5

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

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

1. Introduction

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

For publication reasons this work consists of two companion papers, Part I and Part II respectively. In Part I related papers, concepts and definitions are revised from existing literature to provide this work with a significant survey value and make it self-contained. The survey of past works is completed in Part II Section 2, where differences at the architectural level between different families of existing RS-IUSs, namely, multi-agent hybrid RS-IUSs, two-stage segment-based RS-IUSs and two-stage stratified hierarchical hybrid RS-IUSs, are highlighted.

The original contribution of Part II is to propose novel definitions of objective continuous sub-symbolic sensory data, continuous physical information, subjective discrete semi-symbolic data structure, discrete semantic-square (semantic^2) information (which is naturally generated from the simultaneous combination of three components: (I) an objective continuous sensory data set, (II) an external subjective supervisor (observer) and (III) his/her own subjective prior ontology equivalent to a model of the (3-D) world existing before looking at the objective sensory data at hand) and prior knowledge base.

In practical contexts the aforementioned original definitions imply the following.

a. It is impossible to *extract* semantic[2] information from objective continuous sensory data because the latter, *per se*, are provided with no semantics at all.
b. It is possible to *correlate* discrete semantic[2] information to objective continuous sensory data. Unfortunately, correlation between continuous sensory data and a finite and discrete set of categorical variables, corresponding to independent random variables generating separable data structures (data aggregations, data clusters, data objects), is low in real-world RS image mapping problems at large data scale or fine semantic granularity, other than toy problems at small data scale and coarse semantic granularity.

Some practical conclusions of potential interest to the RS, computer vision (CV), artificial intelligence (AI) and machine learning (MAL) communities stem from these speculations. Firstly, in operational contexts (e.g., RS image classification problems at national/ continental/ global scale), other than toy problems (e.g., RS image mapping at coarse spatial resolution and local/regional scale), inductive classifiers capable of learning from a finite labeled data set are considered structurally inadequate to correlate (rather than extract, see this text above) discrete semantic[2] information with objective sensory data provided, *per se*, with no semantics at all.

Secondly, to increase the operational quality indicators (QIs) of existing two-stage hybrid RS-IUSs (namely, degree of automation, accuracy, efficiency, robustness to changes in input parameters, robustness to changes in the input data set, scalability, timeliness and economy), any first-stage inductive MAL-from-examples approach should be replaced by a deductive Machine Teaching (MAT)-by-rules capable of generating a preliminary classification first stage where small, but genuine image details are well preserved (Baraldi et al., 2006; Baraldi et al., 2010a; Baraldi et al., 2010b; Baraldi, 2011a).

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

As a proof of these concepts, the operational automatic multi-sensor multi-resolution near real-time Satellite Image Automatic Mapper™ (SIAM™), recently presented in RS literature[1] (Baraldi et al., 2006; Baraldi et al., 2010a; Baraldi et al., 2010b; Baraldi et al., 2010c; Baraldi, 2011a; Baraldi, 2011b), is adopted as first stage.

The rest of Part II of this work is organized as follows. Part II Section 3 discusses theoretical inconsistencies and algorithmic drawbacks found in Diamant's works (discussed in Part I Section 2.2 and Part I Section 2.5). Revised/novel definitions of objective continuous sensory data, continuous physical information, discrete semantic[2] information and prior knowledge are provided in Part II Section 4. In Part II Section 5 practical consequences of the novel definitions provided in Part II Section 4 are considered for CV, AI and MAL applications. Part II Section 6 presents the operational automatic multi-sensor multi-resolution near real-time SIAM™ as a proof of the original concepts proposed in this work. Conclusions are reported in Part II Section 7.

[1] SIAM™ - Patent pending - © Andrea Baraldi & University of Maryland.

2. Related works (continued): Taxonomy of hybrid RS-IUS architectures

As reported in Part I Section 2.1, there is a new trend of research and development in both CV (Cootes & Taylor, 2004) and RS literature (Matsuyama & Shang-Shouq Hwang, 1990; Shunlin Liang, 2004) to outperform existing scientific and commercial image understanding systems. This novel trend focuses on the development of quantitative hybrid models for retrieving sub-symbolic continuous variables (e.g., LAI) and symbolic categorical discrete variables (e.g., land cover composition) from multi-spectral (MS) imagery. By definition, hybrid models combine both statistical and physical models to take advantage of the unique features of each and overcome their shortcomings (see Part I Section 2.1). The study of hybrid quantitative models is also called AI systems integration. In this section, the taxonomy of hybrid RS-IUSs is summarized in line with (Baraldi et al., 2010a). It consists of:

- multi-agent hybrid RS-IUSs,
- two-stage segment-based RS-IUSs, whose conceptual foundation is well known in RS literature as as geographic (2-D) object-based image analysis (GEOBIA), including a so-called iterative geographic OO image analysis (GEOOIA) approach (Baatz et al., 2008). and
- two-stage stratified hierarchical hybrid RS-IUSs employing SIAM™ as preliminary classification first stage.

2.1 Multi-agent hybrid RS-IUSs

In existing literature multi-agent hybrid RS-IUSs provide application-specific combinations of inductive and deductive inference mechanisms (Matsuyama & Shang-Shouq Hwang, 1990). A traditional multi-agent hybrid RS-IUS architecture comprises the following modules (see Fig. 1).

1. (3-D) Scene domain knowledge, also called *world model* (Matsuyama & Shang-Shouq Hwang, 1990). It is represented as a semantic network consisting of classes of objects as nodes and relationships between classes as arcs between nodes (refer to Part I Section 2.2.2).
2. A Low-Level Vision Expert (LLVE, refer to Part I Section 2.4.1.2) (Matsuyama & Shang-Shouq Hwang, 1990). In general, an LLVE can be applied either image-wide or within a local image area specified by a Specialized Object Model Selection Expert (SOMSE, see this text below) (Mather, 1994). LLVE includes a battery of low-level sub-symbolic (non-semantic) general-purpose domain-independent inductive-learning (fine-to-coarse, bottom-up) driven-without-knowledge inherently ill-posed image processing algorithms called *image segmentation* for simplicity's sake (also refer to Part I Section 2.4.1.2) (Matsuyama & Shang-Shouq Hwang, 1990). As output, the image segmentation first stage provides image features, namely points and regions (segments, [2-D] objects, parcel or blobs (Carson et al., 1997; Lindeberg, 1993; Yang & Wang, 2007), see Part I Section 2.3) or, vice versa, region boundaries, i.e., edges, provided with no semantic meaning (see Part I Section 2.4.1.2).
3. A high-level interpretation second stage employing a combination of top-down (model-driven) and bottom-up (data-driven) inference mechanisms to establish the correspondence between sub-symbolic (2-D) image features extracted from the image domain and symbolic (3-D) object models stored in the world model to construct plausible structural (semantic) description(s) of the depicted scene (refer to Part I Section

2.3). The combination of top-down with bottom-up inference strategies achieves two operational advantages: (a) provides better conditions for an otherwise ill-posed driven-without-knowledge segmentation first stage (refer to Part I Section 2.3) and (b) allows restriction of intensive processing to a small portion of the image data (Matsuyama & Shang-Shouq Hwang, 1990), analogously to a focus of visual attention in pre-attentive biological vision (Mason & Kandel, 1991; Gouras, 1991; Kandel, 1991). The high-level processing second stage comprises (Matsuyama & Shang-Shouq Hwang, 1990): (I) a Spatial Reasoning Expert (SRE) whose aim is to trigger the instantiation, within a candidate local area, of plausible generic (3-D) object models found in the available world model, e.g., house, and (II) a SOMSE (refer to this text above) which uses domain-dependent knowledge about specific applications to: (i) prune the search space of specialized (3-D) object models (e.g., rectangular house, L-shaped house, etc.) linked by A-KIND-OF relations to the generic target (3-D) object model (e.g., house) provided by SRE; (ii) transform the 3-D appearance properties of the specialized (3-D) object model into a selected set of 2-D appearance properties based on the imaging sensor model; (iii) transform a target spatial relation in fuzzy terms (e.g., in front of) provided by SRE into a local area based on a trial-and-error heuristic search with no concrete theoretical basis and (iv) provide a consistency examination between quantitative absolute image features collected by LLVE in a local area and the target 2-D appearance constraints. In other words, the 2-D appearance properties must be satisfied by image features extracted by LLVE from a local area. Since the image structure in a local area is very simple compared with that of the entire image, image feature extraction performed by an object model-driven and locational constrained LLVE can be very efficient and reliable compared with that performed by the same LLVE run image-wide at the first stage (Matsuyama & Shang-Shouq Hwang, 1990) (p. 41).

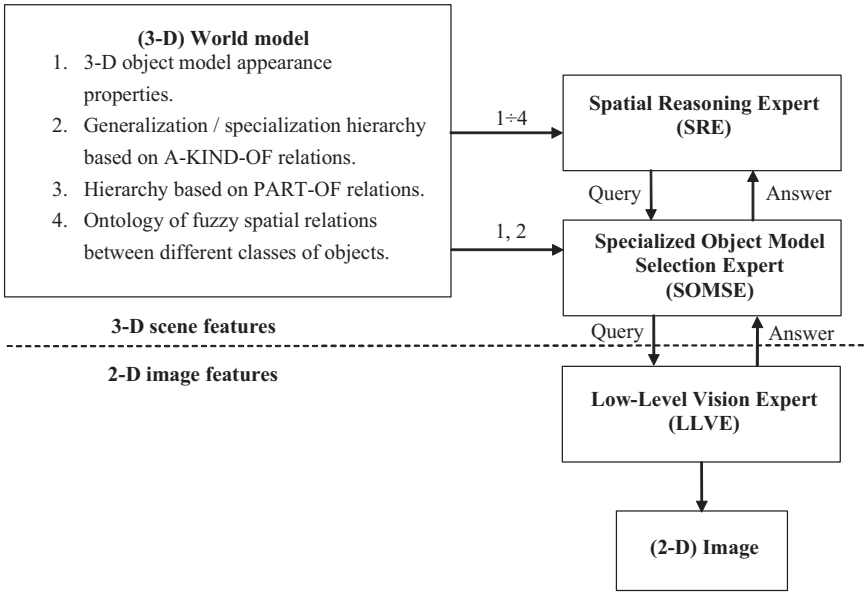

Fig. 1. Multi-agent hybrid systems for RS image understanding (derived from Figure 2.1 in (Matsuyama & Shang-Shouq Hwang, 1990), p. 36).

	Legenda. Y: Yes, N: No, C: Complete, I: Incomplete (radiometric calibration offset parameters are set to zero), (E)TM: (Enhanced) Thematic Mapper, B: Blue, G: Green, R: Red, NIR: Near Infra-Red, MIR: Medium IR, TIR: Thermal IR, SR: Spatial Resolution, Pan: Panchromatic. Blue columns: visible channels typical of water and haze. Green column: NIR band typical of vegetation. Brown columns: MIR channels characteristics of bare soils. Red column: TIR channel.											
SIAM™ system of systems		**B – (E)TM1, 0.45-0.52 (μm)**	**G – (E)TM2, 0.52-0.60 (μm)**	**R – (E)TM3, 0.63-0.69 (μm)**	**NIR – (E)TM4, 0.76-0.90 (μm)**	**MIR1 – (E)TM5, 1.55-1.75 (μm)**	**MIR2 – (E)TM7, 2.08-2.35 (μm)**	**TIR – (E)TM6, 10.4-12.5 (μm)**	**SR (m)**	**Rad. Cal. Y/N, C/I**	**Pan SR (m)**	**Notes**
L-SIAM™ (95/47/18 Sp. Cat.)	Landsat-4/-5 TM	×	×	×	×	×	×	×	30	Y-C		Refer to Table I in (Baraldi et al., 2006a).
	Landsat-7 ETM+	×	×	×	×	×	×	×	30	Y-C	15	Same as above.
	MODIS	×	×	×	×	×	×	×	250, 500, 1000	Y-C		Same as above.
	ASTER		×	×	×	×	×	×	15-30	Y-C		Same as above.
	CBERS-2B	×	×	×	×	×	×	×		N		
S-SIAM™ (68/40/15 Sp. Cat.)	SPOT-4 HRVIR		×	×	×	×			20	Y-I	10	Refer to Table II in (Baraldi et al., 2006a).
	SPOT-5 HRG		×	×	×	×			10	Y-I	2.5 - 5	Same as above.
	SPOT-4/-5 VMI		×	×	×	×			1100	Y-I		Same as above.
	IRS-1C/-1D LISS-III		×	×	×	×			23.5	Y-I		
	IRS-P6 LISS-III		×	×	×	×			23.5	Y-I		
	IRS-P6 AWiFS		×	×	×	×			56	Y-I		
AV-SIAM™ (82/42/16 Sp. Cat.)	NOAA AVHRR			×	×	×		×	1100	Y		Refer to Table II in (Baraldi et al., 2006a).
	MSG			×	×	×		×	3000	Y		Same as above.
AA-SIAM™ (82/42/16Sp. Cat.)	ENVISAT AATSR		×	×	×	×		×	1000	Y		Same as above.
	ERS-2 ATSR-2		×	×	×	×		×	1000	Y		
I-SIAM™ (52/28/12Sp. Cat.)	IKONOS-2	×	×	×	×				4	Y	1	
	QuickBird-2	×	×	×	×				2.4	Y	0.61	
	WorldView-2	×	×	×	×				2.0	Y	0.5	
	GeoEye-1	×	×	×	×				1.64	Y	0.41	
	OrbView-3	×	×	×	×				4	Y	1	
	RapidEye-1 to -5	×	×	×	×				6.5	Y-I		
	ALOS AVNIR-2	×	×	×	×				10	Y		
	KOMPSAT-2	×	×	×	×				4	N	1	
	TopSat	×	×	×	×				5	N	2.5	
	FORMOSAT-2	×	×	×	×				8	N	2	
D-SIAM™ (52/28/12Sp. Cat.)	Landsat-1/-2/-3/-4/-5 MSS		×	×	×				79	Y		
	IRS-P6 LISS-IV		×	×	×				5.8	Y-I		
	SPOT-1/-2/-3 HRV		×	×	×				20	Y-I	10	
	DMC		×	×	×				22-32	N		

Table 1. SIAM™ system of systems. List of spaceborne optical imaging sensors eligible for use as input.

Multi-agent hybrid systems typically suffer from two main limitations.

- In addition to the intrinsic insufficiency of image features, e.g., due to occlusion and dimensionality reduction (refer to Part I Section 2.3), these systems are affected by the so-called artificial insufficiency caused by the inherent ill-posedness of the image segmentation problem (Matsuyama & Shang-Shouq Hwang, 1990) (see Part I Section 2.4.1.2). This means that in RS common practice any first-stage image segmentation algorithm is simultaneously affected by both omission and commission segmentation errors. Although the inherent ill-posedness of image segmentation is acknowledged by a reasonable portion of existing literature (Burr & Morrone, 1992; Corcoran et al., 2010; Corcoran & Winstanley, 2007; Delves et al., 1992; Hay & Castilla, 2006; Matsuyama & Shang-Shouq Hwang, 1990; Petrou & Sevilla, 2006; Vecera & Farah, 1997), this is often forgotten by a large segment of the RS community where literally dozens of "novel" segmentation algorithms are published each year (Zamperoni, 1996) (refer to Part I Section 2.4.1.2).
- Semantic nets lack flexibility and scalability to cope with changes in sensor characteristics and users' changing needs, i.e., they are unsuitable for commercial RS image processing software toolboxes and remain limited to scientific applications.

To overcome these limitations, an alternative two-stage stratified hierarchical hybrid RS-IUS architecture, such as that shown in Fig. 3, was proposed in recent literature (Baraldi et al., 2006; Baraldi et al., 2010a; Baraldi et al., 2010b; Baraldi, 2011a; Baraldi, 2011b; Baraldi et al., 2010c).

2.2 Two-stage segment-based RS-IUSs

Two-stage segment-based RS-IUSs comprise an inductive driven-without-knowledge image segmentation first stage and a second-stage object-based classifier, see Fig. 2. The latter can be implemented based on deductive or inductive inference mechanisms, say, as a prior knowledge-based non-adaptive decision-tree or a supervised data learning classifier (e.g., a Support Vector Machine, SVM (Bruzzone & Carlin, 2006)).

Due to the availability of a commercial GEOBIA software developed by a German company (Definiens Imaging GmbH, 2004; Esch et al., 2008), two-stage segment-based RS-IUSs have recently gained widespread popularity and are currently considered the state-of-the-art in both scientific and commercial RS image mapping application domains (Mather, 1994; Pekkarinen, Reithmaier & Strobl, 2009). In practice, under the guise of 'flexibility' current commercial 2-D object-based software provides overly complicated options to choose from (Hay & Castilla, 2006). This means that with their increasing diffusion commercial two-stage segment-based RS-IUSs show an increasing lack of productivity (Tapsall et al., 2010), consensus and research (Castilla et al., 2008; Hay & Castilla, 2006) (refer to Part I Section 2.4.1.2).

2.3 Two-stage stratified hierarchical hybrid RS-IUS employing SIAM™ as its preliminary classification first stage

Accounting for the customary distinction between a model and the algorithm used to identify it (Baraldi et al., 2010a; Baraldi, 2011a), an original two-stage stratified hierarchical hybrid RS-IUS architecture (see Fig. 3) was identified starting from several RS-IUS

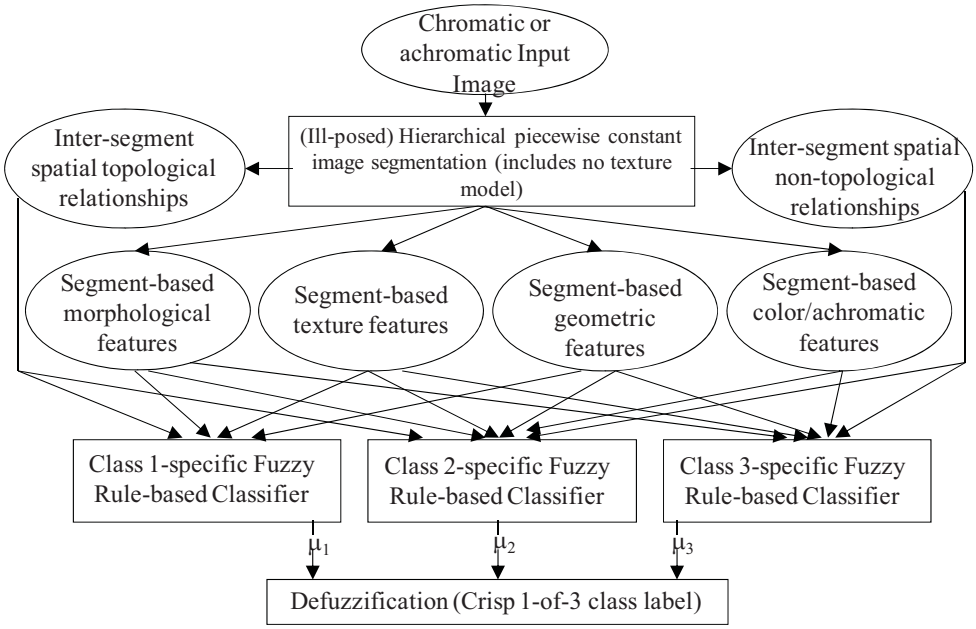

Fig. 2. Two-stage segment-based hybrid RS-IUS architecture adopted, for example, by the eCognition commercial software toolbox (Definiens Imaging GmbH, 2004). Preliminary image simplification is pursued by means of an (ill-posed hierarchical) image segmentation approach which generates as output a segmented (discrete) map, either single-scale or multi-scale. Worthy of note is that first-stage output sub-symbolic informational primitives, namely, labeled segments (2-D objects, parcels), e.g., segment 1, segment 2, etc., are provided with no semantic meaning.

implementations proposed by Shackelford and Davis in recent years (Shackelford & Davis, 2003a; Shackelford & Davis, 2003b). This novel RS-IUS architecture comprises the following phases (Baraldi et al., 2006; Baraldi et al., 2010a; Baraldi et al., 2010b; Baraldi et al., 2010c; Baraldi, 2011a; Baraldi, 2011b).

a. A radiometric calibration pre-processing stage, where DNs are transformed into top-of-atmosphere reflectance (TOARF) or surface reflectance (SURF) values, with TOARF $\supseteq$ SURF, the latter being an ideal (atmospheric noise-free) case of the former. This radiometric calibration constraint not only ensures the harmonization and interoperability of multi-source observational data in line with the Quality Assurance Framework for EO (QA4EO) guidelines (GEO/CEOSS, 2008), but is considered a necessary, although not sufficient, condition for input Earth observation (EO) imagery to be automatically interpreted (see Part I Section 2.7.1). It is worth mentioning that a RS-IUS suitable for mapping TOARF values into surface categories makes the inherently ill-posed (therefore, difficult to solve) atmospheric correction problem an optional MS image pre-processing stage unlike competing classification approaches employing surface reflectance spectra, such as the ERDAS ATCOR3 (Richter, 2006) (see Part I Section 2.7.1).

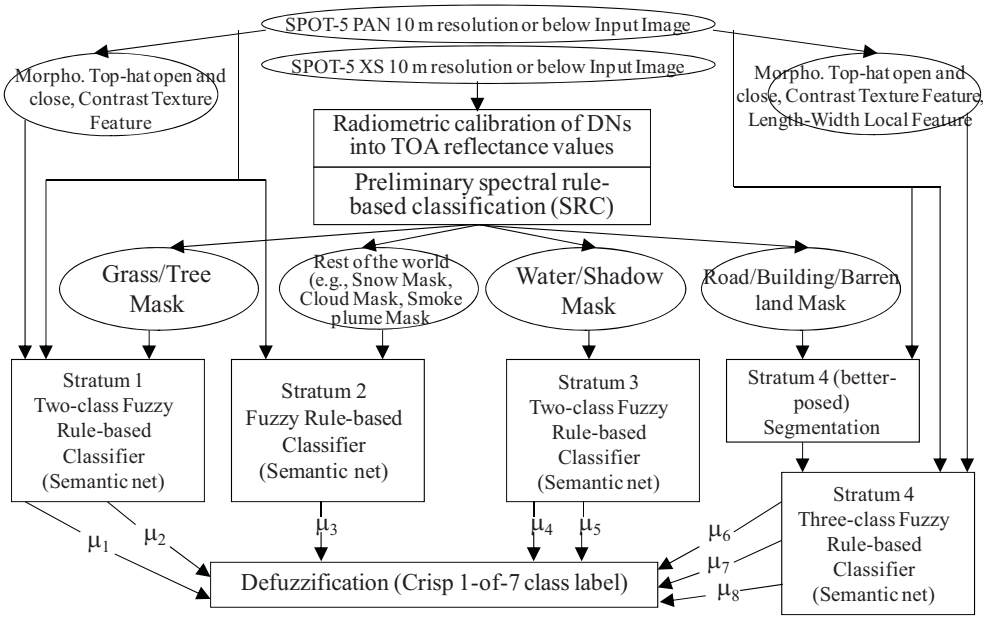

Fig. 3. Novel hybrid two-stage stratified hierarchical RS-IUS architecture. This data flow diagram (DFD) shows processing blocks as rectangles and sensor derived data products as circles. In this example, a SPOT-5 MS image is adopted as input. The panchromatic (PAN) image can be generated from the MS image. The MS image is input to the preliminary classification first stage and, if useful, to second-stage class-specific classification modules. The PAN image is exclusively employed as input to second-stage stratified class-specific context-sensitive classification modules, where color information is dealt with by stratification. For example, stratified texture detection is computed in the PAN image domain, which reduces computation time.

b. A first-stage application-independent per-pixel (non-contextual) top-down (prior knowledge-based, see Part I Section 2.1) preliminary classifier in the Marr sense (Marr, 1982).
c. A second-stage battery of stratified hierarchical context-sensitive application-dependent modules for class-specific feature extraction and classification.

In (Baraldi et al., 2006; Baraldi et al., 2010a; Baraldi et al., 2010b; Baraldi et al., 2010c; Baraldi, 2011a; Baraldi, 2011b), the abovementioned first-stage pixel-based preliminary classifier was designed and implemented as an original operational automatic near-real-time per-pixel multi-source multi-resolution application-independent SIAM™. To employ as input a radiometrically calibrated MS image acquired by almost any of the ongoing or future planned satellite optical missions, SIAM™ is designed as an integrated system of systems. It comprises a "master" 7-band Landsat-like SIAM™ (L-SIAM™) together with five down-scaled ("slave", derived) versions of L-SIAM™ whose input is a MS image featuring a spectral resolution that overlaps with, but is inferior to, Landsat's. To summarize, SIAM™ combines six sub-systems (refer to Table 1).

i. A "master" 7-band L-SIAM™ capable of detecting 95/ 47/ 18 spectral categories at fine/ intermediate/ coarse semantic granularity (see Fig. 4). The legend of the preliminary classification map generated by L-SIAM™ at fine semantic granularity is shown in Table 2.
ii. A four-band Satellite Pour l'Observation de la Terre (SPOT)-like SIAM™ (S-SIAM™), which detects 68/ 40/ 15 spectral categories at fine/ intermediate/ coarse semantic granularity (see Fig. 5).
iii. A four-band National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR)-like SIAM™ (AV-SIAM™), which detects 82/ 42/ 16 spectral categories at fine/ intermediate/ coarse semantic granularity.
iv. A five-band ENVISAT Advanced Along-Track Scanning Radiometer (AATSR)-like SIAM™ (AA-SIAM™), which detects 82/ 42/ 16 spectral categories at fine/ intermediate/ coarse semantic granularity.
v. A four-band IKONOS-like SIAM™ (I-SIAM™), which detects 52/ 28/ 12 spectral categories at fine/ intermediate/ coarse semantic granularity (see Fig. 6). The legend of the preliminary classification map generated by I-SIAM™ at fine semantic granularity is shown in Table 3.
vi. A three-band Disaster Monitoring Constellation (DMC)-like SIAM™ (D-SIAM™), which detects 52/28/12 spectral categories at fine/intermediate/coarse semantic granularity.

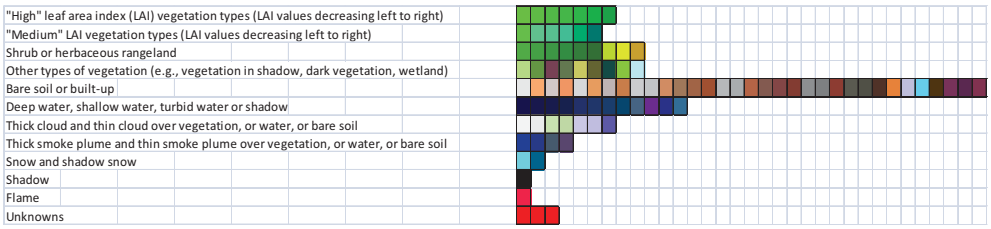

Table 2. Preliminary classification map legend adopted by L-SIAM™ at fine semantic granularity. Pseudo-colors of the 95 spectral categories are gathered based on their spectral end member (e.g., bare soil or built-up) or parent spectral category (e.g., "high" LAI vegetation types). The pseudo-color of a spectral category is chosen as to mimic natural colors of pixels belonging to that spectral category.

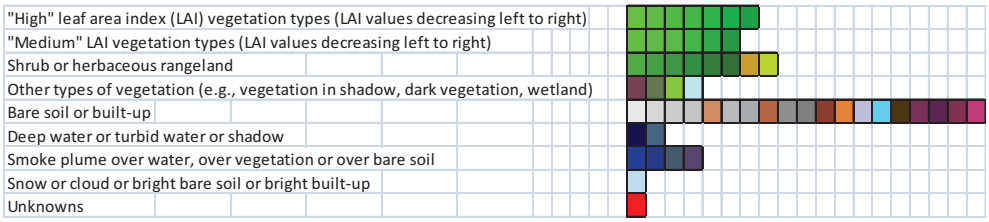

Table 3. Preliminary classification map legend adopted by I-SIAM™ at fine semantic granularity. Pseudo-colors of the 52 spectral categories are gathered based on their spectral end member (e.g., bare soil or built-up) or parent spectral category (e.g., "high" LAI vegetation types). The pseudo-color of a spectral category is chosen as to mimic natural colors of pixels belonging to that spectral category.

Fig. 4 to Fig. 6 show qualitatively that, in disagreement with a common opinion in the RS community where GEOBIA is considered indispensable for spaceborne VHR image understanding (Bruzzone & Carlin, 2006; Bruzzone & Persello, 2009; Persello & Bruzzone, 2010), the pixel-based SIAM™ is very successful in the automatic mapping of RS imagery, including VHR images (Baraldi et al., 2006; Baraldi et al., 2010a; Baraldi et al., 2010b; Baraldi et al., 2010c; Baraldi, 2011a; Baraldi, 2011b). This means that SIAM™ is not affected by the well-known salt-and-pepper classification noise effect which traditionally affects ordinary pixel-based classifiers (e.g., maximum-likelihood classifiers (Cherkassky and Mulier, 2006)), which is tantamount to saying that SIAM™ is successful in modeling the within-spectral-category variance.

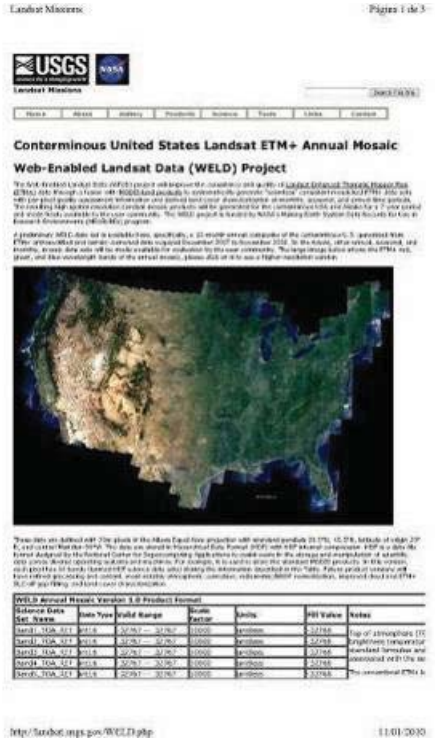

Fig. 4(a). Web-Enabled Landsat Data (WELD) Project (USGS & NASA, 2011). This is a joint NASA and USGS project providing seamless consistent mosaics of fused Landsat-7 Enhanced TM Plus (ETM+) and MODIS data radiometrically calibrated into top-of-atmosphere reflectance (TOARF) and surface reflectance. These mosaics are made freely available to the user community. Each consists of 663 fixed location tiles. Spatial resolution: 30 m. Area coverage: Continental USA and Alaska. Period coverage: 7-year. Product time coverage: weekly, monthly, seasonal and annual composites.

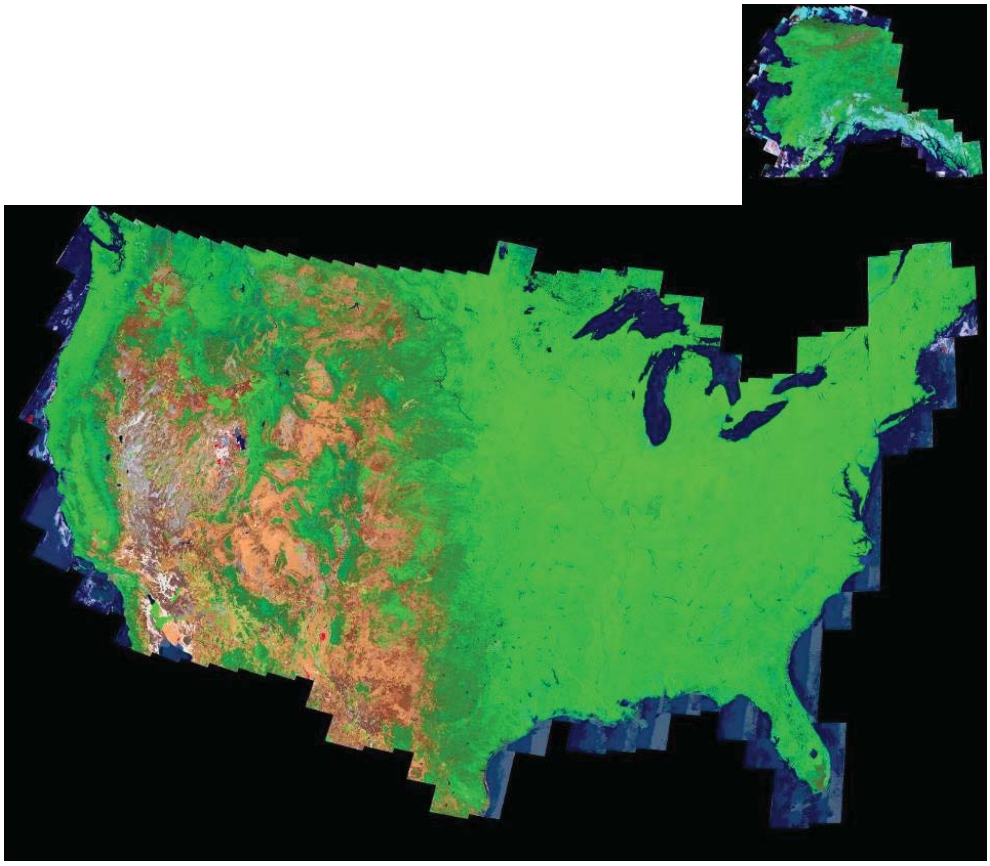

Fig. 4(b). Including the map of Alaska at the top right. Preliminary classification map automatically generated by L-SIAM™ from the 2008 annual WELD mosaic shown in Fig. 4(a). Output spectral categories are depicted in pseudo colors. Map legend: refer to Table 2. To generate this map at national scale L-SIAM™ was run overnight by L. Boschetti (Univ. of Maryland) in Dec. 2010. To the best of this author's knowledge, this is the first example of such a high-level product automatically generated at both the NASA and USGS.

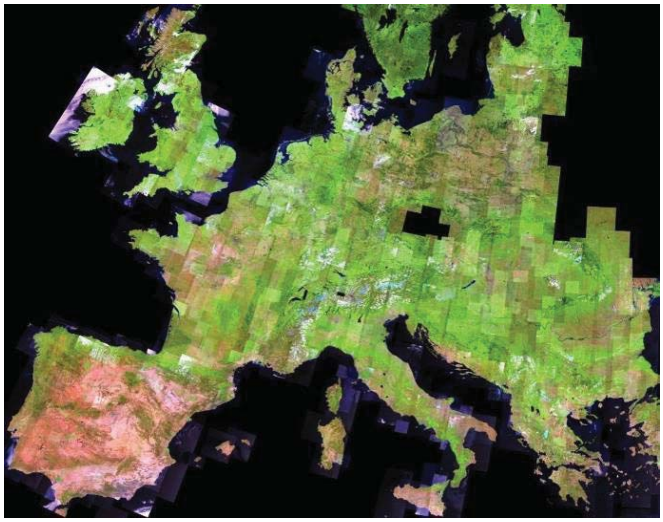

Fig. 5(a). 4-band GMES-IMAGE2006 Coverage 1 mosaic, consisting of approximately two thousand 4-band IRS-P6 LISS-III, SPOT-4, and SPOT-5 images, mostly acquired during the year 2006, depicted in false colors: Red – Band 4 (Short Wave InfraRed, SWIR), Green – Band 3 (Near IR, NIR), Blue – Band 1 (Visible Green). Down-scaled spatial resolution: 25 m.

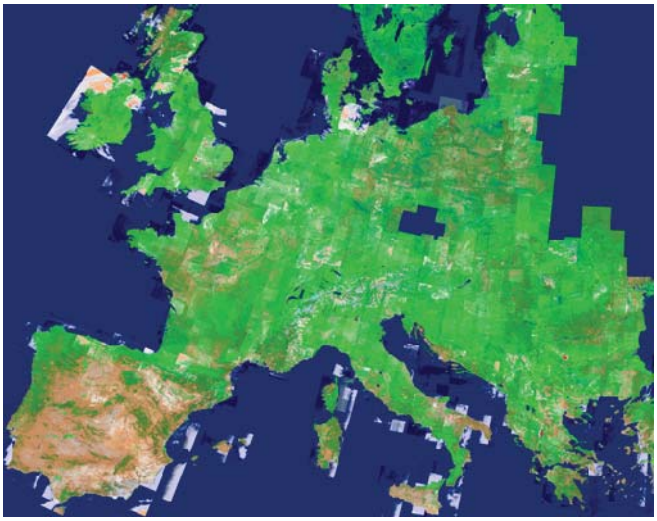

Fig. 5(b). Preliminary classification map automatically generated by S-SIAM™ from the mosaic shown in Fig. 5(a). Output spectral categories are depicted in pseudo colors. A map legend similar to Table **2** is adopted: water and shadow areas are in blue, clouds in white, snow and ice in light blue, vegetation types in different shades of green, rangeland types in different shades of light green, barren land types in different shades of brown and grey. To the best of this author's knowledge, this is the first example of such a high-level product automatically generated at the European Commission – Joint Research Center (EC-JRC).

Fig. 6(a). QuickBird-2 image, 2.4 m spatial resolution, acquisition date 2010-03-16, radiometrically calibrated into TOARF values, depicted in false colors (R: 3, G: 4, B: 1). Default image histogram stretching: ENVI linear stretching 2%.

Fig. 6(b). Automatic Q-SIAM™ preliminary mapping of the QB-2 image shown in Fig. 6(a). Spectral categories are depicted in pseudo colors. Map legend: see Table 3. It is noteworthy that, within the Q-SIAM™ mutually exclusive and completely exhaustive classification scheme, cloud detection is *per se* an interesting operational product with relevant commercial applications and, to the best of these authors' knowledge, without alternative solutions in either commercial or scientific RS-IUSs.

Fig. 7(a). Zoomed area of a Landsat 7 ETM+ image of Virginia, USA (path: 16, row: 34, acquisition date: 2002-09-13), depicted in false colors (R: band ETM5, G: band ETM4, B: band ETM1), 30 m resolution, calibrated into TOARF values.

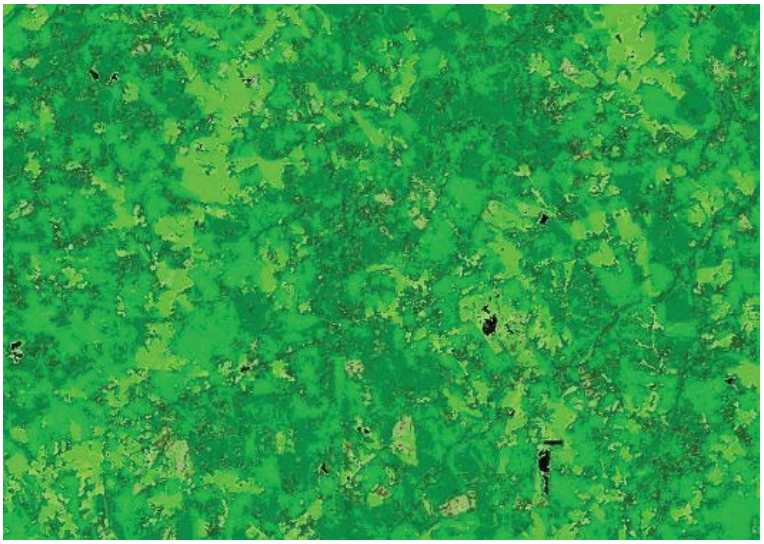

Fig. 7(b). 2nd-stage stratified vegetated land cover classification map generated in series with the L-SIAM™ first stage from Fig. 7(a). This 2nd-stage map consists of 19 vegetated/non-vegetated land cover classes, depicted in pseudo-colors, including: crop field or grassland, broad-leaf forest, needle-leaf forest and non-vegetated pixels (in black). Input features are: spectral layers generated by L-SIAM™, (achromatic) brightness and multi-scale isotropic texture features extracted from the brightness image.

To the best of this author's knowledge no unifying automatic multi-sensor multi-resolution near real-time RS image classification platform alternative to SIAM™ can be found in existing literature. This is tantamount to saying that SIAM™ provides the first operational example of an automatic multi-sensor multi-resolution near real-time EO system of systems envisaged under on-going international research programs such as the Global EO System of Systems (GEOSS) conceived by the Group on Earth Observations (GEO) (GEO, 2005; GEO, 2008a) and the Global Monitoring for the Environment and Security (GMES), which is an initiative led by the European Union (EU) in partnership with the European Space Agency (ESA) (ESA, 2008; GMES, 2011) (see Part I Section 1).

Fig. 7 shows an example of an automatic 2nd-stage stratified rule-based vegetated land cover classification system in series with the L-SIAM™ first stage. The two-stage automatic classifier employing L-SIAM™ as preliminary classification first stage (refer to Fig. 3) is input with a 7-band Landsat image radiometrically calibrated into TOARF values, shown in Fig. 7(a). The 2nd-stage stratified rule-based vegetated land cover classification system in series with the L-SIAM™ first stage employs as input features: spectral-based layers (strata, generated by L-SIAM™ at first stage), (achromatic) brightness and multi-scale isotropic texture extracted from the brightness image. The 2nd-stage classifier provides as output a classification map consisting of 19 vegetated/non-vegetated land cover classes, depicted in pseudo-colors, including: crop field or grassland, broad-leaf forest, needle-leaf forest and non-vegetated pixels (in black), see Fig. 7(b).

3. Inconsistencies and limitations of the Diamant computational theory and algorithms

An original analysis of the Diamant definitions reported in Part I Section 2.2.3 and Diamant's image segmentation and contour detection algorithms summarized in Part I Section 2.5 is provided below.

3.1 Comments on the Diamant definitions of data, information and knowledge

According to this author, the Diamant definitions reported in Part I Section 2.2.3 are affected by three major drawbacks.

i. Diamant states that "information elicitation (extraction) does not require incorporation of any high-level knowledge" (Diamant, 2010a; Diamant, 2010b), which is tantamount to saying that detection of non-semantic primary data structures (data objects), e.g., (2-D) image segments, in an unlabeled data set, e.g., a (2-D) image, does not require incorporation of any high-level (prior) knowledge. Based on this statement it is possible to conclude that despite his theoretical anti-conformism, namely, his willingness to replace the MAL-from-examples paradigm with the MAT-by-rules approach, Diamant is a conformist in practice. In fact, the Diamant image contour detection and image segmentation algorithms (see Part I Section 2.5) fit existing CV system architectures well established in literature, such as, respectively, the Marr CV system architecture, conceived in the 1980s and comprising a zero-crossings (contour detection) primal sketch, and RS-IUSs where an image segmentation first stage is adopted in agreement with the GEOBIA approach (see Part I Section 2.4.1.2). In other words, there is a clear contradiction in terms between the Diamant claim of replacing the MAL-from-examples

with a MAT-by-rules paradigm and his practical proofs of concept, consisting of image segmentation and contour detection algorithms 100% consistent with the same MAL-from-examples paradigm he intends to overcome.

ii. If the Diamant CV system coincides with a Marr CV system or an GEOBIA approach (refer to paragraph (i) above), then, in practical contexts, its operational QIs (see Part I Section 2.8) are expected to score as low as Marr's or OBIA's (refer to Part I Section 1, Part I Section 2.4.1.2 and Part II Section 2). At the level of understanding of an information processing system known as computational theory (system architecture, see Part I Section 2.6), GEOBIA scores low in operational contexts because, according to the present author, it goes symbolic as late as possible, namely, at the output of its second and last stage (see Fig. 2). This is in contrast with an important intuition by Marr stating that "vision goes symbolic almost immediately, right at the level of zero-crossings (first-stage primal sketch)… without loss of information" (Marr, 1982) (p. 343) (see Part I Section 2.3).

iii. To recover from the gap existing between Diamant's theoretical anti-conformism, but practical conformism (refer to paragraphs (i) and (ii) above), it is sufficient to observe that statements such as "information elicitation (aggregation) does not require incorporation of any high-level knowledge" (Diamant, 2010a; Diamant, 2010b), are in clear contradiction with a relevant section of existing literature (see Part I Section 2.4.1.2). In particular, Diamant considers primary data structures, equivalent to non-semantic data objects (e.g., image segments), as "natural data structures which reflect some similarities among neighboring elements in the data. Therefore, defining them is certainly a well-grounded procedure that does not raise any objection, because objective (physical) laws underpin such a procedure" (Diamant, 2010a) (see Part I Section 2.2.3.2). In other words, "physical information, being a natural property of the data, can be extracted instantly from the data, and any special rules for such task accomplishment are not needed" (Diamant, 2010a). Unfortunately, no well-grounded (well-posed) inductive learning-from-unlabeled-data approach exists (see Part I Section 2.1). For example, both unlabeled data clustering and (2-D) image segmentation algorithms are inherently ill-posed (see Part I Section 2.4.1). By adopting the Diamant terminology it is possible to state that detection of "discernable" data structures is not at all a physical problem of objective nature: it is rather a typical semantic problem of a qualitative (subjective) nature, where prior knowledge (provided by an external supervisor) must come into play to make the inherently ill-posed inductive learning-from-data problem better posed, although subjective (see Part I Section 2.1). This is tantamount to saying that the conceptual foundation of GEOBIA, i.e., the relationship between inherently ill-posed sub-symbolic (2-D) image segments and symbolic (3-D) landscape objects, remains affected by a lack of general consensus and research (Hay & Castilla, 2006) (see Part I Section 2.4.1.2).

To conclude, Diamant appears to have totally misunderstood one of two facts about the MAL-from-examples paradigm. These two facts hold true for MAL from unlabeled data and MAL from labeled data algorithms, respectively, as described below.

a. MAL from unlabeled (unsupervised) data (see Part I Section 2.1 and Part I Section 2.4.1). Any machine learning from unlabeled data approach (e.g., unlabeled data clustering, image segmentation) is inherently ill-posed and requires prior knowledge to become better posed. It means that any attempt to extract non-semantic primary data structures (data objects), e.g., image segments and unlabeled data clusters, from an unlabeled data set (e.g., an image) without incorporation of high-level knowledge provided by an

external supervisor is a fatal misconception, committed by Diamant himself, stemming from the fallacies (inherent ill-posedness) of the MAL-from-examples paradigm.

b. MAL from labeled (supervised) data (see Part I Section 2.1 and Part I Section 2.4.2). It is true that, in Diamant's words, "knowledge about the rules that underpin (semantic) secondary (data) structures formation (from primary data structures considered as non-semantic and driven-without-knowledge) is a property of human observers (or their artificial counterparts) and not an inherent property of the data... (therefore) attempts to extract semantics from data are a fatal misconception stemming from the fallacies of the data-processing paradigm..." (Diamant, 2010a). This quote implies that no semantic information can be extracted from objective sensory data, but a correlation function can be established between semantic concepts and objective data for toy data understanding problems exclusively (refer to Part I Section 1 and Part I Section 2.1).

3.2 Comments on the Diamant image segmentation algorithm

In practical terms, the image segmentation algorithm proposed by Diamant can be subjected to the following criticisms.

- Not enough information is provided for the implementation to be reproduced. In practice the Diamant image segmentation algorithm cannot be duplicated and, therefore, cannot be tested by others.
- Diamant does not provide his image segmentation algorithm with QIs such as those listed in Part I Section 2.8. For example, based on Diamant's paper it is impossible to assess the following operational QIs.
  - Degree of automation. The following questions remain unanswered. What is the number of the image segmentation-free parameters to be user-defined? Have these user-defined parameters a physical meaning? What is their range of change?
  - Robustness to changes in input parameters to be user-defined.
  - Robustness to changes in the input data set acquired across time, space and sensors. In his paper (Diamant, 2005) Diamant applies his image segmentation algorithm to a single toy problem whose input data set consists of a panchromatic image 640×480 pixels in size. What about color images? What about satellite imagery? What about synthetic images of known visual properties?
  - Scalability. For example, does this image segmentation algorithm apply to data sets of different spatial scales, e.g., mosaics of hundreds of satellite images to generate classification maps at global scale where small but genuine image details (e.g., one pixel-wide roads) must be well preserved? I am afraid it does not... Does it apply to different sensors and users?
  - Efficiency in computation time and memory occupation.
  - Accuracy in terms of spatial quality of the segment boundaries (Baraldi et al., 2005; Persello & Bruzzone, 2010).

  The conclusion is that based on existing literature the overall quality of the Diamant image segmentation algorithm remains unknown, which is often the case with the dozens of alternative image segmentation algorithms published in RS and CV literature each year (refer to Part I Section 2.4.1.2). Perhaps it is also due to these implementation shortcomings that so many researchers and practitioners ignored or criticized Diamant's methodological speculations.

- The Diamant image segmentation algorithm is not quantitatively compared (see Part I Section 2.8) against at least one alternative approach in a test image set consisting of both real and synthetic images (Baraldi et al., 2010c).
- The image segmentation algorithm proposed in (Diamant, 2005) is not technically sound.
  - In (Diamant, 2005) Diamant writes "segmentation/classification" and then "spatially connected regional groups (of pixels)" as "clusters" rather than segments, blobs or regions (see Part I Section 2.3). It is well known that (2-D) image segmentation, labeled (supervised) data classification and unlabeled (unsupervised) data clustering are completely different inductive learning-from-data problems (see Part I Section 2.4). Mixing these terms is a relevant conceptual mistake.
  - It is well known that image region extraction is the dual task of edge detection, in fact they are both inherently ill-posed inductive learning-from-unlabeled data problems (see Part I Section 2.4.1.2). In (Diamant, 2005), quite surprisingly Diamant acknowledges the ill-posedness of edge detection, but appears to ignore the inherent ill-posedness (subjective nature) of image region extraction acknowledged by a relevant portion of existing literature (see Part I Section 2.4.1.2). In fact, he states: "the efficiency of (my own) unsupervised top-down directed region-based (learning from unlabeled data) image segmentation is hard to disprove today" (Diamant, 2005). For example, by replacing pixels belonging to the same segment with their segment-based mean value (often called mean image), Diamant's image segmentation algorithm provides as output a piecewise constant approximation of the input image. Of course, researchers and practitioners interested in texture segmentation would find the Diamant piecewise constant image segmentation of little utility. In fact, the Diamant image segmentation algorithm incorporates no texture model. In practice, it detects texture elements (textons) rather than textures (made of textons) in the image. This accounts for the subjective nature of the image segmentation problem which is apparently ignored by Diamant.
- Breaking points and failure modes of the implemented algorithm are not documented in the paper.
- Conclusions are not properly supported by results contained in the manuscript. Indeed claims such as "the efficiency of (my own) unsupervised top-down directed region-based (learning from unlabeled data) image segmentation is hard to disprove today" (Diamant, 2005) are completely unjustified in both theoretical and practical terms (see previous comments).

To summarize, the Diamant image segmentation algorithm appears as "yet another image segmentation algorithm" (Baraldi et al., 2010a) based on heuristics whose superiority against alternative approaches is completely unproved. In other words, the image segmentation algorithm proposed by Diamant cannot be considered as adequate proof of his concepts (see Part I Section 2.2.3.2).

3.3 Comments on the Diamant contour detector

In practical terms, the contour detection algorithm proposed by Diamant can be subjected to the following criticisms.

- Status = Eq. (1-4), is nothing new, but a well-known isotropic zero dc-value mexican-hat operator for contrast detection (Canny, 1986; Burt, & Adelson,1983; Marr, 1982; Jain & Healey, 1998).
- Intensity information, I_{int} = Eq. (1-3), is another contrast value. However, it does not feature zero dc-value. This means the following.
  - Correlation between I_{int} = Eq. (1-3) and status = Eq. (1-4) can be relevant, i.e., I_{loc} = Eq. (1-2) = Eq. (1-3) × Eq. (1-5) is the product of two correlated contrast values where one-of-two is absolute valued.
  - Term I_{int} = Eq. (1-3) is not consistent with the psychophysical phenomenon of the Mach bands: where a luminance (radiance, intensity) ramp meets a plateau, there are spikes of brightness (perceived luminance), whereas there are none in the luminance profile. This is the sole case of continuity in the luminance profile capable of generating spikes of brightness (Baraldi & Parmiggiani, 1996a).
- The Diamant contour detection is single scale. On the contrary, it is known that the human visual system employs at least four spatial scales of analysis (Wilson & Bergen, 1979) (see Part I Section 2.3).
- The Diamant contour detector is not quantitatively compared (see Section 2.7) against at least one alternative approach in a test image set consisting of both real and synthetic images (Baraldi et al., 2010c).

To summarize, the Diamant contour detector appears to be neither new nor biologically plausible. It can be considered as "yet another contour detector" (Baraldi et al., 2010a) based on heuristics whose superiority against alternative approaches is completely unproved. In other words, the contour detector proposed by Diamant cannot be considered as adequate proof of his concepts (see Part I Section 2.2.3.2).

4. Revised/novel definitions of objective continuous sub-symbolic sensory data, continuous physical information, subjective discrete semi-symbolic data structure, discrete semantic-square (semantic^2) information and prior knowledge base

As a revision of Diamant's works (Diamant, 2005; Diamant, 2008; Diamant, 2010a; Diamant, 2010b), a new set of definitions of: (i) sub-symbolic objective primary data element in an objective sensory data set, (ii) semi-symbolic subjective secondary data structure, (iii) objective physical information, (iv) subjective semantic-square (semantic^2) information and (v) subjective prior knowledge base (ontology or model of the 3-D world) provided by an external subjective supervisor (human, God or equivalent machine).

4.1 Levels of aggregation of objective continuous sub-symbolic sensory data

There are five fine-to-coarse possible levels of aggregation of objective continuous sub-symbolic sensory data. These levels of aggregation are either sub-symbolic (non-semantic), semi-symbolic or symbolic. Semi-concepts are defined as stable concepts (percepts, classes of 3-D objects in the world) whose semantic meaning is adopted at the bottom level (layer 0) of an ontology (see Part I Section 2.2.2). The semantic information of semi-concepts (e.g., in a RS image, land cover semi-concepts are spectral categories such as *water or shadow*, *snow or ice*, *bare soil or built-up*, *vegetation,* etc.) is superior to that of objective data, whose semantic

information is null, but equal or inferior (i.e., not superior) to that of concepts belonging to higher levels of abstraction (aggregation) in the ontology at hand (e.g., in a RS image classification taxonomy such as the International Global Biosphere Programme (IGBP) land cover classification scheme (FAO, 2000), target (3-D) land cover classes are *water bodies*, *snow or ice*, *barren*, *urban and built-up*, *needle-leaf forest, broad-leaf forest, mixed forest, shrubland, grassland, cropland,* etc.) (Baraldi et al., 2006; Baraldi et al., 2010a; Baraldi et al., 2010b; Baraldi et al., 2010c; Baraldi, 2011a; Baraldi, 2011b). An ontology is a hierarchical abstract representation (model) of the (3-D) world. For example, well-known examples of RS data classification taxonomies are the aforementioned IGBP land cover classification scheme (FAO, 2000), the Co-ordination of Information on the Environment (CORINE) (European Commission Joint Research Center, 2005), the U.S. Geological Survey (USGS) classification hierarchy (Lillesand & Kiefer, 1994) and the Food and Agriculture Organization of the United Nations (FAO) Land Cover Classification System (LCCS) (Di Gregorio & Jansen, 2000; Herold et al., 2006). An ontology can be modeled as a semantic network consisting of a hierarchical class taxonomy, represented as an inverted tree whose leaves are at the bottom layer 0, plus relationships between classes as arcs between nodes (refer to Part I Section 2.2.2).

The five fine-to-coarse possible levels of aggregation of objective sub-symbolic sensory data are listed below.

1. **An unlabeled objective continuous (quantitative) sub-symbolic (non-semantic) sensory *scalar data element*.** For example, a one-band pixel value in an image, a character in a vocabulary, etc. This is a scalar (simple, atomic, elementary, primitive) fact (measurement, sign, symbol, character, element) resulting from an observation (examination, inspection, monitoring, measurement) of the (3-D) world.
2. **An unlabeled objective continuous sub-symbolic *primary data vector / primary data n-tuple / primary data element***, where $n \geq 1$ is the vector dimensionality. Each primary data n-tuple consists of $n \geq 1$ scalar data elements, e.g., a multi-spectral pixel value in an image, a word in a dictionary, etc. In the rest of this paper, if an unlabeled objective data set consisting of primary data elements is discrete and finite (e.g., an image as a 2-D data array), then its cardinality is identified as p (e.g., an image consists of p pixels). In this case primary data elements may be identified by integer numbers, e.g., a pixel is identified by a (row, column) coordinate pair in a (2-D) image domain. A set of sub-symbolic primary data elements (e.g., an image) can be described according to a given mathematical vocabulary/language. For example, a 2-D array of pixels (image) can be encoded as a 2-D spatial frequency function by means of a 2-D fast Fourier transform (FFT).
3. **A finite set** (e.g., a (2-D) image array) **of p unlabeled objective continuous sub-symbolic primary data elements** (e.g., pixels), with $p \in \{1, \infty)$**.** To be described in physical terms, a set of objective sub-symbolic primary data elements requires a mathematical vocabulary/language, e.g., a 2-D FFT of a (2-D) image. This is related to the concept of continuous physical information in an objective sensory data set (refer to this text below).
4. **A labeled subjective discrete semi-symbolic *secondary data structure / secondary data object*.** It consists of one or more primary data elements of a given objective data set grouped together (based on any possible subjective aggregation criterion) and labeled

as one semi-symbolic secondary data structure. Each label belongs to a discrete and finite set of semi-concepts. The semantic meaning of semi-concepts (e.g., *vegetation*) is superior to zero (like that of unlabeled primary data elements) and not superior (i.e., equal or inferior) to that of concepts in the real (3-D) world. A discrete and finite quantitative data set consisting of p unlabeled objective primary data elements (e.g., a multi-spectral image consisting of p pixels, refer to point 3. above) always consists of a discrete and finite set of semi-symbolic secondary data structures whose cardinality is identified hereafter as s, such that inequality ($s \leq p$) always holds. It is noteworthy that if equality ($s == p$) holds, this does not correspond to a trivial case since secondary data structures are semi-symbolic while primary data elements are sub-symbolic. To the best of this author's knowledge, it is at the level of subjective semi-symbolic secondary data structures that the view of the present author starts diverging from *all* existing CV algorithms and implementations, including GEOBIA-based RS-IUSs and Diamant's image segmentation and contour detection algorithms. This degree of novelty is consistent with well-known evidence collected in CV and MAL domains. For example:

- A large section of the scientific community acknowledges that detection of data structures in an unlabeled objective data set, such as the detection of unlabeled data clusters and unlabeled (2-D) image segments (see Part I Section 2.4.1), is an inherently subjective (which is tantamount to saying semantic, since words subjective and semantic are synonyms, refer to Part I Section 2.1) ill-posed problem, therefore difficult to solve, which requires prior (semantic) knowledge to become better posed (tractable) (refer to Part I Section 2.1).
- According to Marr, "vision goes symbolic immediately, right at the level of zero-crossing (primal sketch)... without loss of information" (Marr, 1982) (p. 343) (refer to Part I Section 2.3). Secondary semi-symbolic data structures (e.g., image segments labeled as *vegetation*) can be described (encoded) according to a given pair of one mathematical and one natural vocabulary/language to account for, respectively, their objective (quantitative) and subjective (semantic, qualitative) properties. For example, semi-symbolic image segments can be described by a segment description table whose columns consist of: (a) a segment-specific semantic label belonging to a discrete and finite set of semi-concepts (refer to this text above) and (b) segment-specific quantitative descriptors such as (Matsuyama & Shang-Shouq Hwang, 1990): (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, (v) morphological properties, (vi) spatial non-topological relationships between objects (e.g., distance, angle/orientation, etc.), (vii) spatial topological relationships between objects (e.g., adjacency, inclusion), etc. (Baraldi et al., 2010a).

In practice, the following definition holds.

Discrete semi-symbolic secondary data structure = Continuous sub-symbolic primary data element(s) + discrete semi-symbolic label belonging to a discrete and finite set of semi-concepts (e.g., in RS image understanding, possible semi-concepts are spectral categories equivalent to land cover class sets consisting of one or more land cover classes; examples of spectral categories are *vegetation*, *water or shadow*, *bare soil or built-*

up, etc. (Baraldi et al., 2006; Baraldi et al., 2010a; Baraldi et al., 2010b; Baraldi, 2011a; Baraldi, 2011b; Baraldi et al., 2010c)).

This also means that the set of discrete semi-symbolic secondary data structures incorporates the continuous objective sensory data set.

5. **A finite set of ($s \leq p$) labeled secondary subjective semi-symbolic data structures, which include the objective sensory data set** (refer to point 4. above**),** with $s \in \{1, p\}$. In this author's terminology, it is called ***preliminary classification map*** or ***primal sketch***. These terms are:
   - in line with the CV system proposed by Marr at the level of computational theory (see Part I Section 2.6) when he states: "vision goes symbolic almost immediately, right at the level of zero-crossings (primal sketch)… without loss of information" (Marr, 1982) (p. 343) (refer to Part I Section 2.3)
   - In contrast with the CV system proposed by Marr at the level of algorithm design and implementation (see Part I Section 2.5), where the term primal sketch identifies the non-symbolic output of a zero-crossings algorithm, which is an instance of the unlabeled data learning class of image edge detectors/region extractors (Marr, 1982).

   It is noteworthy that in a (2-D) preliminary classification map domain, a labeled semi-symbolic segment may be defined as a spatially connected set of secondary semi-symbolic data structures featuring the same label, say, connected pixels featuring label *vegetation*. Therefore, in a (2-D) preliminary classification map domain, semi-symbolic pixels belong to semi-symbolic image segments which belong to semi-symbolic image strata (layers) defined as image-wide sets of semi-symbolic segments featuring the same semi-symbolic label. In other words, in the preliminary classification map domain, three spatial types co-exist: **semi-symbolic pixels in semi-symbolic image segments in semi-symbolic image strata**. This would end the bad-faith antagonism between unlabeled pixels versus labeled non-symbolic segments (e.g., segment 1, segment 2, etc.) which affects traditional pixel-based versus object-based RS-IUSs and CV systems (refer to Part I Table 1). A labeled subjective semi-symbolic quantitative data set can be described (encoded) according to a given pair of one mathematical and one natural vocabulary/language capable of accounting for both the quantitative and semantic (qualitative, subjective) nature of labeled subjective semi-symbolic secondary data structures (refer to point 4. above).

4.2 Continuous physical information

Continuous physical (quantitative, objective, sensory) information. This is a hierarchical (i.e., multi-scale, including one-scale as a special case) description (representation), namely, down-scale encoding (decomposition), up-scale decoding (reconstruction) or one-scale transcoding (from one data format to another at the same hierarchical level), of the physical objective data set based on a given mathematical non-natural vocabulary/language. This hierarchical description/ representation of the objective sensory data set can be either lossless or lossy, depending on the exact/non-exact reconstruction (decoding) of the original data set from its representation (encoding). For example, an FFT of a time-signal is a one-scale transcodification of the signal from the time to the frequency domain. A well-known example of down-scale encoding/up-scale decoding is the Gaussian-Laplacian image pyramid (Burt & Adelson, 1983). It means that physical

information stems from the combination of an objective data set with a mathematical non-natural vocabulary/language. To summarize the concept of physical information, we can write the following definition.

Continuous objective data set + (arbitrary) multi-scale down-scale encoding, up-scale decoding or one-scale transcoding/description/data format = hierarchical physical information encompassing down-scale/ fine-to-coarse resolution/ compression/ encoding, up-scale/ coarse-to-fine resolution/ decompression/ decoding, and/or one-scale transcodification (from one data format to another at the same hierarchical level), either lossless or lossy.

4.3 Discrete semantic-square information

Discrete semantic-square (semantic²) (where semantic is a synonym of categorical, symbolic, subjective, abstract, qualitative, vague, but persistent, stable, see Part I Section 2.1) **information (concepts, percepts) stems from the semantic² labeling of an objective data set performed by an external subjective supervisor** (human, God or equivalent machine) **provided with a subjective hierarchical prior knowledge base** (ontology or model of the (3-D) world, equivalent to an inverted tree with leaves at the bottom level 0, see Part I Section 2.2.2). **Semantic² labeling occurs when a subjective supervisor (first source of subjectivity), provided with his/her own subjective ontology (second source of subjectivity), observes and scrutinizes the objective data set, consisting of p sub-symbolic primary data elements (refer to point 3. in Section 4.1), to achieve the following.**

a. At the bottom level 0 of the inverted tree (ontology, see Part I Section 2.2.2), a semi-symbolic label, belonging to a discrete and finite set of semi-concepts (e.g., in a RS image, spectral categories are *vegetation*, *water or shadow, bare soil or built-up*, etc.), is assigned to each sub-symbolic primary data element (e.g., each pixel in a RS image) of a set of p sub-symbolic primary data elements to form a finite and discrete set of s semi-symbolic secondary data elements, with $\boldsymbol{s \leq p}$ (refer to point 5. in Section 4.1).
b. At hierarchical levels ≥ 1 of the inverted tree (see Part I Section 2.2.2), a hierarchical symbolic label is assigned to the set of s semi-symbolic secondary data elements based on symbolic reasoning (Matsuyama & Shang-Shouq Hwang, 1990).

This definition of semantic² labeling disagrees at the level of the aforementioned point a. with the traditional definition of semantic labeling provided by MAL, which encompasses existing CV systems (e.g., Diamant's (Diamant, 2005)) and RS-IUSs (e.g., (Definiens Imaging GmbH, 2004; Matsuyama & Shang-Shouq Hwang, 1990)). In fact, point a. above states that **semantic² information stems naturally (automatically, instantaneously) from the simultaneous interaction of three necessary and sufficient components.**

i. An objective sensory data set (consisting of facts, measures, etc.) described in terms of continuous physical information (representation, description) based on a mathematical vocabulary/language.
ii. A subjective supervisor/actor (human, God or equivalent machine). He/she acts as the first source of subjectivity in the labeling (mapping) process. To be considered as such, a supervisor must be the carrier of a prior semantic knowledge base (ontology). He/she acts as follows:

- observes the objective data set and
- interprets/scrutinizes the objective data set to match (label) data with his/her own ontology.

iii. A subjective hierarchical (multi-scale) prior ontology which exists before looking at the data. Since it deals with semantic information, a prior knowledge base is subjective by definition (since subjective and semantic are synonyms, refer to Part I Section 2.1). In practice, this ontology acts as the second source of subjectivity in the labeling (mapping) process. According to Diamant this hierarchical ontology is equivalent to a narrative story or tale which requires a natural language to comprise, in a top-down representation: the story title, index, sections, paragraphs, sentences and words. It is graphically represented and implemented as a semantic net or inverted tree whose leaves are at the bottom level 0 where physical information is incorporated (refer to this text above and Part I Section 2.2.2).

The aforementioned points i.-iii. imply that **objective sensory data**, *per se*, **do not possess any semantic² information, but physical information exclusively.** Rather, **semantic² information incorporates objective data as one-of-three components**. This also means that nobody should disagree with Diamant when he repeats over and over that sensory data do not possess semantic information, therefore semantic information cannot be *extracted* from sensory data (Diamant, 2010a). On the contrary, Diamant's statement should not be considered original at all because it has been perfectly acknowledged in philosophy for hundreds of years, as well as in psychophysical studies of perception (Matsuyama & Shang-Shouq Hwang, 1990) and MAL in the last 50 years (Cherkassky & Mulier, 2006). This concept is summarized below.

- Philosophy and psychophysical studies of perception. The statement that sensory data do not possess semantic information is tantamount to saying there is an information gap between physical information and semantic information, which is the well-known information gap between (sensory and varying) sensations and (vague, but stable) perceptions. In practice, "we are always seeing objects we have never seen before at the sensation level, while we perceive familiar objects everywhere at the perception level" (Matsuyama & Shang-Shouq Hwang, 1990) (see Part I Section 1 and Part I Section 2.2.2).
- MAL. In unlabeled data learning algorithms (e.g., unlabeled data clustering), no semantics is detected as output (e.g., unlabeled data cluster 1, unlabeled data cluster 2), see Fig. 1. In labeled data learning algorithms for classification applications (see Part I,Fig. 1), no semantic information is extracted from a finite set of training data pairs consisting of an (objective data vector, subjective discrete label), but a correlation function can be estimated between continuous sensory data and a discrete and finite set of subjective labels (refer to Part I Section 2.1 and Part I Section 2.4.2).

The foregoing comments also mean that Diamant is right, although vague, when he states that "semantics is a property of a human observer" (Diamant, 2010a). **To state this more precisely, since semantic² information naturally (automatically, instantaneously) stems from the interaction of three necessary and sufficient components i.-iii. (see above in this text), then semantic² information cannot be separated from any of its three components.** For example, let us think of a piano (symbolic data structure) whose objective presence (fact) requires the simultaneous presence of a subjective human actor (or equivalent machine) to

generate whatever sound (semantic information). The sound (generated semantic information) is neither in the piano, nor in the piano player, nor in his/her prior knowledge of what a piano is all about, but in the instantaneous combination of these three factors. This also means that **semantic² information** quite obviously **changes with the objective data set, the subjective human supervisor and his/her own subjective ontology.** In particular (refer to this text above), **semantic² information means there are two subjective actors in the semantic labeling of objective sensory data, namely, the subjective external observer and scrutinizers (or equivalent machine) and his/her own ontology or semantic (abstract) model of the world**. In fact, it is well known that all humans do not adopt the same ontology and two humans who adopt the same ontology do not apply this ontology the same way through time in interpreting a given observation. For example, two players will never generate the same music when playing the same musical score on the same piano. Not even the same player will ever generate the same music when playing twice the same musical score on the same piano. To summarize these concepts we can write the following definition.

Objective sensory data set + subjective supervisor provided, as such, with a subjective prior hierarchical knowledge base (ontology) = hierarchical semantic² (subjective²) information, which includes physical information at the bottom level 0 of the inverted tree which deals with the semantic granularity of semi-concepts assigned to semi-symbolic secondary data structures.

4.4 Subjective hierarchical (multi-scale) prior knowledge base

Subjective hierarchical (multi-scale) prior knowledge base (ontology, model of the (3-D) world) equivalent to a semantic net or inverted tree with leaves at the bottom level 0 where physical information is incorporated. Refer to this text above.

4.5 Intelligence

Intelligence (cognition) is the system's ability to aggregate bottom-up (from-data-to-concepts) and disassemble top-down (from-concepts-to-data) semantic information (which incorporates physical information) across the hierarchical levels of a subjective prior knowledge base.

4.6 Information processing system

An information processing system, cognitive system or intelligent system transforms an input sensory data set into an output instantiation of a story in natural language whose hierarchical structure is provided by an ontology or inverted tree retained in the system's memory before looking at the sensory data.

To summarize, the aforementioned novel definitions sketch a RS-IUS where information goes symbolic during the pre-attentive vision phase to generate a semi-symbolic primal sketch (preliminary classification map). This is in line with the CV system proposed by Marr at the level of computational theory (see Part I Section 2.6) when he states: "vision goes symbolic almost immediately, right at the level of zero-crossings (primal sketch)" (Marr, 1982), p. 343 (see Part I Section 2.3). However, it differs from the CV system proposed by Marr at the level of primal sketch implementation (see Part I Section 2.6) consisting of a sub-

symbolic zero-crossing algorithm (Marr, 1982). In addition, the novel RS-IUS sketched above differs at the level of both computational theory and algorithm design and implementation from existing CV systems such as GEOBIA systems (Definiens Imaging GmbH, 2004; Esch et al., 2008), including Diamant's (Diamant, 2005; Diamant, 2008; Diamant, 2010a; Diamant, 2010b), where an unlabeled data learning (driven-without-knowledge) algorithm is adopted at the first stage.

5. Practical consequences of the proposed definitions on CV, AI and MAL system design and implementation strategies

Practical consequences of the definitions proposed in Part II Section 4 on CV, AI and MAL system design and implementation strategies are several, more detailed, better posed and, therefore, far more relevant than Diamant's (Diamant, 2010a). Thus, they should benefit from more favorable consideration by the scientific community.

1. Definitions provided in Part II Section 4 are consistent with the Marr statement: "vision goes symbolic almost immediately, right at the level of zero-crossings (primal sketch)… without loss of information" (Marr, 1982) (p. 343) (refer to Part I Section 2.3). This is tantamount to saying that exploitation of the deductive subjective prior knowledge-based inference paradigm must regard the preattentive visual phase whose output, the so-called *primal sketch* (Marr, 1982), must be as follows:
   - semantic in nature (see Part I Section 2.3), therefore it is called *preliminary classification map*;
   - capable of preserving small, but genuine image details (high spatial frequency image components). This requirement is inconsistent with existing image segmentation algorithms which are inherently affected by the *uncertainty principle* according to which, for any contextual (neighborhood) property, we cannot simultaneously measure that property while obtaining accurate localization (Corcoran & Winstanley, 2007; Petrou & Sevilla, 2006) (see Part I Section 2.4.1.2).

   Although he stated that vision goes symbolic right at the output of the preattentive vision phase, which has to affect the architectural level of understanding of a CV system (see Part I Section 2.6), Marr selected a sub-symbolic edge detection (zero-crossing) algorithmic for primal sketch generation (Marr, 1982). By embracing the Marr computational theory rather than his algorithmic solutions, the present author concludes that, as output, **the preattentive visual phase no longer generates sub-symbolic image primitives, namely, non-semantic points and edges or, vice versa, image regions (which is what was implemented by Marr (Marr, 1982)), but semi-symbolic secondary data structures, namely, semi-symbolic pixels in semi-symbolic segments in semi-symbolic strata** (see Part II Section 4) (Baraldi et al., 2006; Baraldi et al., 2010a; Baraldi et al., 2010b; Baraldi et al., 2010c; Baraldi, 2011a; Baraldi, 2011b).

2. **It is impossible to *extract* semantic^2 information from objective continuous sensory data because the latter, *per se*, are provided with no semantics at all.** This is the well-known information gap between semantic^2 information and physical information (refer to Part I Section 2.2.2 and Part I Section 2.3).

3. **Although it is impossible to *extract* semantic^2 information from objective continuous sensory data, it is possible to *correlate* discrete semantic^2 information to objective**

continuous sensory data. This conclusion is by no means novel as it is well known in literature. For example, Shunlin Liang summarizes this concept in a few words: statistical pattern recognition systems are based on correlation relationships between objective sensory (e.g., RS) data and either continuous (e.g., LAI) or categorical (e.g., land surface) variables (see Part I Section 2.1) (Shunlin Liang, 2004). **Unfortunately, low or no correlation can be found between continuous sensory data and a finite and discrete set of categorical variables, corresponding to independent random variables generating "distinguishable" data structures (data aggregations, data clusters) in real-world data mapping problems at large data scale or fine semantic granularity, other than toy problems at small data scale and coarse semantic granularity.** This low correlation effect is due to the combination of two factors.

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

4. As an extension of points 2. and 3. above, **unlabeled (unsupervised) data learning algorithms,** namely, driven-without-knowledge image segmentation algorithms and unlabeled data clustering algorithms (see Part I Section 2.4.1), **should be considered highly inappropriate** (like using a fork for cutting food: unless the food is particularly soft, it will never work) **when the objective sensory data acquisition occurs in the domain of real-world data mapping problems at large data scale or fine semantic granularity** (where the separability of "distinguishable" data structures in a given objective sensory data set belonging to a given measurement space is expected to be low), other than toy problems at small data scale and coarse semantic granularity.

Does this mean the relevant effort spent by the MAL community to develop driven-without-knowledge image segmentation algorithms (Castilla et al., 2008) or, say, self-organizing topology-preserving unlabeled data clustering algorithms (Fritzke, 1997; Martinetz & Schulten, 1994), has been worthless? Fortunately, not. It rather means the following.

i. The main application domain of, say, self-organizing topology-preserving unlabeled data clustering algorithms should remain the modeling of stationary and non-stationary distributions, see Part I Fig. 1.
ii. When an unlabeled (unsupervised) data learning algorithm, either a driven-without-knowledge image segmentation algorithm or an unlabeled data clustering algorithm (see Part I Section 2.4.1), is adopted as the first stage of a two-stage hybrid cognitive system, CV system or RS-IUS, it should be considered highly inappropriate. In particular:
    I It should be replaced by a deductive MAT-by-rules approach where community-agreed prior knowledge is conveyed to generate as output a lossless semi-symbolic product (consisting of semi-concepts). For example, in a RS-IUS, the MAT-by-rules first stage should generate a preliminary classification map (see Part II Section 4) where small, but genuine image details are well preserved (refer to this text above).
    II If useful, it should be:
        a. adapted to work on a driven-by-knowledge stratified (semantic masked) basis and
        c. next, moved to the second stage of a two-stage stratified hierarchical hybrid cognitive system. For example, a two-stage stratified hierarchical hybrid RS-IUS architecture has been proposed in recent literature, see Fig. 3 (Baraldi et al., 2006; Baraldi et al., 2010a; Baraldi et al., 2010b; Baraldi et al., 2010c; Baraldi, 2011a; Baraldi, 2011b).

5. As an extension of points 2. and 3. above, **labeled (supervised) data learning classifiers (**see Part I Section 2.4.2**) should be considered highly inappropriate** (like using a fork for cutting food; unless the food is particularly soft, it will never work) **in real-world data mapping problems at large data scale or fine semantic granularity** (where within-class variability is monotonically non-decreasing (i.e., it increases or remains equal) with the cardinality of the objective sensory data set), other than toy problems at small data scale and coarse semantic granularity. This conclusion is by no means novel. Rather, it is well known in literature. For example, Shunlin Liang summarizes this concept in few words: statistical model are usually site-specific (see Part I Section 2.1) (Shunlin Liang, 2004). Does this mean the relevant effort spent by the MAL community to develop supervised data learning classifiers has been worthless? Fortunately, no. It rather means the following.
    i. The main application domain of supervised data learning algorithms should be considered function regression where input and output variables are continuous non-semantic, see Fig. 1.
    ii. When a supervised data learning classifier (see Part I Section 2.4.2) is adopted as the first stage of a two-stage hybrid cognitive system, CV system or RS-IUS, it should be considered highly inappropriate. An experimental proof of this concept is that supervised MAL algorithms (say, SVMs), either context-insensitive (e.g.,

pixel-based) or context-sensitive (Bruzzone & Carlin, 2006; Bruzzone & Persello, 2009; Persello & Bruzzone, 2010), considered successful in terms of operational QIs (refer to Part I Section 2.7.2) at local/regional scale, become impracticable in mapping RS image mosaics consisting of hundreds of images at national/continental/global scale (Chengquan Huang et al., 2008). In these real world problems the cost, timeliness, quality and availability of adequate reference (training) data sets derived from field sites, existing maps and tabular data are currently considered the most limiting factors on RS data product generation and validation (Gutman et al., 2004). In particular, the first-stage supervised data learning classifier of a two-stage hybrid RS-IUS should be:

I replaced by a deductive MAT-by-rule approach where community-agreed prior knowledge is conveyed to generate a preliminary classification map (see Part II Section 4) where small, but genuine image details are well preserved (refer to this text above);

II if useful, it should be:
   a. adapted to work on a driven-by-knowledge stratified (semantic masked) basis and
   d. next, moved to the second stage of a two-stage stratified hierarchical hybrid RS-IUS architecture proposed in recent literature, see Fig. 3 (Baraldi et al., 2006; Baraldi et al., 2010a; Baraldi et al., 2010b; Baraldi, 2011a; Baraldi, 2011b; Baraldi et al., 2010c).

6. SIAM™ as a proof of the efficacy of the required shift of learning paradigm from MAL-from-examples to MAT-by-rules at the first stage of two-stage hybrid RS-IUSs

To the best of this author's knowledge SIAM™ provides the first experimental proof of the efficacy of the required switch of learning paradigm from MAL-from-examples to MAT-by-rules at the first stage of a two-stage hybrid RS-IUS architecture (refer to Part II Section 2.3), see Table 4. SIAM™ is an operational (good-to-go, press-and-go, turnkey) software button (executable). In particular, SIAM™ is automatic, efficient, scalable, accurate and robust to changes in the input data acquired across time, space and sensors. For example, the automatic SIAM™ is consistent and accurate across sensors at the national/ continental/ global scale (refer to Part II Section 2.3) (Baraldi et al., 2006; Baraldi et al., 2010a; Baraldi et al., 2010b; Baraldi et al., 2010c; Baraldi, 2011a; Baraldi, 2011b), whereas semi-automatic inductive data learning neural network approaches, such as SVMs, require to be re-trained (supervised) image-wide (Chengquan Huang et al., 2008).

SIAM™ belongs to the family of physical models that follow the physical laws of the real (3-D) world to represent an abstract of the reality (see Part I Section 2.1) (Shunlin Liang, 2004). In particular, SIAM™ follows the physical laws of spaceborne optical imaging devices to provide a two-stage hybrid RS-IUS with a first-stage deductive prior knowledge-based inference mechanism. Unfortunately, it takes a long time for human experts to learn physical laws of the real (3-D) world and tune physical models based on human intuition, domain expertise and evidence from data observations (Mather, 1994; Shunlin Liang, 2004). For example, the development of the SIAM™ dates back to the year 2002 (Baraldi, 2011a).

Quality Indicators (Qis)	State-of-the-art RS-IUSs	SIAM™
Degree of automation: (a) number, physical meaning and range of variation of user-defined parameters, (b)collection of the required training data set, if any.	VL, L	VH (fully automatic, it cannot be surpassed)
Effectiveness : (a) semantic accuracy and (b) spatial accuracy.	M, H, VH	VH
Semantic information level	Land cover class (e.g., *deciduous forest*)	Spectral semi-concept (e.g., *vegetation*)
Efficiency: (a) computation time and (b) memory occupation.	VL, L in training (hours per images)	VH (5 m to 30 s per Landsat image in a laptop)
Robustness to changes in input image	VL (specific training per image)	VH
Robustness to changes in input parameters	VL	VH (it cannot be surpassed)
Scalability to changes in the sensor's specifications or user's needs.	VL	VH (it works with any existing spaceborne sensor)
Timeliness (from data acquisition to high-level product generation, increases with manpower and computing power).	VH (e.g., the collection of reference samples is a difficult and expensive task)	VL, i.e., timeliness is reduced to almost zero
Economy (inverse of costs increasing with manpower and computing power).	VL, L, high costs in manpower and also computing power	VH, i.e., costs in manpower and computing power are reduced to almost zero

Table 4. QIs of SIAM™ versus state-of-the-art RS-IUSs' (refer to Part I Section 2.8). Legend of fuzzy sets: Very low (VL), Low (L), Medium (M), High (H), Very High (VH). Legend of colors: Red-Bad, Blue-Average, Green-Good

Part I Section 2.2.2 reported the question: is human biology as irrelevant to AI research as bird biology is to aeronautical engineering? Actually, biological vision has always represented a fundamental source of inspiration for the CV community. While SIAM™ considers its degree of biological plausibility as a value added, straightforward imitation of biological vision solutions is not always possible. This is the reason why SIAM™ cannot be considered highly plausible in biological terms although it is very useful in practice. For example, SIAM™ cannot work with panchromatic imagery whereas the human visual system is perfectly able to interpret gray-tone images.

7. Conclusions

It is well known that semantic information is not in objective sensory data, which is tantamount to saying there is a well-known information gap between semantic^2 information and physical information. This conceptual work observes that semantic^2 information is naturally (automatically, instantaneously) generated by the simultaneous interaction of a subjective external supervisor who observes and scrutinizes an objective sensory data set based on his/her own subjective prior knowledge base (ontology, model of the 3-D world). Semantic^2 information resulting from this interaction takes the intermediate form of semi-symbolic secondary data structures that incorporate physical information at the bottom level (layer 0) of an ontology represented as an inverted tree.

A shift of learning paradigm from MAL-from-examples to MAT-by-rules in the first stage of two-stage hybrid RS-IUSs is recommended. Experimental proof of this concept is provided by the operational automatic SIAM™ recently proposed in RS literature.

The practical conclusion of this conceptual work is twofold.

1. In line with a relevant section of existing literature (Shunlin Liang, 2004), labeled (supervised) data learning classifiers (see Part I Section 2.4.2) should be considered highly inappropriate, being affected by low operational QIs (see Part I Section 2.8), in dealing with real-world data mapping problems at large data scale (e.g., RS image mapping at national/ continental/ global scale) or fine semantic granularity, except in the case of toy problems at small data scale and coarse semantic granularity (e.g., RS image mapping at coarse spatial resolution and local/regional scale). This awareness should be divulged among the RS, CV, AI and MAL communities.
2. Any inductive MAL-from-examples algorithm, whether labeled (supervised, e.g., SVMs) or unlabeled (e.g., image segmentation, unlabeled data clustering), whether context-insensitive (e.g., pixel-based) or context-sensitive (e.g., (2-D) object-based), employed as the first stage of a two-stage hybrid cognitive system, CV system or RS-IUS, should be:
    a. replaced by a deductive MAT-by-rules approach where community-agreed prior knowledge is conveyed and,
    b. if useful, adapted to work on a driven-by-knowledge stratified (semantic masked) basis and moved to the second stage of a two-stage stratified hierarchical hybrid cognitive system. For example, a two-stage stratified hierarchical hybrid RS-IUS architecture has been proposed in recent literature, see Fig. 3 (Baraldi et al., 2006; Baraldi et al., 2010a; Baraldi et al., 2010b; Baraldi et al., 2010c; Baraldi, 2011a; Baraldi, 2011b).

This required shift of the learning paradigm from MAL-from-examples to MAT-by-rules adopted in the first stage of a two-stage hybrid RS-IUS is similar in nature to previous conceptual shifts occurring between deductive coarse-to-fine (from symbolic concepts to sub-symbolic data) AI/MAI and inductive fine-to-coarse (from sub-symbolic data to symbolic concepts) Cybernetics/MAL, see Part I Section 2.2. What is novel about the proposed shift of the learning paradigm from MAL-from-examples to MAT-by-rules at the first stage of a two-stage hybrid RS-IUS is the following.

- Its aim is to accomplish the following fundamental observation by Marr: "vision goes symbolic almost immediately, right at the level of zero-crossings (primal sketch)… without loss of information" (Marr, 1982) (p. 343) (see Part I Section 1, Part I Section 2.2.2 and Part I Section 2.3), which means that exploitation of the deductive subjective prior knowledge-based inference paradigm must regard the preattentive visual phase whose output product, known as *primal sketch*, must be: (i) semantic in nature (in disagreement with the Marr algorithmic solution of zero-crossings), therefore it is called *preliminary classification map* (see Part II Section 4) and (ii) capable of preserving small, but genuine image details, unlike existing image segmentation algorithms affected by the *uncertainty principle* (Corcoran & Winstanley, 2007; Petrou & Sevilla, 2006) (see Part I Section 2.4.1.2).
- It comes together with a novel conceptual framework consisting of explicit definitions of: (i) sub-symbolic objective primary data element in an objective sensory data set, (ii) semi-symbolic subjective secondary data structure, (iii) objective physical information, (iv) subjective semantic[2] information and (v)

subjective prior knowledge base (ontology or model of the 3-D world (Matsuyama & Shang-Shouq Hwang, 1990)) provided by an external subjective supervisor (human, God or equivalent machine), refer to Part II Section 4.

- It affects exclusively the inductive learning-from-data first stage of traditional two-stage hybrid CV systems (e.g., Marr's (Marr, 1982), Diamant's (Diamant, 2005; Diamant, 2008; Diamant, 2010a; Diamant, 2010b)) or RS-IUSs, whether or not this first stage is implemented as an inductive algorithm capable of learning from either unlabeled (unsupervised) or labeled (supervised) data, whether context-insensitive (e.g., pixel-based) or context-sensitive (e.g., (2-D) object-based). If useful, these inductive data learning algorithms may be adapted to run on a driven-by-knowledge stratified (semantic masked, layered) basis and moved to the second stage of a novel two-stage stratified hierarchical hybrid RS-IUS architecture proposed in recent literature, see Fig. 3 (Baraldi et al., 2006; Baraldi et al., 2010a; Baraldi et al., 2010b; Baraldi et al., 2010c; Baraldi, 2011a; Baraldi, 2011b).
- It comes together with a novel two-stage stratified hierarchical hybrid RS-IUS architecture employing a first-stage spectral rule-based preliminary classification algorithm based on prior spectral knowledge, see Fig. 3 (Baraldi et al., 2006; Baraldi et al., 2010a; Baraldi et al., 2010b; Baraldi et al., 2010c; Baraldi, 2011a; Baraldi, 2011b).
- It comes together with an operational (namely, automatic, efficient, accurate, robust, scalable, see Part I Section 2.8) Satellite Image Automatic Mapper™ (SIAM™) implementation (software executable), equivalent to an automatic (good-to-go, press-and-go, turnkey) software button, provided as an experimental proof of the efficacy of the required shift in learning paradigm from MAL-from-examples to MAT-by-rules at the first stage of a two-stage hybrid RS-IUS architecture, see Fig. 3 (Baraldi et al., 2006; Baraldi et al., 2010a; Baraldi et al., 2010b; Baraldi et al., 2010c; Baraldi, 2011a; Baraldi, 2011b).

To summarize, to the best of this author's knowledge this is the first time a novel computational theory (RS-IUS architecture) is supported by operational (good-to-go, press-and-go, turnkey) algorithmic and implementation solutions as proofs of concept. For example, this was not the case of the Marr (Marr, 1982) or the Diamant CV systems (Diamant, 2005; Diamant, 2008; Diamant, 2010a; Diamant, 2010b), whose computational theories (see Part I Section 2.6) are both inconsistent with algorithmic solutions adopted by their authors. As a consequence, these two CV systems become two more instances of the well-known class of two-stage segment-based hybrid CV systems, also termed GEOBIA systems, traditionally affected by a lack of general consensus and research (Hay & Castilla, 2006; Matsuyama & Shang-Shouq Hwang, 1990).

The proposed conclusions of potential interest to the RS, CV, AI and MAL communities are supported by unquestionable independent sources of evidence listed below.

- Since the late 1950s, the original ambitious goals of AI/MAI and Cybernetics/MAL have been fragmented into “practical” and “manageable” problems equivalent to "a family of relatively disconnected efforts” (Diamant, 2005; Diamant, 2008; Diamant, 2010a; Diamant, 2010b).

- It is well-known in literature that inductive learning-from-examples "is an inherently difficult (ill-posed) problem and its solution requires a priori knowledge in addition to data" (Cherkassky & Mulier, 2006) (p. 39) (see Part I Section 2.1). In practical contexts this means the following.
  - Unlabeled (unsupervised) data learning algorithms, namely, unlabeled data clustering (Backer & Jain, 1981; Baraldi & Alpaydin, 2002a; Baraldi & Alpaydin, 2002b; Cherkassky & Mulier, 2006; Fritzke, 1997) and unlabeled (2-D) image segmentation algorithms (Burr & Morrone, 1992; Corcoran et al., 2010; Corcoran & Winstanley, 2007; Delves et al., 1992; Hay & Castilla, 2006; Matsuyama & Shang-Shouq Hwang, 1990; Petrou & Sevilla, 2006; Vecera & Farah, 1997), are recognized as inherently ill-posed problems subjective in nature by a relevant portion of existing literature.
  - Labeled (supervised) data learning classifiers are unable to establish correlation relationships between objective sensory (e.g., RS) data and categorical variables (e.g., land cover classes) at large data scale or fine semantic granularity. For example, in (Chengquan Huang et al., 2008) a forest/non-forest one-class SVM battery of classifiers must be re-trained and re-selected for every image in an image mosaic at global scale. Vice versa, labeled data learning classifiers are exclusively suitable for finding correlation relationships between objective sensory data and categorical variables at small data scale and coarse semantic granularity (e.g., in RS data mapping problems at coarse spatial resolution and local/regional scale). In fact, in practical RS data applications where supervised data learning algorithms are employed at large spatial scale, fine spatial resolution or fine semantic granularity (Chengquan Huang et al., 2008), the cost, timeliness, quality and availability of adequate reference (training/testing) datasets derived from field sites, existing maps and tabular data have turned out to be the most limiting factors on RS data product generation and validation (Gutman et al., 2004).
- The prior knowledge-based SIAM™ is provided with unsurpassed operational QIs (see Part I Section 2.8) in the mapping of RS image mosaics at national/ continental/ global scale (e.g., refer to Table 4).

To the best of this author's knowledge, while the proposed practical conclusions of potential interest to the RS, CV, AI and MAL communities are supported by the aforementioned independent sources of evidence, these conclusions are not contradicted by any practical achievement gained by the RS, CV, AI and MAL communities in recent years. Thus, rather than being agreed or disagreed upon, these conclusions ought to be accepted by the scientific community unless proved otherwise when the increasing rate of collection of RS data of enhanced spatial, spectral and temporal quality will no longer outpace our capability of generating (rather than extracting) semantic[2] information from RS data provided, *per se*, with no semantics at all.

8. Acknowledgments

This material is partly based upon work supported by the National Aeronautics and Space Administration under Grant/Contract/Agreement No. NNX07AV19G issued through the Earth Science Division of the Science Mission Directorate. The research leading to these results has also received funding from the European Union Seventh Framework Programme

FP7/2007-2013 under grant agreement n° 263435. This author wishes to thank the Editorial Board of InTech for its competence and willingness to help.

9. References

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

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

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

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

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

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

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

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

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

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

Baraldi, A. (2011a). Levels of understanding and degrees of novelty of a two-stage remote sensing image understanding system employing the Satellite Image Automatic Mapper™ (SIAM™) as its preliminary classification first stage. *IEEE Trans. Geosci. Remote Sensing*, submitted for consideration for publication, TGRS-2011-00241.

Baraldi, A. (2011b). Fuzzification of a crisp near real-time operational automatic spectral rule-based decision-tree preliminary classifier of multi-source multi-spectral

remotely-sensed images, *IEEE Trans. Geosci. Remote Sensing*, accepted for publication, July 2011.

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

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

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

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

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

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

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

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

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

Cootes, T. F. & Taylor, C. J. (2004). *Statistical Models of Appearance for Computer Vision, Imaging Science and Biomedical Engineering*, University of Manchester, 17.04.2011, Available from: www.isbe.man.ac.uk/~bim/Models/app_model.ps.gz.

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

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

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

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

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

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

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

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

Di Gregorio, A. & Jansen, L. (2000). *Land Cover Classification System (LCCS): Classification concepts and user manual*, FAO Corporate Document Repository, 17.04.2011, Available from: http://www.fao.org/DOCREP/003/X0596E/X0596e00.htm

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

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

Forestry Department, Food and Agriculture Organization (FAO) of the United Nations (2000). *FRA 2000 - Forest Cover Mapping & Monitoring with NOAA-AVHRR & Other Coarse Spatial Resolution Sensors*, Rome, 2000.

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

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

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

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

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

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

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

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

Herold, M.; Woodcock, C.; Di Gregorio, A.; Mayaux, P.; Belward, A. S.; Latham, J. & Schmullius, C. (2006). A joint initiative for harmonization and validation of land cover datasets. *IEEE Trans. Geosci. Remote Sensing*, Vol. 44, No. 7, pp. 1719–1727.

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

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

Lillesand, T. & Kiefer, R. (1994). *Remote Sensing and Image Interpretation*. Wiley & Sons, New York.

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

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

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

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

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

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

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

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

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

Shackelford, A. K. & Davis, C. H. (2003a). A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas. *IEEE Trans. Geosci. Remote Sensing*, Vol. 41, No. 9, pp. 1920–1932.

Shackelford, A. K. & Davis, C. H. (2003b). A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas. *IEEE Trans. Geosci. Remote Sensing*, Vol. 41, No. 10, pp. 2354–2363.

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

Tapsall, B.; Milenov, P. & Tasdemir, K. (2010). Analysis of RapidEye imagery for annual land cover mapping as an aid to European Union (EU) Common agricultural policy, W. Wagner, B. Székely, (Eds.*): ISPRS TC VII Symposium – 100 Years ISPRS*, Vienna, Austria, July 5–7, 2010, IAPRS, Vol. XXXVIII, Part 7B. 17.04.2011, Available from: http://mars.jrc.it/mars/content/download/1648/8982/file/P4-5_Milenov_Tapsall_BG_RapidEye_Project_JRC_ver2.pdf

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

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

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

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

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

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