-posed [69]. For example, the inverse heat equation is not well-posed. In addition, a system is called *well-conditioned* when the solution depends continuously on the input data, in some reasonable topology. Otherwise the model is called *ill-conditioned*, meaning that a small error in the initial data can result in much larger errors in the answers. Even if a problem is well-posed, it may still be ill-conditioned. The requirement of continuity of changes of the solution with the input data is related to the requirement of stability or robustness of the solution with respect to changes in the input data set. Continuity, however, is a necessary but not sufficient condition for stability [69]. If the problem is well-posed, then it stands a good chance of solution on a computer using a stable algorithm. *If it is not well-posed, it needs to be re-formulated to become better conditioned for numerical treatment*. Typically, this involves *including additional assumptions, equivalent to prior knowledge*, to make the problem better posed, e.g., smoothness of solutions known as (Tikhonov) regularization.

➢ In the words of Shunlin Liang ([56]; p. 2), "*statistical models* (e.g., eligible for assessing categorical or continuous biophysical variables from EO sensory data) *are based on correlation relationships and... cannot account for cause-effect relationships*", refer to Section 4.1.

➢ In the words of Shunlin Liang ([56]; p. 2), *statistical models in RS data analysis are effective for summarizing local data exclusively*. This means that statistical models are usually site-specific, *i.e.*, they tend to be effective locally with small data sets exclusively. This is a consequence of the well-known central limit theorem [71]: The sum of distributions generated by a large number of independent random variables (equivalent to, say, different LC classes depicted in a RS image) tends to form a Gaussian distribution, where no "meaningful" or "natural" data entity, cluster or (sub-)structure can be identified [23]. For example, in the framework of the Global Forest Cover Change (GFCC) Project [72], a pixel-based support vector machine (SVM) [32] model selection strategy is run for each (!) image of a multi-temporal Landsat image mosaic at global scale and 30 m spatial resolution. This image-based SVM model selection strategy, which is extremely time-consuming, is required to counterbalance the aforementioned well-known limitation of statistical models, which tend to be site-specific [56].

➢ In the RS application domain it is well known that supervised data learning algorithms [32,71], whether context-insensitive (e.g., pixel-based) or context-sensitive (e.g., 2-D object-based) [73,74], require the collection of reference training samples, which are typically scene-specific, expensive, tedious and difficult or impossible to collect [1,59]. This means that in RS common practice 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-derived product generation [1].

➢ Inductive data learning decision-tree classifiers, developed by statistics and MAL (e.g., Classification And Regression Tree (CART) [75], C5.0 [76], *etc.*) to overcome limitations of traditional (static) syntactic inference systems developed by AI (refer to Section 4.1), have provided the basis for a rising interest in data mining. Their typical advantages include the following:

   o In general, inductive decision-trees are non-parametric distribution-free.
   o The tree structure enables interpretation of the explanatory nature of the independent input variables. For example, adaptive decision-trees have been widely used in RS data classification applications at regional scale [77].

Typical disadvantages of inductive data learning decision-tree classifiers developed by statistics and MAL include the following:

   ▪ The problem of learning an optimal decision tree is known to be NP-complete under several aspects of optimality and even for simple relations such as XOR. Consequently, practical decision-tree learning algorithms are based on heuristic algorithms. The result is that decision-tree learners can create over-complex trees that do not generalize the training data well.

   ▪ Inductive data learning systems available to date are unable to find even simple class grammars or discover semantic networks consisting of concepts and inter-concept relations (e.g., part-of, subset-of) [14] (refer to Section 4.1).

   ▪ Sufficient training data usually consists of hundreds or even thousands of training samples to be independently identically distributed (iid).

To recapitulate, statistical models are typically affected by the following limitations:

- ✓ They are based on correlation relationships, not to be confused with cause-and-effect relationships.
- ✓ They are usually site-specific.
- ✓ They are inherently ill-posed and require *a priori* knowledge in addition to data to become better posed for numerical treatment. It means that, to become better posed, they are semi-automatic, *i.e.*, the user, considered as a source of prior knowledge, is required to define the system free-parameters based on heuristics.
- ✓ They are unable to construct class grammars and semantic networks consisting of (semantic) concepts (as network nodes) and relations between concepts (as arcs between nodes, e.g., part-of, subset-of).
- ✓ They require adequate reference (training/testing) datasets whose cost, timeliness, quality and availability can soon become serious limiting factors on data-derived product generation.

5. Critical Review of Biological and Artificial Vision Concepts and Terminology

This section reviews the basic principles of biological vision, consisting of a pre-attentive and an attentive vision phase, to highlight their possible links to artificial vision systems encompassing CV systems and RS-IUSs.

The main role of any biological or artificial visual system is to back-project the information in the (2-D) image domain to that in the 3-D scene domain [25], see Figure 1. In greater detail, the goal of an image understanding system is to provide plausible (multiple) symbolic description(s) of a 3-D viewed-scene, which belongs to the (4-D) world-through-time and is acquired in a (2-D) image at a given time, by finding associations of sub-symbolic *image features* with symbolic classes of 4-D objects-through-time (4-D concepts-through-time, e.g., buildings, roads, *etc.*), that belongs to a so-called *world model* [25]. The *world model*, also called 4-D *spatio-temporal ontology* of the world-through-time [65], can be graphically represented as a *semantic (concept) network* (refer to Section 4.1). According to the Open Geospatial Consortium (OGC) Simple Feature Specification [65], sub-symbolic (2-D) *image features* are (0-D) points, (1-D) lines, (2-D) polygons, multi-part polygons (strata) or, *vice versa*, region boundaries (edges, contours, either closed or non-closed) provided with no semantic meaning. In the literature, image plane entities are also called image-polygons, image-objects, 2-D segments, 2-D regions, patches, parcels, blobs or tokens [78–80], considered as inputs to intermediate-level vision known as full primal sketch [13] or perceptual grouping [33,52,80].
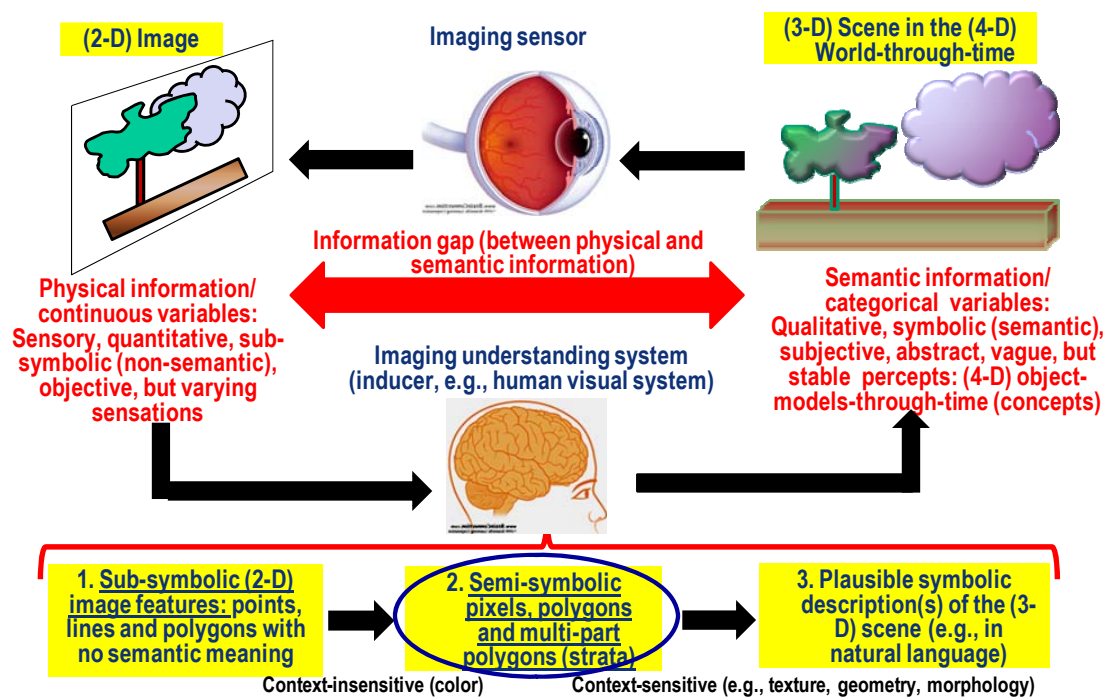
With regard to the terminology commonly adopted in the CV and RS literature, it is noteworthy that the use of the generic term 'object' is very ambiguous and, therefore, little informative because it may mean either '2-D object' in the image domain or '4-D object-model-through-time' in the *world model*.

In the rest of this paper, for simplicity's sake and in line with [25], *since sub-symbolic image-region extraction is the dual problem of sub-symbolic image-contour detection, operators for image-region extraction together with those for image-contour detection are called sub-symbolic 'segmentation' algorithms.*

Figure 1 shows there is a well-known *information gap* between the sub-symbolic image features (points, lines and polygons) and the (multiple) symbolic description(s) of a 3-D viewed-scene. This is the same *information gap* existing between continuous sub-symbolic sensory *sensations* and discrete

symbolic, semantic, linguistic, qualitative, vague, 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" [25].

Figure 1. Previously shown in [23]. Inherently ill-posed image understanding problem (vision). There is a well-known information gap between sub-symbolic (2-D) *image features* (points, lines, polygons) as input and a symbolic description (e.g., in natural language) of the 3-D viewed-scene as output [23,25,55]. To fill this gap, a pre-attentive vision first stage is expected to provide as output an image preliminary classification (pre-classification, primal sketch [13]) consisting of symbolic *semi-concepts* (e.g., spectral categories, say, '*vegetation*') [16–24]. The semantic meaning of a semi-concept is: (a) superior to zero, which is the semantic value of sub-symbolic *image features*; and (b) equal or inferior to the semantic meaning of the attentive vision *concepts* (e.g., land cover classes, say, '*needle-leaf forest*'), belonging to a *world model*, equivalent to a 4-D spatio-temporal ontology of the physical world-through-time.

In the terminology of philosophical hermeneutics, this information gap is that between the concept of '*information-as-thing*', which has nothing to do with meaning, and the concept of '*information-as-*(*an interpretation*)*process*' [4,5], refer to Section 3.

In addition to the information gap between low-level sub-symbolic sensory data and high-level symbolic information, a biological or artificial vision system has to cope with the so-called *intrinsic insufficiency* of image features [25]. It means that, due to dimensionality reduction and occlusion phenomena, image features ('*information-as-thing*') cannot be considered sufficient for a vision system to generate as output a unique (unequivocal) interpretation ('*information-as-*(*an interpretation*)*process*') of the 3-D viewed-scene, but these symbolic descriptions (for example, in natural language) of the viewed-scene can be, in general, more than one.

Finally, it is well known from the MAL literature that any inductive data-learning problem "is an inherently difficult (ill-posed) problem whose solution requires *a priori* knowledge in addition to data" ([32]; p. 39), refer to Section 4.2.

The first conclusion is that the problem of image understanding (vision), from sub-symbolic imagery to symbolic description(s) of the 3-D viewed-scene, belongs to the family of symbolic inductive data learning problems (refer to Section 4.2). As such, it is inherently ill-posed in the Hadamard sense and, consequently, very difficult to solve due to the *information gap* and the *intrinsic insufficiency* of image features [18,25]. In practice, vision is a circular (*chicken-and-egg*) dilemma, like the well-known problem of RS image topographic correction (TOC). About the latter, while image classification should be run only after TOC takes place, TOC requires a priori knowledge of surface roughness which is land cover class specific [20]. About the former, *a RS-IUS cannot detect RS image-objects without prior knowledge of the types of 3-D objects-through-time depicted by the imaging sensor; at the same time the RS-IUS cannot adopt an image-object-based classification approach without preliminary detection of the RS image-objects*. To break this circular dilemma, the only solution for image-object detection and image-object (pre-)classification would be to be solved simultaneously.

The second conclusion is a corollary of the first one. Since vision is an (inherently ill-posed) symbolic inductive inference problem, its solution requires symbolic prior knowledge in addition to (sub-symbolic) sensory data to become better posed (conditioned). For example Figure 1 shows that, before (prior to) looking at a 3-D scene, any human observer is provided with a (mental) prior knowledge of the 4-D world-through-time, called *world model* [23,65,68] (refer to Section 4.1). This is tantamount to saying that, like the human vision system, an artificial vision system must be a symbolic hybrid inference system (refer to Section 2) where concepts, to be detected as output by an attentive vision second stage, belong to a *world model* that exists before the acquisition of sensory data takes place.

With regard to design and implementation specifications, a vision system in mammals is known to comprise a pre-attentive and an attentive vision phase summarized as follows.

1. Pre-attentive (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 [61–63]. It is known that the human visual system employs at least four spatial scales of analysis [64].

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 observed in sequence and each step is examined quickly (20–80 ms) [61–63].

At this point of the analysis, an important question to answer would be: When does vision go symbolic? Since attentive vision is symbolic (to generate as output a symbolic description of the viewed-scene), a better question would be: Has pre-attentive vision anything to do with the meaning of image features? If the answer is no, then pre-attentive vision provides as output image features irrespective of their meaning (related to the concept of '*information-as-thing*', refer to this section above). If the answer is yes, then the pre-attentive meaning of image features (related to the concept of '*information-as-*(*an interpretation*)*process*', refer to this section above) must be not superior

(*i.e.*, equal or inferior) to the attentive meaning of image features, called *concepts*. To answer this question, let us consider the following inter-disciplinary contributions.

In the literature of psychophysics, according to Vecera and Farah pre-attentive image segmentation is an interactive (hybrid, see Section 2) inference process "in which top-down knowledge partly guides lower level processing" ([33]; p. 1294). For example, familiarity of stimuli, say, shape, influences image segmentation (also refer to Section 1).

In the CV literature, according to Marr "vision goes symbolic almost immediately, right at the level of zero-crossing (pre-attentive primal sketch)... without loss of information" ([13]; p. 343) (refer to Section 1), which is consistent with the Vecera and Farah quote. The Marr conjecture implies the following.

(a) The output of pre-attentive vision is a symbolic *primal sketch*, also called *preliminary classification* (*pre-classification*) map. This is tantamount to saying that:

   i.   Vision goes symbolic within the pre-attentive vision phase. This means that first-stage image segmentation (image feature extraction) and image pre-classification are solved simultaneously.

   ii. The primal sketch is a preliminary semantic map consisting of pre-attentive symbolic labels that belong to the *world model,* which exists before (prior to) looking at a 3-D viewed-scene.

   iii. The meaning of (the degree of symbolic information collected by) the pre-attentive symbolic labels must be superior to zero and not superior (*i.e.*, equal or inferior) to that of the attentive symbolic labels. The attentive symbolic labels are related to *concepts* (refer to this section above). Hence, the pre-attentive symbolic labels are called *semi-concepts*.

(b) The symbolic output of pre-attentive vision is lossless (or lossy). To be lossless, the pre-attentive mapping of a continuous sub-symbolic variable (e.g., surface reflectance) into a discrete categorical variable (semi-concepts, e.g., '*vegetation*') must be reversible. If the input image is reconstructed (synthesized) from its discrete semantic description by inverse mapping, then the reconstructed image is a piecewise constant approximation of the input image. The reconstructed image must satisfy the following constraints.

   i. The image-wide discretization (quantization) error (summary statistic) of the reconstructed image in comparison with the original image must be low.

   ii. Locally, small but genuine image details (high spatial frequency image components) of the original image must be well preserved in the reconstructed image.

   This is tantamount to saying that the inverse mapping of the symbolic pre-attentive vision pre-classification map back to the input image domain generates a piecewise constant approximation of the input image equivalent to an edge-preserving smoothing filter.

It is noteworthy that, in contradiction with his own intuitions about what a CV system design should be, the CV system implementation proposed by Marr is unable to accomplish either of the two aforementioned CV system requirements specification (a) or (b) inspired to human vision. For example, the Marr pre-attentive vision phase provides as output a sub-symbolic raw and a sub-

symbolic full primal sketch. In particular: (I) the raw primal sketch consists of a hierarchy of sub-symbolic primitives, namely, multi-scale zero-crossings ([13]; pp. 54–59), zero-crossing segments ([13]; p. 60) and level 1 image-tokens, comprising blobs (closed contours), edges, bars and discontinuities (terminations) ([13]; pp. 70–73), and (II) a full primal sketch, equivalent to perceptual grouping [33,52,80], where level 2 boundaries (e.g., texture boundaries) are detected between groups of tokens ([13]; pp. 53, 91–95). Unfortunately, Marr provided neither raw nor full primal sketch implementation details. This apparent contradiction between Marr's intuitions about the CV system design (computational theory) and his proposed CV system implementation is not at all surprising. It accounts in general for the customary distinction between a model and the algorithm used to identify it [18,23]. In particular, the inconsistency between what Marr wrote and what he implemented in practice accounts for the seminal nature of the work by Marr followed by his premature death [14]. To conclude, long-lasting inspiration from Marr's work should stem from his level of understanding of the CV system design that he considered the linchpin of success of a CV system, rather than algorithms or implementation [13,14] (also refer to Section 1).

It is also important to mention that, according to the terminology adopted in this work (refer to this section above), (sub-symbolic) *image feature extraction and* (sub-symbolic) *image segmentation are synonyms for sub-symbolic pre-attentive vision*, which is not the symbolic pre-attentive vision first stage in the Marr sense described in this section above.

To recapitulate, if the Marr quote holds true, then the third conclusion of this review section is that, in the symbolic hybrid human vision system, the ignition of symbolic prior knowledge starts at the pre-attentive vision first stage subjected to the following constraints.

(I) Symbolic pre-attentive vision is general-purpose (application-independent), parallel and rapid (efficient) to generate as output a (symbolic) preliminary classification (pre-classification) map of the input image. Hence, symbolic pre-attentive vision accomplishes image feature extraction (image segmentation) and image pre-classification simultaneously (refer to Section 1).

(II) Pre-attentive semantic labels belong to a discrete and finite set of semi-concepts whose degree of semantic information must be superior to zero and equal or inferior to that of concepts detected by the attentive vision second phase.

(III) The inverse mapping of the pre-classification map back to the input image domain generates a piecewise constant approximation of the input image equivalent to an edge-preserving smoothing filter.

The fourth conclusion holds as an extension of the third conclusion to an artificial pre-attentive vision sub-system. It states that, if an artificial pre-attentive vision first stage fails to accomplish a hybrid combination of symbolic prior knowledge with statistical inference to comply with the aforementioned constraints (I) to (III), then the inherently ill-posed image feature extraction problem cannot become better posed (conditioned) for numerical treatment (refer to Section 4.2).

To this regard it is important to stress that *image-region extraction* is the dual problem of *image-contour detection* and these complementary visual problems are both inherently ill-posed in the Hadamard sense [70] (refer to Section 4.2). Unfortunately, the inherent ill-posedness of any *image-region/image-contour* detection algorithm is explicitly acknowledged by a small portion of the CV and RS literature, e.g., it is explicitly mentioned in [6,25,33–35,69,81–84]. This may explain why,

although no "best" *image-region extraction*/*image-contour* detection approach exists, literally dozens of "novel" (supposedly better) segmentation algorithms are published each year (e.g., refer to the proprietary image segmentation algorithm implemented in the Definiens Developer GEOBIA/GEOOIA commercial software [7,10,11], which was later adapted as proposed in [9] and whose input parameters can be tuned statistically as proposed in [85]). On the other hand, in the RS literature, fortunately there are several authors whose quotes seem to agree with the aforementioned fourth conclusion, although their RS-IUS implementations do not.

For example, in the context of RS image segmentation Hay and Castilla observe that: (i) Changing the bit depth of any similarity/heterogeneity measure can lead to different image segmentation results; and (ii) even human photo-interpreters will not delineate exactly the same image segments [34]. According to Castilla *et al.* "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" [86]. Well-known driven-without-knowledge image segmentation algorithms adopted at the first stage of GEOBIA systems "... are conceptually inconsistent with the object-oriented (OO) approach... an underlying hypothesis of any segmentation method is that there is a correspondence between radiometric similarity in the (2-D) image and semantic similarity in the viewed (3-D) landscape. Thus, it is expected that (2-D) image-objects coincide with (3-D) landscape-objects" [86].

In the words of Baatz *et al.*, who are among the developers of the Definiens Developer GEOBIA/GEOOIA commercial software [7–10], "the correct extraction and shaping of objects of interest typically requires more advanced models, domain knowledge and semantics, in order to cope with the specific characteristics of the structure and to sort out ambiguities that often occur. The more or less simple and knowledge-free segmentation procedures used to produce object clusters or object primitives almost never succeed in extracting objects of interest in a robust and reliable manner. Furthermore, different types of target objects also need different strategies for their extraction" [6], *i.e.*, in RS common practice EO image segmentation is a function of the target land cover class depicted by the imaging sensor model at hand.

Like in the case of Marr (refer to this section above), the RS-IUS implementations promoted by these authors do not comply with their own intuitions at the level of computational theory (system design). This accounts for the customary distinction between a model and the algorithm used to identify it [18,23], which stands for the difference between words and facts.

For example, the Size-Constrained Region Merging (SCRM) algorithm proposed by Castilla *et al.* makes no exception to their criticism, in fact SCRM is prior knowledge-free and its "correspondence between radiometric similarity and semantic similarity is not straightforward" [86].

The same consideration holds for the driven-without-knowledge multi-scale image segmentation algorithm implemented at the pre-attentive vision first stage of the Definiens Developer GEOBIA/GEOOIA commercial software by Baatz *et al.* [7–10], which is in contradiction with these authors' own statement that "the correct extraction and shaping of objects of interest typically requires more advanced models, domain knowledge and semantics, in order to cope with the specific characteristics of the structure and to sort out ambiguities that often occur" [10].

6. The GEOBIA Paradigm

Paradigm is intended here as the generally accepted perspective of a given discipline at a particular time [35]. In this section the GEOBIA objectives and definitions are proposed before presenting the GEOBIA system design and implementation.

6.1. Review of the GEOBIA Objectives and Definitions

In [34,35,67], GEOBIA is defined as a sub-discipline of Geographic Information science (GIScience), also known as *geomatics* engineering [66], devoted to:

(1) The automatic or semi-automatic partitioning (segmentation, aggregation, simplification) of a raster RS image, consisting of sub-symbolic unlabeled pixels, into discrete sub-symbolic labeled image-objects, where the sub-symbolic label is a segment identifier (e.g., an integer number, say, Segment 1, Segment 2, *etc*.), such that each discrete image-object is a connected set of pixels whose visual (appearance, pictorial) properties are considered relatively homogeneous with respect to their surroundings according to a measure of similarity chosen subjectively based on its ability to create "interesting" ("meaningful") image-objects.

(2) The automatic or semi-automatic mapping (projection) of sub-symbolic labeled image-objects onto a discrete and finite set of symbolic 4-D object-models-through-time belonging to a *world model* (refer to Section 5) [25,65,68], depending on the image-object-specific spatial, spectral and temporal characteristics, so as to generate as output symbolic vector geospatial information in a Geographic Information System (GIS)-ready format.

In terms of *induction* and *deduction* rules of inference [32] (see Section 4), the GEOBIA system architecture can be summarized as follows:

(Inductive) Unsupervised data learning (e.g., image segmentation, unlabeled data clustering) pre-attentive vision first stage
+ (in series with)
Attentive vision second stage implemented as an (inductive) 2-D object-based supervised data learning classifier or a (deductive) 2-D object-based syntactic classifier.
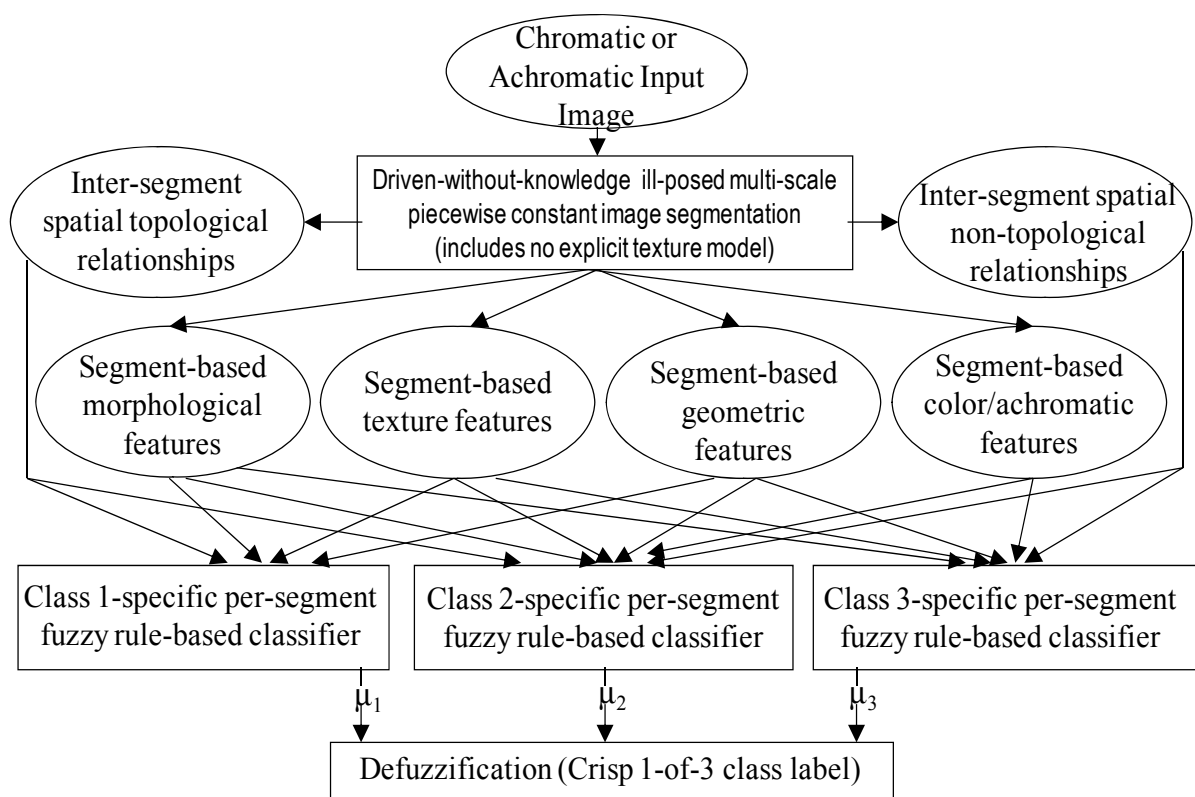
About the GEOBIA commitments, Hay and Castilla propose that "the primary objective of GEOBIA as a discipline is to develop appropriate theory, methods and tools sufficient to replicate (and or exceed experienced) human interpretation of RS images in automated/semi-automated ways, that will result in increased repeatability and production, while reducing subjectivity, labor and time costs" [34,35]. In [67], Lang states that since *automation* is the overall aim of GEOBIA (like that of any other computer-based technique), the ultimate benchmark of GEOBIA mimics human perception.

Since expressions like 'cognitive', 'semantic' and 'equivocal' are synonyms of "human interpretation" (refer to Section 3) [4,12,23], there is a contradiction of terms in the GEOBIA aim of developing automatic cognitive systems by reducing subjectivity. In addition, the GEOBIA claim of mimicking human vision (refer to Section 5) remains more an expression of intentions than a fact.

At least in part, it may be due to these conceptual incongruities, found at a level of abstraction as high as the GEOBIA objectives, if GEOBIA still lacks general consensus and research in the scientific

community [34,35] while, in common practice, it is affected by low productivity (refer to Section 2). For a detailed discussion of the GEOBIA weaknesses, refer to Section 8.2 below.

Figure 2. Previously shown in [18]. Data flow diagram (DFD) of a two-stage non-iterative geographic object-based image analysis (GEOBIA) architecture according to [6], based on the GEOBIA terminology introduced in [34,35]. In a DFD, processing blocks are shown as rectangles and sensor-derived data products as circles [87]. Pre-attentive vision image simplification is pursued by means of an inherently ill-posed driven-without-knowledge image segmentation approach that generates as output a sub-symbolic segmentation map, either single-scale or multi-scale, where each image-object is identified by a sub-symbolic (e.g., numerical) label (e.g., segment 1, segment 2, *etc*.) featuring no semantic meaning.

6.2. Two-Stage Non-Iterative GEOBIA Architecture

In the last fifty years a huge variety of RS-IUS architectures and implementations has been proposed in the literature, e.g., refer to seminal works about hybrid inference systems for high spatial resolution (HR) and very HR (VHR) image understanding by Nagao, Matsuyama and Shang-Shouq Hwang [25,26], Shackelford and Davis [27–29], *etc*. To overcome the well-known limitations of traditional pixel-based (non-contextual) supervised data learning classifiers [88,89], which are typically affected by a salt-and-pepper classification noise effect especially when dealing with VHR spaceborne imagery (<5 m), context-sensitive RS-IUSs, including GEOBIA as a special case, have been investigated in the last thirty years [26]. In recent years the term GEOBIA was proposed by Hay and Castilla to differentiate GEOBIA from 2-D object-based image analysis (OBIA) in CV and biomedical imaging [34,35].

The GEOBIA paradigm has traditionally been identified with a two-stage non-iterative GEOBIA architecture where the pre-attentive vision first stage is implemented as an inductive driven-without-knowledge (knowledge-free [6]) image segmentation algorithm followed by an attentive vision second stage implemented as a 2-D object-based classification module provided with no feedback mechanism, see Figure 2 [6,78].

Since the year 2000, mainly due to the availability of a series of commercial GEOBIA software products developed by the German company Definiens (e.g., eCognition v1, presented in the year 2000, to eCognition v4, followed by Professional 5 and Developer v4, launched in the year 2003, up to Developer v8.64 proposed in 2011) [7–11], GEOBIA systems have quickly gained widespread popularity (especially in Europe) and are currently considered the state-of-the-art in both scientific and commercial thematic mapping of VHR spaceborne imagery.

Unfortunately, despite its commercial success, the GEOBIA approach remains affected by a lack of general consensus and research, as acknowledged by the existing literature [6,34,35]. In other words, in RS common practice traditional GEOBIA systems score low in at least one of their OQIs (refer to Section 2). For example, to the best of these authors' knowledge no existing GEOBIA system has ever been successful in generating thematic maps from RS image mosaics at a continental or global scale [90], although GEOBIA projects of regional/national spatial extent with spatial resolution ranging from medium ($\approx$30 m) to HR ($\approx$10 m) have been implemented in recent years [91].
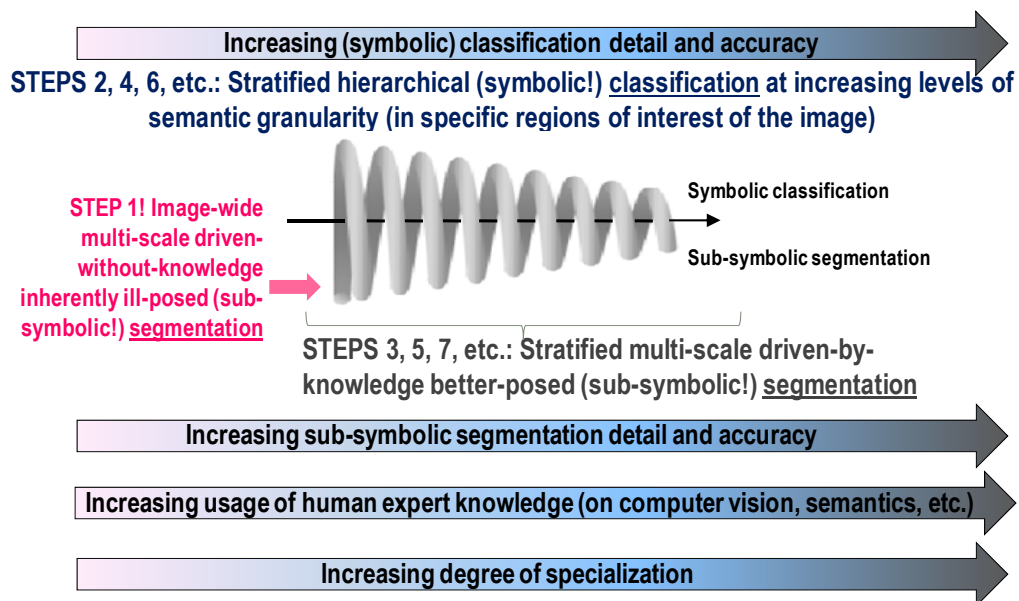
7. Three-Stage Iterative GEOOIA Architecture

To better deal with different applications, users and target classes, *i.e.*, to increase its OQIs, the process of GEOBIA should rather be considered cyclic (iterative) [67]. To reach this objective, the three-stage iterative GEOOIA approach was proposed by Baatz *et al.* [6] (see Figure 3). To be compared with the traditional two-stage non-iterative GEOBIA design depicted in Figure 2, the three-stage iterative GEOOIA architecture, shown in Figure 4, consists of a series of:

(1) An (inherently ill-posed) driven-without-knowledge (non-stratified, symbolic mask-unconditioned) image segmentation pre-attentive vision first stage, in common with the two-stage non-iterative GEOBIA design (refer to Section 8.2 below),

(2) An attentive vision second-stage battery of 2-D object-based class-specific classification modules that introduce semantics, implemented as (inductive) 2-D object-specific supervised data learning classifiers (e.g., SVMs) and/or (deductive) 2-D object-specific decision rule-set classifiers (refer to Section 4),

(3) A battery of stratified (symbolic mask-conditioned) class-specific driven-by-knowledge segmentation algorithms eligible for improving the segmentation locally for each specific class, where steps (2) and (3) can be iterated hierarchically according to the well-known problem-solving principle of divide-and-conquer (*dividi et impera*) [71], which is typically adopted by decision-trees (where it is known as the "stratified" or "layered" approach) [89,92]. The principle of *stratification* is also well known in statistics [60] (refer to Section 2 above and also to Section 5.1 in [12]). In the Definiens GEOOIA commercial software products the "stratification" principle is called "class filter" such that image objects will be part of the search domain, called the "image object domain", if they are classified with one of the classes selected

in the class filter [7,8]. In practice, this iterative approach approximates the "focus of visual attention" mechanism adopted by the human attentive vision second phase [16–25] (refer to Section 5). On the contrary, any "layered" approach is absent from the traditional two-stage non-iterative GEOBIA design shown in Figure 2, which is the reason why GEOBIA is outperformed by the GEOOIA scheme [6].

Figure 3. Sketch of the GEOOIA iterative procedure.

It is noteworthy that:
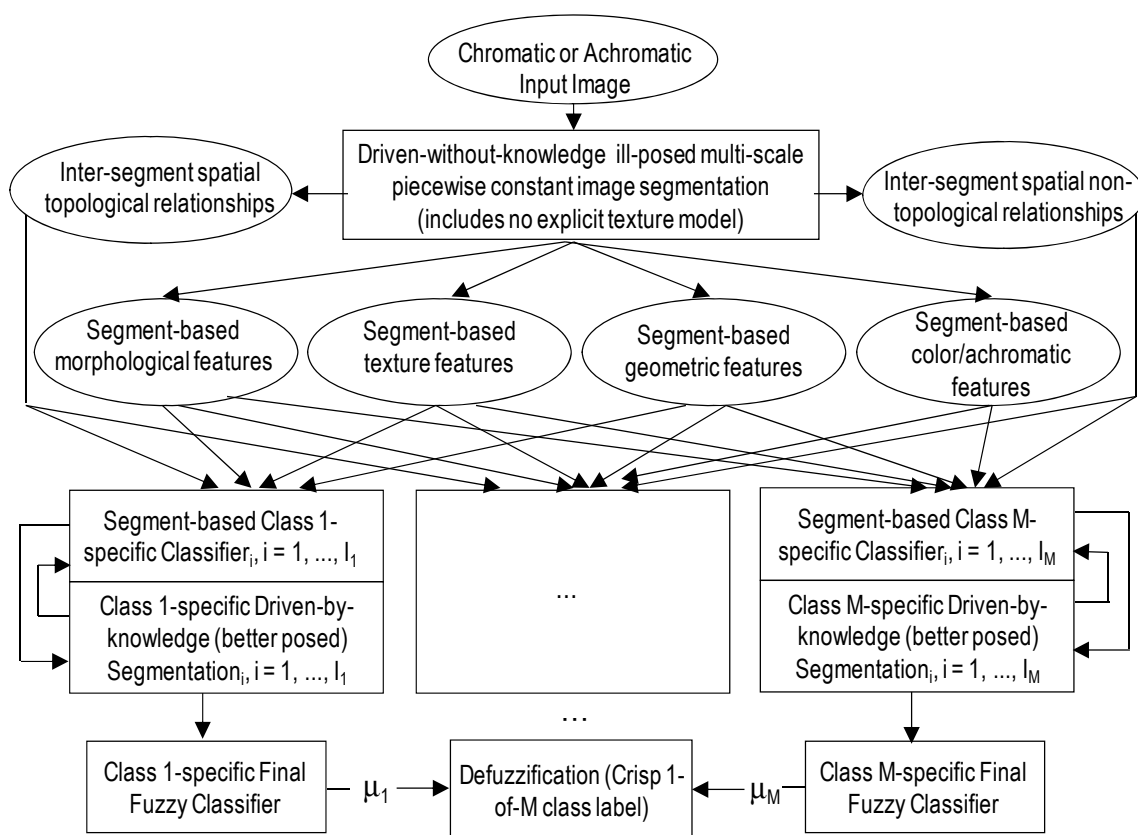
(i) the GEOBIA architecture shown in Figure 2 can be considered a special case of the GEOOIA scheme depicted in Figure 4, *i.e.*, GEOOIA ⊃ GEOBIA.

(ii) Whereas GEOOIA inherits from GEOBIA the inherent ill-posedness of the driven-without-knowledge image segmentation pre-attentive vision first stage (refer to Section 5), the iterative GEOOIA second and third stages are expected to hierarchically introduce additional supervised knowledge (for example, in the form of user-defined parameters, labeled data sets for the training/testing of inductive systems for classification-from-examples, prior knowledge-based syntactic rule sets or prior knowledge-based selection of symbolic strata, *etc.*). The amount and costs of this supervised knowledge are expected to decrease monotonically with iterations according to a divide-and-conquer problem solving principle. This supervised knowledge is equivalent to assumptions eligible for making the inherently ill-posed RS image mapping problem better posed (conditioned) for numerical treatment (refer to Section 4.2).

(iii) In terms of *induction* and *deduction* rules of inference [32] (see Section 4), the GEOOIA system architecture can be summarized as follows:

(Inductive) Unsupervised data learning (e.g., image segmentation, unlabeled data clustering) pre-attentive vision first stage
+ (in series with)

Attentive vision second stage implemented as an (inductive) 2-D object-based supervised data learning classifier or a (deductive) 2-D object-based syntactic classifier + (in series with)

Optional iteration(s): driven-by-knowledge (stratified, symbolic mask-conditioned) sub-symbolic pre-attentive vision first stage followed by driven-by-knowledge attentive vision second stage.

Figure 4. DFD of a three-stage iterative GEOOIA architecture derived from the sketch shown in Figure 3. In this DFD, processing blocks are shown as rectangles and sensor-derived data products as circles [87]. For more details about this RS-IUS scheme, refer to the text.

8. SWOT Analysis of GEOBIA/GEOOIA

In [34] and [35] Hay and Castilla propose a SWOT analysis (see Table 2) of GEOBIA to provide a better understanding of its current status and potential strategies to achieve its stated objectives. This section expands those previous analyses to the GEOOIA scheme depicted in Figure 4, where GEOOIA ⊃ GEOBIA (refer to Section 7).

Table 2. Strengths, Weaknesses, Opportunities and Threats (SWOT) matrix [34].

	Helpful in Achieving the Objective	Harmful to Achieving the Objective
Internal (attributes of the organisation)	Strengths	Weaknesses
External (attributes of the environment)	Opportunities	Threats

8.1. GEOBIA/GEOOIA Strengths (*Due to Internal Drivers*)

The following analysis of GEOBIA/GEOOIA internal strengths is inspired in part by that found in [34,35].

➢ According to Section 6.1, the aim of GEOBIA/GEOOIA is to partition an image into discrete sub-symbolic 2-D objects and provide a structural description of these 2-D objects in a way akin to that of a human photo interpreter observing a 3-D viewed-scene of the 4-D world-though-time [34,35]. To perform symbolic reasoning together with spatial reasoning as successfully as in biological vision (refer to Section 5) [25], object-based image analysis in place of traditional pixel-based image analysis is required, since image-objects exhibit useful contextual features (e.g., per-object shape, within-segment texture) and spatial topological relationships (e.g., adjacency, inclusion, *etc.*) that single pixels lack [34,35].

➢ Using image-objects reduces the number of information primitives of a classifier by orders of magnitude when basic units are the image pixels [34].

➢ Image-objects can be more readily integrated in vector-based GIScience (geomatics [66]) than thematic maps generated from pixel-based classifiers [34], which are typically affected by salt-and-pepper classification noise effects and are difficult to transform into a vector data format [16–24].

➢ Several GEOBIA/GEOOIA methods/commercial software packages, built upon the powerful object-oriented (OO) paradigm, exist [34,35], e.g., refer to [7–10,81].

8.2. GEOBIA/GEOOIA Weaknesses (*Due to Internal Drivers*)

The following analysis of GEOBIA/GEOOIA internal weaknesses enhances that found in [6,34,35].

➢ Both GEOBIA and GEOOIA commercial software products listed in Table 1 do not comply with the QA4EO requirements (refer to Section 2). The relaxation of the system requirement of radiometric consistency of multi-source, multi-temporal and MS imagery brings with it, as an inevitable consequence, that these RS-IUS commercial software products are based on (inherently ill-posed) statistical rather than physical models, which means they are intrinsically semi-automatic and site-specific (refer to Section 4.2).

➢ The fourth conclusion of Section 5 is that, in a hybrid RS-IUS, semantic prior knowledge should be ignited starting at the pre-attentive vision first stage under several functional requirements derived from human vision. For this ignition to occur, a MAT-by-rules paradigm, also called knowledge engineering in AI and "fusion of horizons" in philosophical hermeneutics, must be adopted (refer to Section 4.1). Unfortunately, syntactic pattern recognition requires significant human interaction, but once a physical model-based rule set is tuned and proved to be transferable, the effort pays off (refer to Section 4.1). On the contrary, GEOBIA/GEOOIA systems do the following.

  ✓ The pre-attentive vision first stage is nearly always implemented as a sub-symbolic statistical model-based image segmentation algorithm (refer to Section 2).

- ✓ At the attentive vision second stage, symbolic syntactic inference may or may not be employed. If it is not, the GEEOBIA/GEOOIA system is fully statistical.
- ✓ An attempt to convey additional user knowledge into the GEOBIA framework is provided by the three-stage iterative GEOOIA architecture (see Section 7), but GEOOIA shares with GEOBIA the sub-symbolic statistical model-based pre-attentive vision first stage [6].

To summarize, both GEOBIA and GEOOIA systems are not biologically plausible (refer to Section 5), which is in contrast with their original goal of attempting to replicate human vision (refer to Section 6.1).

➤ The inherent ill-posedness of any sub-symbolic inductive image-object extraction/image-contour detection algorithm adopted at the GEOBIA/GEOOIA pre-attentive vision first stage is the driver of both systematic and accidental errors. The former are related to the so-called *intrinsic insufficiency* of image features (refer to Section 5), the latter are related to the fact that image-objects are always affected by a so-called *artificial insufficiency* due to the image segmentation algorithm at hand [25]. This second source of segmentation errors is also known as the *uncertainty principle* according to which, *for any contextual (neighborhood) property, we cannot simultaneously measure that property while obtaining accurate localization* [82,83]. In practical contexts the inherent ill-posedness of any knowledge-free image segmentation algorithm implies the following.

- ▪ In real-world applications (other than toy problems), it is inevitable for erroneous segments to be detected while genuine segments are omitted ([25]; p. 18).
- ▪ System free-parameters are required to work as additional assumptions necessary to make the inherently ill-posed image segmentation problem better posed for numerical treatment (refer to Section 4.2). Unfortunately, image segmentation parameters are always site-specific and must be user-defined based on heuristics and a trial-and-error approach. For example, in the case of the popular Baatz *et al*. segmentation algorithm adopted by the pre-attentive vision first stage of the Definiens GEOBIA/GEOOIA commercial software products [10], statistical methods have been developed to automatically optimize the parameters based on a site-specific training set of reference image-objects [9,85,93].

To summarize, in general, with regard to the set of OQIs introduced in Section 2, any sub-symbolic, inductive, driven-without-knowledge image segmentation algorithm tends to score as follows: (i) it is difficult to use because its degree of automation, which is monotonically decreasing with the number of system free-parameters to be user-defined, tends to be low; (ii) accuracy tends to be low; (iii) robustness to changes in the input data set is low; (iv) robustness to changes in input parameters tends to be low; and (v) timeliness tends to be high.

➤ Under the guise of 'flexibility', current GEOBIA/GEOOIA commercial software products provide at both the pre-attentive vision first stage and the attentive vision second stage overly complicated collections of algorithms to choose from based on heuristics (e.g., the Definiens Developer v8 process list comprises: 6× segmentation, 4× classification, 6× advanced classification, 4× variable operation, 9× reshaping, 3× level operation, *etc*.) [34,35]. In RS

common practice commercial GEOBIA/GEOOIA software products appear affected by a combination of three limitations.

✓ Options to choose from mainly consist of statistical models for retrieving land surface variables, either sub-symbolic continuous variables or symbolic categorical variables, from RS imagery. Peculiar properties and limitations of inductive inference in RS data applications are well known in the existing literature, refer to Section 4.2.

✓ Lack of physical models, based on prior observations of the physical world-through-time, for retrieving land surface variables, either sub-symbolic continuous variables or symbolic categorical variables, from RS imagery [56]. This holds so true that none of the existing commercial GEOBIA/GEOOIA software products listed in Table 1 considers RS data radiometric calibration, namely, the transformation of DNs into physical units of radiometric measure, as a pre-processing step mandatory before investigating RS images acquired across space, time and sensors. In practice, none of the existing commercial GEOBIA/GEOOIA software products listed in Table 1 agrees with the QA4EO guidelines [2] (refer to Section 2).

✓ The RS-IUS free-parameter selection and the combination of pre-attentive vision first-stage segmentation and attentive vision second-stage classification algorithms are delegated to the full responsibility of the user whose scientific rationale and expertise may be extremely subjective, empirical and/or inadequate for such a complex task. This freedom of choice makes the definition and implementation of the GEOBIA and GEOOIA workflows more similar to (subjective, qualitative) art than (objective, quantitative) science.

➤ Image-segments can be described by a segment description table [26], whose columns consist of: (a) a segment sub-symbolic label or identifier, typically an integer number; (b) a segment symbolic label, if any, belonging to a 4-D spatio-temporal ontology; and (c) segment-specific quantitative descriptors (primitives) such as [25]: (i) locational properties (e.g., minimum enclosing rectangle); (ii) photometric properties (e.g., mean, standard deviation, *etc.*); (iii) geometric/shape properties (e.g., area, perimeter, compactness, straightness of boundaries, elongatedness, rectangularity, number of vertices, *etc.*); (iv) texture properties [94]; (v) morphological properties [95]; (vi) spatial non-topological relationships between objects (e.g., distance, angle/orientation, *etc.*); (vii) spatial topological relationships between objects (e.g., adjacency, inclusion); (viii) temporal relationships between objects, *etc.* In common practice image segmentation algorithms are demanding in terms of both computation time and memory occupation. For example, since the second-stage classifiers of both GEOBIA and GEOOIA (see Figures 2 and 4, respectively) employ sub-symbolic image-objects as information units exclusively, when pixel-based spectral properties are sufficient for classification purposes the image segmentation first stage of both GEOBIA and GEOOIA requires the transformation of each pixel into a one-pixel segment, which is trivial and time-consuming. It is noteworthy that, alternative to the GEOBIA/GEOOIA systems, RS-IUS instances found in the existing literature, such as the Shackelford and Davis RS-IUS implementations proposed in [27–29], provide examples of a combined pixel- and 2-D object-based classification approach where pixels and 2-D objects co-exist as spatial information primitives.

➢ In general, there are numerous challenges involved in the segmentation of very large data sets such as complex tiling and restricted memory availability, which require close monitoring of the number of image-objects in a project. Recent developments in hardware (e.g., availability of 64-bit central processing units, multiple processing, *etc*.) and software (e.g., the Definiens 64 bit-based Developer v8.64 software product [8]) may mitigate operational limitations of GEOBIA and GEOOIA systems in dealing with large data sets.

➢ As a result of the bullets listed above, to date there is a lack of consensus and research on the conceptual foundations of GEOBIA/GEOOIA [34,35], together with an unquestionable lack of productivity (refer to Section 2). For example, it is acknowledged in the literature that the rule base developed at the attentive vision second stage of a GEOBIA/GEOOIA scheme tends to be non-transferable to other applications [91].

8.3. GEOBIA/GEOOIA Opportunities (*Due to External Drivers*)

The following analysis of GEOBIA/GEOOIA external opportunities is largely inspired by that found in [34,35].

➢ Object-oriented concepts and methods have been successfully applied to many different problems, not only computer languages, and they can easily be adapted to GEOBIA/GEOOIA even when they stem from biomedical imaging and CV which, unfortunately, remain unknown to most of the RS community [34,35].

➢ There are new information technology (IT) tools (e.g., wikis), which may accelerate consensus and cohesion of GEOBIA/GEOOIA [34,35].

➢ There is a steadily growing community of RS and GIS practitioners who currently use image segmentation for different geographic information applications. Thus, as GEOBIA/GEOOIA matures, new commercial/research opportunities will come into existence to customize 2-D object-based solutions for specific fields, disciplines and user needs [34,35].

➢ Image segmentation is traditionally computation intensive and requires large memory occupation to deal with a segment description table [26], see Section 8.2. Hardware developments such as symmetric multiprocessing, parallel processing and grid computing, together with software developments (e.g., refer to the Definiens Developer v8.64 software product [8]), are recent technologies that GEOBIA/GEOOIA methods may benefit from in tackling the analysis of large data sets [34,35].

8.4. GEOBIA/GEOOIA Threats (*Due to External Drivers*)

The following analysis of GEOBIA/GEOOIA external threats is largely inspired by that found in [34,35].

➢ Since much remains to be solved, GEOBIA/GEOOIA is far from being an operationally established paradigm [34,35]. In particular, the inherent ill-posedness of image-region extraction/image-contour detection continues to be largely underestimated or, worse, ignored by a large portion of the RS community, see Section 5.

➢ Trying to make GEOBIA/GEOOIA distinct from other object-oriented concepts and methods (e.g., by using terms like 'object-based' instead of the traditional expression 'object-oriented') may contribute to insulation of GEOBIA/GEOOIA users in an esoteric world of 2-D 'objects' and isolation of the GEOBIA/GEOOIA paradigm rather than to its consolidation [34,35].

➢ The visual appeal of discrete geographic image-objects (geo-objects [65], geons [67]), their easy integration with GIScience and the enhanced classification possibilities of GEOBIA/GEOOIA systems with respect to traditional pixel-based classifiers have attracted the attention of major RS image processing vendors, who are increasingly incorporating new segmentation tools in their packages. This provides a wider choice for practitioners, but promotes confusion (among different packages, options, syntax, *etc*.) and makes it more difficult to reach a consensus on what GEOBIA/GEOOIA is all about. A lack of protocols, formats, and standards may lead to a splitting of the GEOBIA/GEOOIA field into sub-fields rather than a consolidation of GEOBIA/GEOOIA as a discipline [34,35].

9. Conclusions

Split into two parts for publication purposes, this methodological work provides the remote sensing (RS), computer vision (CV), artificial intelligence (AI) and machine learning (MAL) communities with several multi-disciplinary conclusions of practical interest for developing a new generation of RS image understanding systems (RS-IUSs) whose quality indicators (QI) of operativeness (OQIs) (refer to Section 2) are expected to score high in real-world RS applications, including RS image understanding at large (e.g., global) spatial scale and fine semantic granularity, in compliance with the Group on Earth Observations (GEO)-Committee on Earth Observation Satellites (CEOS) Quality Assurance Framework for Earth Observation (QA4EO) guidelines [2].

This section provides a useful summary of the multi-disciplinary conclusions of the first part of this theoretical work together with links to the text.

(1) Vision is a symbolic inductive learning problem (from sub-symbolic true facts to symbolic generalizations). As such, to cope with its inherent ill-posedness due to the *information gap* and the *intrinsic insufficiency* of sub-symbolic image features (image-objects or, *vice versa*, image-contours), any vision system, either biological or artificial, requires symbolic prior knowledge in addition to sub-symbolic data to become better posed (conditioned). It means that any vision system must be a *symbolic hybrid inference* system, refer to Section 5.

(2) In the CV literature, according to Marr "vision goes symbolic almost immediately, right at the level of zero-crossing (pre-attentive primal sketch) ... without loss of information" [13] (p. 343). If this conjecture holds true in compliance with evidence provided by Vecera and Farah (image segmentation is an "interactive" (hybrid) inference process "in which top-down knowledge partly guides lower level processing") [33] (p. 1294), then the symbolic hybrid human vision system comprises a symbolic hybrid pre-attentive vision sub-system subjected to the following constraints (refer to Section 5).

 (I)   Symbolic pre-attentive vision is general-purpose (application-independent), parallel and rapid (efficient). It generates as output a (symbolic) preliminary classification (pre-classification)

map of the input image. Hence, the symbolic pre-attentive vision first stage accomplishes image feature extraction (image segmentation) and image pre-classification simultaneously.

(II) Symbolic pre-attentive labels belong to a discrete and finite set of semi-concepts whose degree of semantic information must be superior to zero and equal or inferior to that of concepts detected by the attentive vision second phase.

(III) The inverse mapping of the pre-classification map back to the input image domain generates a piecewise constant approximation of the input image equivalent to an edge-preserving smoothing filter where image details featuring high spatial-frequency components are well preserved.

(3) To be considered inspired to human vision, an artificial pre-attentive vision sub-system should comply with the aforementioned requirements (I) to (III), refer to Section 5.

(4) Despite their commercial success, state-of-the-art two-stage non-iterative Geographic Object-Based Image Analysis (GEOBIA) systems (refer to Section 6) and three-stage iterative Geographic Object-Oriented Image Analysis (GEOOIA) systems, where GEOOIA ⊃ GEOBIA (refer to Section 7), remain affected by a lack of productivity, general consensus and research, as pointed out in existing literature [6,34,35] (refer to Section 2). An original Strengths, Weaknesses, Opportunities and Threats (SWOT) analysis of the GEOBIA/GEOOIA systems highlights the following (see Section 8.2).

✓ Popular GEOBIA and GEOOIA commercial software products, like those listed in Table 1, do not comply with the QA4EO requirements (refer to Section 2). The relaxation of the requirement of radiometric consistency of multi-source, multi-temporal and multi-spectral (MS) imagery brings, as an inevitable consequence, that these RS-IUS commercial software products are based on (inherently ill-posed) statistical rather than physical models, which means they are intrinsically semi-automatic and site-specific (refer to Section 4.2).

✓ Both GEOBIA and GEOOIA are not biologically plausible, which is in contrast with their original goal of attempting to replicate human vision (refer to Section 6.1).

♦ In place of a symbolic pre-attentive vision first stage capable of accomplishing the aforementioned requirements (I) to (III) inspired to replicate human vision, both GEOBIA and GEOOIA adopt the same sub-symbolic statistical approach.

♦ At the attentive vision second stage, both GEOBIA and GEOOIA may or may not employ symbolic syntactic inference. If they do not, they are fully statistical systems.

✓ Any structural ill-posedness of GEOBIA, which is inherited by GEOOIA at the sub-symbolic pre-attentive vision first stage, is eventually mitigated at the GEOOIA second and third stages iteratively by additional ignitions of user supervision. This iterative process, where human supervision is expected to monotonically decrease with iterations, is equivalent to a well-known divide-and-conquer problem solving approach. In practice, it approximates a "focus of visual attention" mechanism adopted by the human attentive vision second phase (refer to Section 7).

To recapitulate, when compared to human vision, GEOBIA and GEOOIA systems lack deductive inference mechanisms starting at their pre-attentive vision first stage.

The degree of novelty of the proposed conclusions can be considered relevant because:

- They encompass the four levels of understanding of a CV system or RS-IUS considered as an information processing system, namely: (a) computational theory (system architecture), (b) information/knowledge representation, (c) algorithm design and (d) implementation (refer to Section 1).
- They are complementary to conclusions proposed by a large portion of the existing literature where RS data mapping algorithms are tested in toy problems at small (e.g., local) spatial scale or coarse semantic granularity. Unfortunately, scalability of these latter approaches to real-world RS image understanding problems at a large (e.g., global) spatial scale and fine semantic granularity appears questionable or remains unknown (refer to Section 2).

To comply with the QA4EO requirements and the symbolic pre-attentive vision sub-system constraints (I) to (III) listed in this section above, a novel hybrid RS-IUS design and implementation, where the operational, automatic, near real-time SIAM™ decision-tree preliminary classifier is adopted as its symbolic pre-attentive vision first stage, is selected from the existing literature [16–24] and discussed in the second part of this work.

In the RS literature, expert systems have been (almost) exclusively proposed in the attentive vision second-stage classification [25–31]. To the best of these authors' knowledge, this is the first time a symbolic syntactic inference system, like SIAM™, is made available to the RS community for operational use in a RS-IUS pre-attentive vision first stage, to accomplish multi-scale image segmentation and multi-granularity image pre-classification simultaneously, automatically and in near real-time.

Acknowledgments

References

1. Gutman, G., Janetos, A.C., Justice, C.O., Moran, E.F., Mustard, J.F., Rindfuss, R.R., Skole, D., Turner, B.L., Cochrane, M.A., Eds.; *Land Change Science*; Kluwer: Dordrecht, The Netherlands, 2004.

2. GEO/CEOSS. *A Quality Assurance Framework for Earth Observation: Operational Guidelines Version 3.0*. Available online: http://calvalportal.ceos.org/cvp/c/document_library/get_file?p_l_id=17516&folderId=17835&name=DLFE-304.pdf (accessed on 10 January 2012).

3. Shannon, C. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423, 623–656.

4. Capurro, R.; Hjørland, B. The concept of information. *Annu. Rev. Inform. Sci.* **2003**, *37*, 343–411.

5. Capurro, R. Hermeneutics and the Phenomenon of Information. In *Metaphysics, Epistemology, and Technology. Research in Philosophy and Technology*; JAI/Elsevier: Amsterdam, The Netherlands, 2000; Volume 19, pp. 79–85.

6. 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, USA, 2008; Chapter 1.4, pp. 29–42.

7. Definiens Imaging GmbH. *eCognition Elements User Guide 4*; Definiens Imaging GmbH: Munich, Germany, 2004.

8. Definiens AG. *Developer 8 Reference Book*; Definiens AG: Munich, Germany, 2011.

9. Esch, T.; Thiel, M.; Bock, M.; Roth, A.; Dech, S. Improvement of image segmentation accuracy based on multiscale optimization procedure. *IEEE Geosci. Remote Sens. Lett.* **2008**, *5*, 463–467.

10. Baatz, M.; Schäpe, A. Multiresolution Segmentation: An Optimization Approach for High Quality Multi-Scale Image Segmentation. In *Angewandte Geographische Informationsverarbeitung XII*; Strobl, J., Ed.; Herbert Wichmann Verlag: Berlin, Germany, 2000; Volume 58, pp. 12–23.

11. Nuebert, M.; Herold, H.; Meinel, G. Assessing Image Segmentation Quality-Concepts, Methods and Apllication. 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, USA, 2008; Chapter 8.3, pp. 769–784.

12. Baraldi, A.; Boschetti, L. Operational automatic remote sensing image understanding systems: Beyond Geographic Object-Based and Object-Oriented Image Analysis (GEOBIA/GEOOIA). Part 2: Novel system architecture, information/knowledge representation, algorithm design and implementation. *Remote Sens*. **2012**, accepted.

13. Marr, D. *Vision*; Freeman and Company: New York, NY, USA, 1982.

14. Sonka, M.; Hlavac, V.; Boyle, R. *Image Processing, Analysis and Machine Vision*; Chapman & Hall: London, UK, 1994.

15. Baraldi, A.; Bruzzone, L.; Blonda, P. Quality assessment of classification and cluster maps without ground truth knowledge. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 857–873.

16. Baraldi, A. Impact of radiometric calibration and specifications of spaceborne optical imaging sensors on the development of operational automatic remote sensing image understanding systems. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens*. **2009**, *2*, 104–134.

17. Baraldi, A.; Puzzolo, V.; Blonda, P.; Bruzzone, L.; Tarantino, C. Automatic spectral rule-based preliminary mapping of calibrated Landsat TM and ETM+ images. *IEEE Trans. Geosci. Remote Sens*. **2006**, *44*, 2563–2586.

18. Baraldi, A.; Durieux, L.; Simonetti, D.; Conchedda, G.; Holecz, F.; Blonda, P. 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 Sens.* **2010**, *48*, 1299–1325.

19. Baraldi, A.; Durieux, L.; Simonetti, D.; Conchedda, G.; Holecz, F.; Blonda, P. 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 Sens.* **2010**, *48*, 1326–1354.

20. Baraldi, A.; Gironda, M.; Simonetti, D. Operational two-stage stratified topographic correction of spaceborne multi-spectral imagery employing an automatic spectral rule-based decision-tree preliminary classifier. *IEEE Trans. Geosci. Remote Sens*. **2010**, *48*, 112–146.

21. Baraldi, A.; Wassenaar, T.; Kay, S. Operational performance of an automatic preliminary spectral rule-based decision-tree classifier of spaceborne very high resolution optical images. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 3482–3502.

22. Baraldi, A. Fuzzification of a crisp near-real-time operational automatic spectral-rule-based decision-tree preliminary classifier of multisource multispectral remotely sensed images. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 2113–2134.

23. Baraldi, A. 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 Three-Stage Hybrid Remote Sensing Image Understanding System–Part I and Part II. In *Earth Observation*; Rustamov, R.; Salahova, S., Eds.; InTech Open Access Publisher: Rijeka, Croatia, 2012; pp. 63–98, 99–136.

24. Baraldi, A. Satellite Image Automatic Mapper™ (SIAM™)—A turnkey software button for automatic near-real-time multi-sensor multi-resolution spectral rule-based preliminary classification of spaceborne multi-spectral images. *Recent Pat. Space Technol.* **2011**, *1*, 81–106.

25. Matsuyama, T.; Hwang, S.-S.V. *SIGMA: A Knowledge-based Aerial Image Understanding System*; Plenum Press: New York, NY, USA, 1990.

26. Nagao, M.; Matsuyama, T. *A Structural Analysis of Complex Aerial Photographs*; Plenum Press: New York, NY, USA, 1980.

27. Shackelford, A.K.; Davis, C.H. A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1920–1932.

28. Shackelford, A.K.; Davis, C.H. A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas. *IEEE Trans. Geosci. Remote Sens*. **2003**, *41*, 2354–2363.

29. Shackelford, A.K. Development of Urban Area Geospatial Information Products from High Resolution Satellite Imagery Using Advanced Image Analysis Techniques. Ph.D. Dissertation, University of Missouri, Colombia, MO, USA, 2004.

30. Pakzad, K.; Bückner, J.; Growe, S. Knowledge based Moorland Interpretation Using a Hybrid System for Image Analysis. In *Proceedings of ISPRS WG III/ 2 & 3 Workshop "Automatic Objects from Digital Imagery"*, Munich, Germany, 8–10 September 1999.

31. Growe, S. Knowledge based interpretation of multisensor and multitemporal remote sensing images. *Int. Arch. Photogramm. Remote Sens*. **1999**, *32*, 130–138.

32. Cherkassky, V.; Mulier, F. *Learning from Data: Concepts, Theory, and Methods*; Wiley: New York, NY, USA, 1998.

33. Vecera, S.P.; Farah, M.J. Is visual image segmentation a bottom-up or an interactive process? *Percept. Psychophys*. **1997**, *59*, 1280–1296.

34. Hay, G.J.; Castilla, G. Object-based Image Analysis: Strengths, Weaknesses, Opportunities and Threats (SWOT). In *Proceedings of 1st International Conference on Object-based Image Analysis* (*OBIA*), Salzburg, Austria, 4–5 July 2006.

35. Hay, G.J.; Castilla, G. Geographic Object-Based Image Analysis (GEOBIA): A New Name for a New Discipline. 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, USA, 2008; Chapter 1.4, pp. 81–92.

36. Group on Earth Observations. *GEO Announces Free and Unrestricted Access to Full Landsat Archive: Universal Availability of Cost-Free Satellite Data and Images will Revolutionize The Use of Earth Observations for Decision-Making*. Available online: www.fabricadebani.ro/userfiles/GEO_press_release.doc (accessed on 9 September 2012).

37. Sart, F.; Inglada, J.; Landry, R.; Pultz, T. Risk Management Using Remote Sensing Data: Moving from Scientific to Operational Applications. In *Proceedings of SBSR Workshop*, Natal, Brazil, 23–27 April 2001.

38. GEO. *GEOSS Strategic Targets*. Available online: http://www.earthobservations.org/documents/geo_vi/12_GEOSS%20Strategic%20Targets%20Rev1.pdf (accessed on 9 September 2012).

39. GEO. *The Global Earth Observation System of Systems* (*GEOSS*) *10-Year Implementation Plan*. Available online: http://www.earthobservations.org/documents/10-Year%20Implementation%20Plan.pdf (accessed on 9 September 2012).

40. Sjahputera, O.; Davis, C.H.; Claywell, B.; Hudson, N.J.; Keller, J.M.; Vincent, M.G.; Li, Y.; Klaric, M.; Shyu, C.R. GeoCDX: An Automated Change Detection and Exploitation System for High Resolution Satellite Imagery. In *Proceedings of IEEE 2008 International Geoscience and Remote Sensing Symposium (IGARSS)*, Boston, MA, USA, 6–11 July 2008; pp. 467–470.

41. ESA. *About GMES—Overview*. Available online: http://www.esa.int/esaLP/SEMRRI0DU8E_LPgmes_0.html (accessed on 9 September 2012).

42. GMES. *GMES Info*. Available online: http://www.gmes.info (accessed on 10 January 2012).

43. USGS. *Web-Enabled Landsat Data* (*WELD*) *Project*. Available online: http://landsat.usgs.gov/WELD.php (accessed on 9 September 2012).

44. D'Elia, S. Personal communication. 2012.

45. Schaepman-Strub, G.; Schaepman, M.E.; Painter, T.H.; Dangel, S.; Martonchik, J.V. Reflectance quantities in optical remote sensing-definitions and case studies. *Remote Sens. Environ.* **2006**, *103*, 27–42.

46. Herold, M.; Woodcock, C.; Di Gregorio, A.; Mayaux, P.; Belward, A.S.; Latham, J.; Schmullius, C. A joint initiative for harmonization and validation of land cover datasets. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1719–1727.

47. De Lima, M.V.N.; Bielski, C.; Nowak, C. IMAGE2006: A Component of the GMES Precursor Fast Track Service on Land Monitoring. In *Proceedings of IEEE 2007 International Geoscience and Remote Sensing Symposium* (*IGARSS*), Barcelona, Spain, 23–28 July 2007; pp. 2669–2672.

48. Richter, R.; Schlapfer, D. *Atmospheric/Topographic Correction for Satellite Imagery—ATCOR-2/3 User Guide*, Version 8.0.2; August 2011. Available online: http://www.rese.ch/pdf/ atcor3_manual.pdf (accessed on 10 January 2012).

49. Richter, R.; Schlapfer, D. *Atmospheric/Topographic Correction for Airborne Imagery—ATCOR-4 User Guide*; Version 6.2 BETA; February 2012. Available online: http://www.dlr.de/eoc/ Portaldata/60/Resources/dokumente/5_tech_mod/atcor4_manual_2012.pdf (accessed on 10 January 2012).

50. Dorigo, W.; Richter, R.; Baret, F.; Bamler, R.; Wagner, W. Enhanced automated canopy characterization from hyperspectral data by a novel two step radiative transfer model inversion approach. *Remote Sens.* **2009**, *1*, 1139–1170.

51. Schlapfer, D.; Richter, R.; Hueni, A. Recent Developments in Operational Atmospheric and Radiometric Correction of Hyperspectral Imagery. In *Proceedings of 6th EARSeL SIG IS Workshop*, Tel Aviv, Isreal, 16–18 March 2009. Available online: http://www.earsel6th.tau.ac.il/~earsel6/ CD/PDF/earsel-OCEEDINGS/3054%20Schl%20pfer.pdf (accessed on 14 July 2012).

52. McCafferty, J.D. *Human and Machine Vision*, *Computing Perceptual Organization*; Ellis Horwood Limited: Chichester, UK, 1990.

53. Iqbal; Q.; Aggarwal, J.K. Image Retrieval via Isotropic and Anisotropic Mappings. In *Proceedings of IAPR Workshop on Pattern Recognition in Information Systems*, Setubal, Portugal, 6–8 July 2001; pp. 34–49.

54. Zamperoni, P. Plus ça va, moins ça va. *Pattern Recogn. Lett.* **1996**, *17*, 671–677.

55. Diamant, E. Machine Learning: When and Where the Horses Went Astray? In *Machine Learning*; Zhang, Y., Ed.; InTech Open Access Publisher: Rijeka, Croatia, 2010; pp. 1–18.

56. Liang, S. *Quantitative Remote Sensing of Land Surfaces*; John Wiley & Sons: Hoboken, NJ, USA, 2004.

57. Cootes, T.F.; Taylor, C.J. *Statistical Models of Appearance for Computer Vision*, *Imaging Science and Biomedical Engineering*; University of Manchester: Manchester, UK, 2004.

58. Chavez, P.S. An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sens. Environ.* **1988**, *24*, 459–479.

59. Stehman, S.V.; Czaplewski, R.L. Design and analysis for thematic map accuracy assessment: Fundamental principles. *Remote Sens. Environ.* **1998**, *64*, 331–344.

60. Hunt, N.; Tyrrell, S. *Stratified Sampling*. Available online: http://www.coventry.ac.uk/ec/~nhunt/ meths/strati.html (accessed on 10 January 2012).

61. Mason, C.; Kandel, E.R. Central Visual Pathways. In *Principles of Neural Science*; Kandel, E., Schwartz, J., Eds.; Appleton and Lange: Norwalk, CT, USA, 1991; pp. 420–439.

62. Gouras, P. Color Vision. In *Principles of Neural Science Principles of Neural Science*; Kandel, E., Schwartz, J., Eds.; Appleton and Lange: Norwalk, CT, USA, 1991, pp. 467–479.

63. Kandel, E.R. Perception of Motion, Depth and Form. In *Principles of Neural Science*; Kandel, E., Schwartz, J., Eds.; Appleton and Lange: Norwalk, CT, USA, 1991; pp. 441–466.

64. Wilson, H.R.; Bergen, J.R. A four mechanism model for threshold spatial vision. *Vision Res.* **1979**, *19*, 19–32.

65. Goodchild, M.F.; Yuan, M.; Cova, T.J. Towards a general theory of geographic representation in GIS. *Int. J. Geogr. Inf. Sci.* **2007**, *21*, 239–260.

66. Laurini, R.; Thompson, D. *Fundamentals of Spatial Information Systems*; Academic Press: London, UK, 1992.

67. Lang, S. Object-based Image Analysis for Remote Sesning Applications: Modeling Reality-Dealing with Complexity. 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, USA, 2008; Chapter 1.1, pp. 3–27.

68. Lüscher, P.; Burghardt, D.; Weibel, R. Ontology-Driven Enrichment of Spatial Databases. In *Proceedings of 10th ICA Workshop on Generalisation and Multiple Representation*, Moscow, Russia, 2–3 August 2007.

69. Bertero, M.; Poggio, T.; Torre, V. Ill-posed problems in early vision. *Proc. IEEE* **1988**, *76*, 869–889.

70. Hadamard, J. Sur les problemes aux derivees partielles et leur signification physique. *Princet. Univ. Bull.* **1902**, *13*, 49–52.

71. Bishop, C.M. *Neural Networks for Pattern Recognition*; Clarendon Press: Oxford, UK, 1995.

72. GLCF. *Global Forest Cover Change* (*GFCC*) *Project*. Available online: http://landcover.org/research/portal/gfcc (accessed on 9 September 2012).

73. Bruzzone, L.; Carlin, L. A multilevel context-based system for classification of very high spatial resolution images. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2587–2600.

74. Bruzzone, L.; Persello, C. A novel context-sensitive semisupervised SVM classifier robust to mislabeled training samples. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 2142–2154.

75. Salford Systems. *CART® Classification and Regression Trees*. Available online: http://www.salford-systems.com/en/products/cart (accessed on 9 September 2012).

76. RuleQuest Research Pty Ltd. *Data Mining Tools See5 and C5.0*. Available online: http://www.rulequest.com/see5-info.html (accessed on 9 September 2012).

77. Hansen, M.C.; Roy, D.P.; Lindquist, E.; Adusei, B.; Justice, C.O.; Altstatt, A. A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo Basin. *Remote Sens. Environ.* **2008**, *112*, 2495–2513.

78. Lindeberg, T. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention. *Int. J. Comput. Vis.* **1993**, *11*, 283–318.

79. Carson, C.; Belongie, S.; Greenspan, H.; Malik, J. Region-Based Image Querying. In *Proceedings of IEEE Workshop on Content-Based Access of Image and Video Libraries*, San Juan, Puerto Rico, 20 June 1997; pp. 42–49.

80. Yang J.; Wang, R.S. Classified road detection from satellite images based on perceptual organization. *Int. J. Remote Sens.* **2007**, *28*, 4651–4669.

81. Ruiz, L.A.; Recio, J.A.; Fernández-Sarría, A.; Hermosilla, T. A feature extraction software tool for agricultural object-based image analysis. *Comput. Electron. Agric.* **2011**, *76*, 284–296.

82. Corcoran, P.; Winstanley, A. Using Texture to Tackle the Problem of Scale in Landcover Classification. 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, USA, 2008; pp. 113–132.

83. Petrou, M.; Sevilla, P. *Image Processing: Dealing with Texture*; John Wiley & Sons: Chichester, UK, 2006.

84. Burr, D.C.; Morrone, M.C. A Nonlinear Model of Feature Detection. In *Nonlinear Vision*: *Determination of Neural Receptive Fields*, *Functions*, *and Networks*; Pinter, R.B., Bahram, N., Eds.; CRC Press: Boca Raton, FL, USA, 1992; pp. 309–327.

85. Computer Vision Lab. *Segmentation Parameter Tuner* (*SPT*). Available online: http://www.lvc.ele.puc-rio.br/wp/?p=904 (accessed on 3 July 2011).

86. Castilla, G.; Hay, G.J.; Ruiz-Gallardo, J.R. Size-constrained region merging (SCRM): An automated delineation tool for assisted photointerpretation. *Photogramm. Eng. Remote Sensing* **2008**, *74*, 409–429.

87. Page-Jones, M. *The Practical Guide to Structured Systems Design*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1988.

88. Pekkarinen, A.; Reithmaier, L.; Strobl, P. Pan-european forest/non-forest mapping with Landsat ETM+ and CORINE Land Cover 2000 data. *ISPRS J. Photogramm.* **2009**, *64*, 171–183.

89. Mather, P. *Computer Processing of Remotely-Sensed Images—An Introduction*; John Wiley & Sons: Chichester, UK, 1994.

90. Tapsall, B.; Milenov, P.; Tasdemir, K. Analysis of RapidEye Imagery for Annual Land Cover Mapping as an Aid to European Union (EU) Common Agricultural Policy. In *Proceedings of ISPRS TC VII Symposium: 100 Years ISPRS*, Vienna, Austria, 5–7 July 2010; Volume XXXVIII, Part 7B, pp. 568–573.

91. Lucas, R.; Medcalf, K.; Brown, A.; Bunting P.; Breyer J.; Clewley, D.; Keyworth, S.; Blackmore, P. Updating the Phase 1 habitat map of Wales, UK, using satellite sensor data. *ISPRS J. Photogramm.* **2011**, *66*, 81–102.

92. Crocetto, N.; Tarantino, E. A class-oriented strategy for features extraction from multidate ASTER imagery. *Remote Sens.* **2009**, *1*, 1171–1189.

93. Novack, T.; Esch, T.; Kux, H.; Stilla, U. Machine learning comparison between WorldView-2 and QuickBird-2-simulated imagery regarding object-based urban land cover classification. *Remote Sens.* **2011**, *3*, 2263–2282.

94. Baraldi, A.; Parmiggiani, F. Combined detection of intensity and chromatic contours in color images. *Opt. Eng.* **1996**, *35*, 1413–1439.

95. Pesaresi, M.; Benediktsson, J.A. A new approach for the morphological segmentation of high-resolution satellite imagery. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 309–320.