

AUTOMATED NEAR REAL-TIME EARTH OBSERVATION LEVEL 2 PRODUCT GENERATION FOR SEMANTIC QUERYING

A. Baraldi^{a,*}, Dirk Tiede^b, Martin Sudmanns^b, Mariana Belgiu^b, and Stefan Lang^b

^a Department of Agricultural and Food Sciences, University of Naples Federico II, Italy - andrea6311@gmail.com

^b Paris-Lodron Universität Salzburg, IFFB Geoinformatik, Schillerstraße 30, A-5020 Salzburg – (dirk.tiede, martin.sudmanns, mariana.belgiu, stefan.lang@sbg.ac.at)

KEY WORDS: Algebra for spatiotemporal objects and events, Cognitive science, Earth observation Level 2 product, Semantic content-based image retrieval, Semantic network, Vision

ABSTRACT:

Existing Earth observation (EO) content-based image retrieval (CBIR) systems support human-machine interaction through queries by metadata text information, image-wide summary statistics or either image, object or multi-object examples. No semantic CBIR (SCBIR) system in operating mode has ever been developed by the remote sensing (RS) community. At the same time, no EO data-derived Level 2 prototype product has ever been generated systematically at the ground segment, in contrast with the visionary goal of the Global Earth Observation System of Systems (GEOSS) implementation plan for years 2005-2015. Typical EO Level 2 products include: (i) a multi-spectral (MS) image corrected for atmospheric, adjacency and topographic effects, and (ii) a scene classification map (SCM), encompassing cloud and cloud-shadow quality layers. This work presents an original hybrid (combined deductive and inductive) feedback EO image understanding for semantic querying (EO-IU4SQ) system as a proof-of-concept where systematic multi-source EO big data transformation into Level 2 products is accomplished as a pre-condition for SCBIR, in agreement with the object-based image analysis (OBIA) paradigm and the Quality Assurance Framework for Earth Observation (QA4EO) guidelines. In the hybrid EO-IU4SQ system, statistical model-based/bottom-up/inductive/machine learning-from-data algorithms and physical model-based/top-down/deductive/human-to-machine knowledge transfer approaches are combined with feedback loops to take advantage of the complementary features of each and overcome their shortcomings.

1. INTRODUCTION

Existing Earth observation (EO) content-based image retrieval (CBIR) systems support human-machine interaction through queries by metadata text information (e.g., EO image acquisition time, geographic area of interest, etc.), summary statistics (e.g., image-wide cloud cover quality index), or by either image, object or multi-object examples (Smeulders et al., 2000), see Figure 1. To our best knowledge, no semantic CBIR (SCBIR) system, capable of processing semantic queries such as “retrieve all EO images in the image base where there is a lake adjacent to a highway”, has ever been developed by the remote sensing (RS) community in operating mode (Dumitru et al., 2015). To be considered in operating mode an EO SCBIR system is required to score “high” in a set of community-agreed outcome and process (OP) quantitative quality indicators (Q²Is) including degree of automation, accuracy, efficiency, scalability to changes in sensor specification and user requirements, robustness to changes in input parameters, robustness to changes in input data, timeliness from data acquisition to product generation, and costs in manpower and computer power (Baraldi and Boschetti, 2012), in agreement with the Quality Assurance Framework for Earth Observation (QA4EO) Calibration/Validation (*Cal/Val*) guidelines (Group on Earth Observation, 2010). According to these definitions the Earth Observation Image Librarian (EOLib), considered the state-of-the-art in SCBIR system prototypes (Dumitru et al., 2015), is affected by several structural limitations in operating mode. EOLib adopts an inductive supervised data learning algorithm, specifically, a support vector machine (SVM), to accomplish 1D image classification, where the vector data sequence consists of image convolutional values generated by 2D spatial filters (Dumitru et al., 2015). Due to the inherent ill-posedness of machine learning-from-data algorithms (Cherkassky and Mulier, 1998), EOLib is semi-automatic and site-specific; it

requires user interaction, first, to collect supervised training samples and, second, to define the SVM’s free-parameters based on heuristics. In addition, any 1D image classifier is an orderless encoder invariant to permutations, where spatial topological information in the image domain is lost. This is in contrast with an unequivocal fact: human chromatic and achromatic visions are nearly as affective in scene-from-image representation. It means that spatial information, either topological (e.g., adjacency, inclusion, etc.) or non-topological (e.g., metric distance, angle measure, etc.), dominates color information in both the 4D scene domain and the (2D) image domain (Matsuyama and Hwang, 1990). Unlike traditional 1D image interpretation approaches, such as EOLib’s, topology preserving mapping of convolutional local filters into 2D retinotopic maps explains the increasing popularity of deep convolutional neural networks in computer vision applications (Cimpoi et al, 2014).

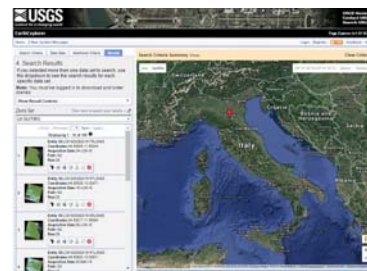


Figure 1. Existing EO CBIR systems support image retrieval by metadata text information, summary statistics or image example. No SCBIR capability is supported in operating mode.

Our conjecture is that existing EO CBIR systems support no semantic querying because they lack EO image understanding

capabilities, where spatial information dominates color information (Matsuyama and Hwang, 1990). This conjecture has two corollaries. First, it implies that a pre-condition of SCBIR is the computational solution of the multi-disciplinary cognitive problem of image interpretation (vision), i.e., the complexity of the SCBIR problem is not inferior to the complexity of vision. Vision is an inherently ill-posed problem (Matsuyama and Hwang, 1990) in the Hadamard sense, i.e., it is NP-hard. To infer stable symbolic percepts in the 4D spatiotemporal scene domain from ever-varying sub-symbolic sensations (quantitative sensory data) in the (2D) image domain, vision is affected by a semantic information gap and by a 4D-to-2D data dimensionality reduction problem.

The second corollary of our conjecture about an ongoing lack of SCBIR solutions is that existing EO image understanding systems (EO-IUSs) score “low” in operating mode, i.e., they fall short in transforming multi-source EO big data into comprehensive, timely and operational information products. If this thesis holds true, then existing EO-IUSs do not comply with the visionary goal of the Group on Earth Observation (GEO), formulated in the Global Earth Observation System of Systems (GEOSS) implementation plan for years 2005-2015 and in the QA4EO *Cal/Val* guidelines (Group on Earth Observation, 2010). The thesis that existing EO-IUSs score “low” in operating mode is supported by the fact that no EO data-derived Level 2 prototype product has ever been generated systematically at the ground segment (European Space Agency, 2015). By definition an EO Level 2 product includes an EO image corrected for atmospheric, adjacency and topographic effects, and a general-purpose user- and application-independent scene classification map (SCM).

To contribute toward filling an ongoing information gap from EO big data to EO value-adding products and services, an innovative EO Image Understanding for Semantic Querying (EO-IU4SQ) system is presented as a proof-of-concept, where automated multi-source EO big data spatiotemporal analytics is accomplished as a pre-condition for SCBIR, see Figure 2. In particular the EO-IU4SQ system consists of two *hybrid* (combined deductive and inductive) inference subsystems provided with feedback loops: an EO-IU subsystem is followed by an EO-SQ subsystem. The former transforms automatically (i.e., without user interaction) and in near real-time any multi-source EO multi-spectral (MS) image, provided with a radiometric calibration metadata file in agreement with the QA4EO *Cal/Val* requirements, into an EO Level 2 product consisting of: (a) an enhanced EO image corrected for atmospheric and topographic effects, and (b) an SCM whose legend includes quality layers, such as cloud and cloud-shadow, in addition to the general-purpose user- and application-independent 8-class land cover (LC) taxonomy adopted by the initial Dichotomous Phase (DP) of the two-phase Food and Agriculture Organization of the United Nations (FAO) - Land Cover Classification System (LCCS) (Di Gregorio and Jansen, 2000). The 8-class LCCS-DP taxonomy consists of the three “nested” dichotomous layers: (i) vegetation/non-vegetation, (ii) terrestrial/aquatic and (iii) managed/natural. The three DP layers deliver as output the following eight LCCS-DP classes. (1) Cultivated and Managed Terrestrial (non-aquatic) Non-vegetated Areas. (2) Natural and Semi-Natural Terrestrial Vegetation. (3) Cultivated Aquatic or Regularly Flooded Areas. (4) Natural and Semi-Natural Aquatic or Regularly Flooded Vegetation. (5) Artificial Surfaces and Associated Areas. (6) Bare Areas. (7) Artificial Waterbodies, Snow and Ice. (8) Natural Waterbodies, Snow and Ice. The general-purpose 8-class LCCS-DP legend is preliminary to the LCCS Modular

Hierarchical Phase (MHP), consisting of a hierarchical battery of application- and user-specific one-class LC classifiers.

Unlike alternative hierarchical LC class taxonomies whose Level 1 is multi-class, such as the CORINE Land Cover (CLC) and the EOLib’s LC map legend (Dumitru et al., 2015), the two-phase LCCS-DP and -MHP taxonomies are “nested” end-to-end, starting from the first-level LCCS-DP layer vegetation/non-vegetation, whose quality assurance becomes paramount according to an intuitive garbage in garbage out (GIGO) quality principle. To date EO-IUSs typically detect vegetation by inspecting a vegetation spectral index and this spectral index analysis, either single-date or multi-temporal, is typically conducted pixel-based. Conceptually a spectral index is a scalar value equivalent to the slope of one tangent to a multi-spectral (MS) signature; no MS signature characterized by a spectral shape and intensity can be parameterized by one or more angular coefficient in one point without an information loss in both shape and intensity. In practice, an n-channel MS signature is projected onto a scalar variable, equivalent to a panchromatic image affected by an nD-to-1D data dimensionality reduction. In addition, as a special case of 1D image analysis pixel-based image analysis completely ignores both non-topological and topological spatial information in the image domain. The conclusion is that existing EO-IUSs are likely to fall short at level 1 of the 3-level LCCS-DP taxonomy.

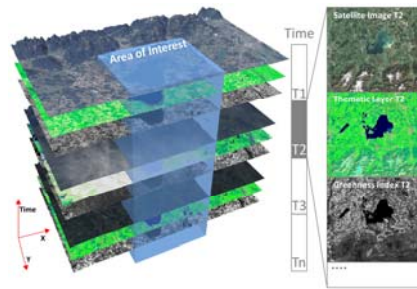


Figure 2. The EO-IU4SQ system for spatiotemporal analytics of multi-source EO big data. Each EO image is automatically provided with numeric/quantitative and categorical/qualitative information layers.

To accomplish multi-source EO big data transformation into Level 2 products the EO-IU4SQ system complies with the object-based image analysis (OBIA) paradigm (Blaschke et al., 2014), where spatial information, either topological or non-topological, dominates color information. In addition the two EO-IU and EO-SQ subsystems adopt a *hybrid* inference approach provided with feedback loops, where statistical/ bottom-up/ inductive/ machine learning-from-data algorithms and physical/ top-down/ deductive/ human-to-machine knowledge transfer approaches are combined to take advantage of the complementary features of each and overcome their shortcomings (Baraldi and Boschetti, 2012). These properties make the EO-IU4SQ system architecture alternative to the design adopted by the majority of EO-IUSs systems, including EOLib’s, consisting of a feedforward inductive inference system architecture.

For the sake of brevity, in the rest of this paper the proposed EO-IU4SQ system is described at the two levels of understanding of architecture (design) and knowledge/information representation, whereas the two levels of system understanding of algorithms and implementation are merely outlined, see Figure 3. According to Marr, the linchpin of success of any information processing system is system

design and knowledge/information representation, rather than algorithms and implementation (Marr, 1982).

2. EO-IU4SQ SYSTEM DESIGN AND INFORMATION REPRESENTATION

2.1 System design

The EO-IU4SQ system architecture encompasses a hybrid feedback EO-IU subsystem preliminary to a hybrid feedback EO-SQ subsystem, see Figure 3 and Figure 4. These two subsystems share: (i) a fact base of EO data and data-derived products, (ii) a knowledge base available in addition to facts, and (iii) an inference engine, which applies the knowledge base to the fact base to infer new information, such as EO Level 2 products.

2.2 Knowledge/information representation

The EO-IU4SQ system's knowledge base consists of four types of knowledge, see Figure 3.

(1) 4D spatiotemporal scene-domain knowledge in user-speak, known as world model. In the 4D scene domain, spatial information typically dominates color information (Matsuyama and Hwang, 1990). Represented as a semantic network in a graphic user interface (GUI) (Grove, 1999), the world model can be thought of as an entity-relationship (ER) conceptual model where classes of real-world 4D objects, represented as nodes, are provided with attributes, either spatiotemporal or not, including appearance properties, while inter-class relationships, represented as arcs, are either spatiotemporal or not, such as part-of and subset-of (Sonka et al., 1994; Laurini and Thompson, 1992). In EO data applications, the world model consist of land cover (LC) classes belonging to a discrete and finite LC class taxonomy to be community-agreed upon in advance, such as the FAO-LCCS taxonomy (Di Gregorio and L. Jansen, 2000). To date grammars (syntactic models) can rarely be inferred from supervised (labeled) or unsupervised (unlabeled) data by statistical approaches. Hence, grammars typically require significant human interaction, e.g., to identify non-spatial inter-class relationships including part-of and subset-of (Sonka et al., 1994). In addition, "no amount of syntax will ever produce semantics" (Mayo, 2003). Hence, the world model is expected to be passed on to the EO-IU4SQ system by human domain-experts. This is a top-down scene-domain knowledge transfer, from human-to-machine. Typically investigated by artificial intelligence (Laurini and Thompson, 1992), top-down knowledge transfer is the dual problem of (complementary not alternative to) inductive machine learning-from-data, which is currently dominating the RS literature.

(2) Image-domain knowledge in techno-speak, accounting for quantitative visual features such as color, texture (perceptual spatial grouping of texture elements), local shape and inter-object spatial relationships, either topological such as inclusion, adjacency, etc., or non-topological, such as metric distances and inter-angle measures (Smeulders et al., 2000; Matsuyama and Hwang, 1990; Sonka et al., 1994). In the (2D) image domain, spatial information, i.e., local shape, texture and inter-object spatial relationships, typically dominates color information.

(3) Knowledge about the sensor transfer functions (translation rules) capable of mapping the scene-domain knowledge in user-speak (e.g., in meter units) onto the image-domain knowledge in techno-speak (e.g., in pixel units).

(4) Standard rule base consisting of physical laws, methods, processes, first-principle models, if-then rules, etc., to be employed by the EO-IU4SQ inference engine to generate new

information about the observed 4D scene domain from the existing fact base of EO images and products.

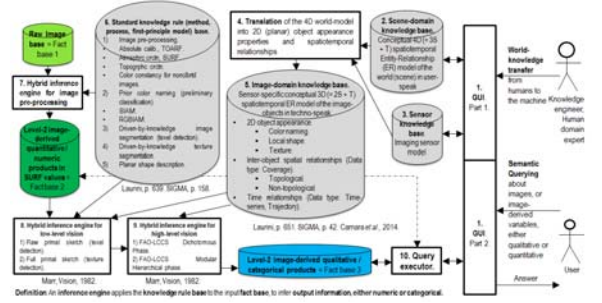


Figure 3. EO-IU4SQ system architecture.

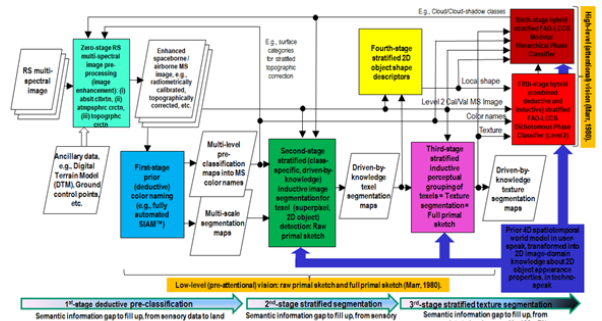


Figure 4. Hybrid feedback EO-IU subsystem based on a convergence-of-evidence approach to low- and high-level EO image understanding (classification). Visual evidence stems from color, local shape, texture and inter-object spatial relationships. Low-level EO image classification is general-purpose, user- and application independent, such as an EO Level 2 SCM product provided with an 8-class LCCS-DP map legend. High-level EO image classification is user- and application-specific, such as an EO SCM product provided with an LCCS-MHP legend.

2.3 Standard rule base

Well-known visual features to be investigated by standard computer vision algorithms are color, local shape, texture and inter-object spatial relationships (Matsuyama and Hwang, 1990; Smeulders et al., 2000; Sonka et al., 1994). Visual attributes in the image domain should never be confused with materials (surface types) in the 4D scene domain. For example, MS color names, just like spectral endmembers in hyper-spectral images, cannot always be inverted to unique LC class names (Baraldi and Boschetti, 2012). According to a convergence-of-evidence computer vision approach it is the combination of visual attributes, with special emphasis on spatial information, eligible for identifying a target material in the scene domain. To support a convergence-of-evidence vision approach, the EO-IU4SQ's standard rule base encompasses algorithms for EO data enhancement (pre-processing) and pre-attentional (low-level) vision. In agreement with the Marr terminology (Marr, 1982) pre-attentive vision consists of two phases: a raw primal sketch for zero-crossing (ZX) segment and texel (texture element) detection followed by a full primal sketch for texture segmentation (perceptual spatial grouping of texels). These standard image pre-processing and low-level vision algorithms are summarized below.

(1) EO image pre-processing (EO data enhancement).

(i) Absolute radiometric *Cal*, in compliance with the QA4EO *Cal* requirements. A battery of sensor-specific EO image radiometric calibrators, e.g., AVHRR, SPOT-1 to SPOT-7, Pleiades, Landsat-4 to Landsat-8, MeteoSat, RapidEye, WorldView, QuickBird, Ikonos, etc., is required to transform digital numbers into a radiometric unit of measure, specifically, top-of-atmosphere reflectance (TOARF). Sensory data *Cal* is a well-known “prerequisite for physical model-based analysis of airborne and satellite sensor measurements in the optical domain” (Schaeppman-Strub et al., 2006). In other words, EO data *Cal* is a necessary not sufficient condition for both physical model-based and hybrid inference-based data understanding systems, including the hybrid EO-IU4SQ system. Whereas physical variables can be investigated by physical, statistical and hybrid inference systems, uncalibrated sensory data can be investigated by statistical data models exclusively. Although they do not require physical variables as input, statistical models can benefit from physical variables, since physical units of measure harmonize data acquired across time, space and sensors. Irrespective of this unquestionable benefit, in the RS literature, where statistical models dominate physical ones, EO data *Cal* is largely overlooked, in disagreement with the QA4EO *Cal* requirements.

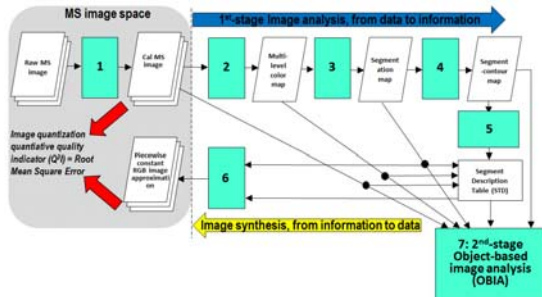


Figure 5. The SIAM software toolbox for prior knowledge-based vector quantization, VQ. Phase 1-of-2 = Encoding phase/Image analysis - Stage 1: MS data calibration into TOARF or SURF values. Stage 2: Prior knowledge-based SIAM color space partitioning. Stage 3: Well-posed two-pass connected-component multi-level image labeling. Stage 4: Well-posed segment-contour extraction. Stage 5: Well-posed Segment Description Table (STD) allocation and initialization. Phase 2-of-2 = Decoding phase/Image synthesis - Stage 6: Piecewise constant MS image approximation (“Object-mean view”) and per-pixel VQ error estimation. Stage 7: Object-based image analysis (OBIA).

(ii) “Stratified” (“layered”, class-conditional) approach to atmospheric correction of TOARF into surface reflectance (SURF) values. Atmospheric correction is a typically ill-posed physical model-based inversion problem. To become better conditioned for numerical treatment, an atmospheric correction algorithm can be run on class-conditional EO data layers (Vermote and Saleous, 2007), such as MS color names detected by a static decision tree for MS reflectance space polyhedralization (Baraldi and Boschetti, 2012).

(iii) “Stratified” topographic correction (TOC). TOC is a typical chicken-and-egg dilemma. To become better posed for automated numerical solution, an inherently ill-conditioned TOC algorithm is run on informative EO image strata. These image masks comprise MS color names, automatically detected by a deductive (prior knowledge-based) MS reflectance space polyhedralization such as the Satellite Image Automatic

Mapper (SIAM) (Baraldi and Boschetti, 2012), see Figure 5, overlapped with illumination-specific information layers inferred from ancillary data sources, specifically, a digital surface model (DSM) and the image metadata sun and sensor positions. Illumination-specific image masks are threefold: horizontal areas where no TOC is required, slopes facing the sun and slopes facing away from the sun. In the latter two illumination-specific layers, a stratified TOC is applied per MS color name (Baraldi et al., 2010).

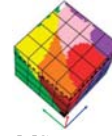


Figure 6. Unlike a MS space polyhedralization impossible to visualize when the number of channels is superior to three, an RGB data cube polyhedralization is intuitive to display. For example, a “universal” inventory of eleven basic color (BC) names in 20 human languages (black, white, gray, red, orange, yellow, green, blue, purple, pink and brown) can be mapped onto the RGB cube based on perceptual evidence (Griffin, 2006).



Figure 7. Left: Sentinel-2A (S2A) image of Austria, acquired on 2015-08-18, 10 m resolution, radiometrically calibrated into TOARF values and depicted in true colors: R = visible red channel, G = visible red, B = visible blue. No histogram stretching for visualization purposes. Right: Automatic SIAM mapping of the S2A image onto a legend of MS color names (spectral categories), depicted as pseudo-colors (green as vegetation, blue as water or shadow, etc.).

(iv) Color constancy in non-calibrated images. In human vision, color constancy ensures that the perceived color of objects remains relatively constant under varying illumination conditions, so that they appear identical to a “canonical” (reference) image, subject to a “canonical” (known) light source of controlled quality, such as a white light source (Gijsenij et al., 2010). In short, solution of the color constancy problem is the recovery of an illuminant-independent representation of the reflectance values in a scene. In practice color constancy supports image harmonization and interoperability when no radiometric calibration parameter is available. A novel self-organizing statistical algorithm for image color constancy (patent pending), inspired by human vision, is typically applied to MS images provided with no radiometric calibration metadata file, such as those acquired by consumer-level RGB color cameras, either true- or false-color, mounted onboard Unmanned Aerial Vehicles (UAVs) (Vo et al., 2016).

(2) Pre-attentive vision: Raw and full primal sketch.

(i) Raw primal sketch: Prior knowledge-based MS reflectance space polyhedralization into color names. When a numeric/quantitative color variable is mapped onto a categorical/nominal color variable, color attributes of target LC classes can be called by name, in line with symbolic human

reasoning. Proposed to the RS community in recent years (Baraldi and Boschetti, 2012), the SIAM software product is an expert system (prior knowledge-based decision tree) for deductive/top-down vector quantization (VQ) and VQ quality assessment in a MS color space radiometrically calibrated into TOARF or SURF values, see Figure 5. Since it relies exclusively on *a priori* knowledge available in addition to data, any expert system, such as SIAM, is fully automated, i.e., it requires neither user-defined parameters nor training data to run. In the VQ encoding phase, SIAM partitions the MS reflectance space into static (non-adaptive to data) hyperpolyhedra, not necessarily convex or connected, equivalent to a dictionary (codebook) of MS color names (codewords). Each MS pixel is mapped onto one MS polyhedron associated with a color name. Unfortunately, when the MS space dimensionality is greater than three, a prior dictionary of mutually exclusive and totally exhaustive hyperpolyhedra is difficult to think of and impossible to visualize, see Figure 6. Next, for image analysis purposes a 2D multi-level VQ map is automatically generated, see Figure 7, together with its segmentation map, see Figure 8. In the VQ decoding phase, equivalent to image synthesis, a VQ quality assessment is accomplished in compliance with the QA4EO *Val* requirements, via a segmentwise-constant MS image reconstruction, see Figure 5. In VQ problems, a typical community-agreed Q^2I is the per-vector encoding-decoding Euclidean distance (Cherkassky and Mulier, 1998). According to the existing literature (Baraldi et al., 2016), SIAM provides a very substantial contribution to the detection of the first two-of-three levels of the LCCS-DP taxonomy, namely, vegetation/non-vegetation and terrestrial/aquatic, based on context-insensitive spectral properties exclusively. To improve LC classification accuracy, reliability and informativeness, preliminary spectral-based classification by SIAM must be combined with additional evidence stemming from spatial information according to a convergence-of-visual-evidence approach, refer to Figure 4.

(ii) Raw primal sketch: Prior knowledge-based RGB cube polyhedralization into color names, see Figure 6. The RGB Image Automatic Mapper (RGBIAM, patent pending) is a novel expert system capable of partitioning an RGB data cube, either true- or false-color, into a pre-defined dictionary of RGB color names. The RGBIAM software design is the same as SIAM's, see Figure 5. When a color space dimensionality is equal to three, a prior dictionary of mutually exclusive and totally exhaustive polyhedra, either convex or not, either connected or not, is intuitive to think of and easy to visualize, see Figure 6. The novel RGBIAM's static decision tree requires as input an RGB image subject to color constancy for data harmonization when no radiometric *Cal* parameter is available, in compliance with the QA4EO *Cal* guidelines.

(iii) Raw primal sketch: Prior knowledge-based MS reflectance space polyhedralization into color names through time. Any post-classification LC change/no-change (LCC) detection overall accuracy, $OA-LCC_{1,2} \in [0, 1]$, is such that $OA-LCC_{1,2} \leq (OA-LC_1 \times OA-LC_2)$, where $OA-LC_1$ and $OA-LC_2$ are the two single-date OAs. For example, if $OA-LC_1 = 0.90$, $OA-LC_2 = 0.90$, then $OA-LCC_{1,2} \leq 0.81$. It means that, in general, post-classification LC change/no-change detection-through-time is recommended if and only if single-date thematic map accuracies score "high". When two SIAM maps of the same geographic area of interest are available at the same spatial resolution at times $T1$ and $T2 > T1$, post-classification through time is straightforward according to a two-way square contingency table (association matrix, cross-tabulation matrix)

to be community-agreed upon, whose two input categorical variables at times $T1$ and $T2$ are the same SIAM legend.

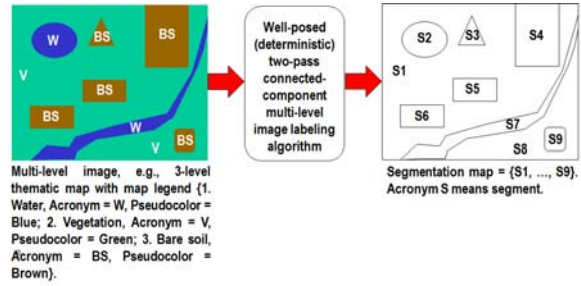


Figure 8. One segmentation map is deterministically generated from one multi-level image, such as a thematic map, but the vice versa does not hold, i.e., many multi-level images can generate the same segmentation map. In this example, nine image-objects/segments S1 to S9 can be detected in the thematic map. Each segment consists of a connected set of pixels sharing the same thematic map label. Each stratum/layer/level consists of one or more segments, e.g., stratum Vegetation (V) consists of the two disjoint segments S1 and S8. In any multi-level image domain, three labeled spatial primitives co-exist and are provided with parent-child relationships: pixel with level label, segment with level label and stratum with level label. This overcomes the ill-fated dichotomy between traditional pixel-based versus segment-based image analysis (OBIA) in the numeric image domain.

(3) Raw primal sketch: Two-pass connected-component multi-level image labelling, see Figure 8. Any SCM whose legend is a discrete and finite dictionary of LC class labels is a multi-level image (Sonka et al., 1994). As such, it can be partitioned into planar spatial units (image-objects), either (0D) pixel, (1D) line, (2D) polygon or multi-part image-object (Open Geospatial Consortium, 2015), where each image-object is a connected set of pixels featuring the same LC class label, by a well-posed (deterministic) two-pass connected component multi-level image labeling algorithm (Sonka et al., 1994). Each SCM-derived planar object in the image domain is provided with both a segment identifier and an LC class label in the SCM's taxonomy, in compliance with the increasingly popular object-based image analysis (OBIA) paradigm (Blaschke et al., 2014).

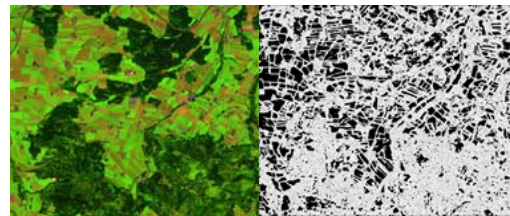


Figure 9. Left: Sentinel-2A image of Austria, radiometrically calibrated into TOARF values, depicted in false colors (R = MIR, G = NIR, B = Visible Blue), 10 m resolution. Right: Automated image segmentation into ZX segments. Segment contours automatically depicted with 8-adjacency cross-aura values in range $\{0, 8\}$ (Baraldi and Boschetti, 2012).

(4) Raw primal sketch: Consistent with the Mach band illusion (Pessoa, 1996), novel physical model-based automated ZX contour-pixel detection and ZX segment detection. If we

require that a computer vision system should be able to predict perceptual effects, such as the well-known Mach bands illusion where bright and dark bands are seen at ramp edges, then the number of published vision models becomes surprisingly small (Pessoa, 1996). The novel ZX contour-pixel detector consists of a multi-scale multi-orientation quasicomplete even-symmetric spatial wavelet transform, eligible for image analysis, coding and synthesis (Baraldi and Parmiggiani, 1996). In Marr's words an intermediate information primitive in the image domain, called ZX segment, was defined as "a piece of the contour whose intensity slope (rate at which the convolution changes across the segment) and local orientation are roughly uniform" (Marr, 1982, p. 60). Hence, it is at the level of detection of ZX segments that 0D ZX pixels turn into sub-symbolic discrete 2D image-objects (polygons). To automatically extract 2D ZX segments from 0D ZX pixels, an original physical model-based algorithm, featuring no free-parameter to be user-defined, was implemented in operating mode, see Figure 9.

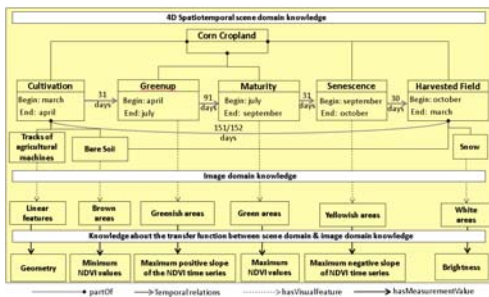


Figure 10. Semantic network of a real-world object with a cyclic behavior, such as a corn cropland in the northern hemisphere.

(5) Full primal sketch: “stratified” texture segmentation. A numeric/quantitative texture variable (perceptual spatial grouping of texels) is mapped onto a categorical/nominal texture variable, suitable for human symbolic reasoning. A multi-scale texture profile of the spatial distribution of ZX contour-pixels is accomplished in line with (Nagao and Matsuyama, 1980). For each spatial scale, move an $N \times N$ window over the picture of binary ZX contour-pixels. If the window contains more than $\sqrt{2} \times N$ binary ZX contour-pixels, mark the central point of the window as “high” texture.

(6) Planar shape description. Given an image segmentation map consisting of planar objects, see Figure 8, each planar object can be described in geometric (shape) and size terms by a minimally dependent and maximally informative (mDMI) set of planar shape and size indexes, such as area, characteristic spatial scale, scale-invariant roundness, convexity, rectangularity, elongatedness, straightness of boundaries and simple connectivity (Soares et al., 2014). The implemented mDMI set of planar shape indexes is alternative to existing libraries of planar shape descriptors available in commercial EO image processing software toolboxes, provided with no quality assessment in terms of feature independence and informativeness.

2.4 GUI

The EO-SQ subsystem’s GUI supports the following human-machine interactions.

(1) Scene-domain knowledge transfer from human-to-machine. The world model can be graphically represented as a semantic network with LC classes as nodes and inter-class relationships,

including events, as arcs between nodes (Growe, 1999). Observations of the 4D real world are discrete n-tuples (space x , y and z , time t ; “theme”, plus other numeric or categorical attributes, e.g., weight, size, etc.), where the 4-tuple (x, y, z, t) is the location in space and time of the observation, while attribute “theme” identifies the real-world phenomenon or object being observed (Reis Ferreira et al., 2014). Hence, “theme” may account for semantics involved with the observed object or phenomenon. All possible combinations of attributes (space, time, theme), can be modeled as three data types, called time series, coverage, and trajectory, where one attribute is measured, the second is fixed and the third is controlled. According to Reis Ferreira et al. (2014), a time series represents the measured variations of a theme over a controlled time in a fixed location; a trajectory measures locations of a fixed theme over a controlled time and a coverage measures attribute theme within a controlled spatial extent at a fixed time. A world model is an ontology of real-world geospatiotemporal objects/continuants and events/occurrences derived from the three spatiotemporal data types time series, coverage and trajectory. A real-world object/continuant (e.g., a car) is an identifiable (discrete) entity, i.e., it is provided with a unique “identity”, which remains constant during its lifetime, while its attributes, whether spatiotemporal or not, including semantics or “theme”, can change during its lifetime. An object is present as a whole unit at each moment of its existence. In EO applications, a real-world object can be: (i) a periodic object whose identity is fixed while its attributes change with a cyclic behavior, i.e., the object’s identity comprises a given sequence of different states in a fixed time periodicity, see Figure 10, or (ii) persistent/non-periodic objects, e.g. forest area. An event is an individual episode with a definite beginning and an end. It only exists as a whole across the interval over which it occurs, either instant or durative. An event does not change over time. While an event can involve one or more objects, the same object can be involved in any number of events. Events can be: (i) instant, or (ii) durative, including short-term transition events and slowly transient events.

(2) Graphic selection of existing semantic queries/decision rules or generation of new semantic queries/decision rules, in compliance with the world model. There are two types of queries. (i) To accomplish SCBIR operations, where the fact base is investigated for EO image retrieval purposes. e.g., retrieve EO images that are cloud-free across the selected geographic area of interest (AOI). (ii) To infer new information layers from the fact base, e.g., detect flooded areas as a post-classification time-series analysis of single-date thematic maps.

2.5 Array fact base

To accomplish efficient geospatial data querying and analysis through space and time within a user-defined AOI and a target time interval, the EO-IU4SQ system prototypical implementation stores its fact base, consisting of multi-sensor multi-temporal EO images, e.g., Landsat-5/7/8 and Sentinel-2A images, each one provided with its information products, either numeric or categorical, in an array database implemented as the Rasdaman (Baumann et. al, 2015). An array database instantiates multiple spatiotemporal data cubes in compliance with the Open Geospatial Consortium (OGC) standards (OGC, 2015), to guarantee inter-system harmonization and compatibility. In a spatiotemporal data cube the third dimension is time, defined in eXtensible Markup Language (XML) as a 1D temporal coordinate system accessible Unified Resource Identifiers (URIs). Time overlays the 2D spatial coordinate

system, specified by European Petroleum Survey Group (EPSG) codes defined in XML using URIs. Similar to standard relational databases investigated by the Structured Query Language (SQL), an array database consisting of data cubes can be queried by a declarative query language. In an array database storage-related characteristics, such as indexing, tiling and horizontal scaling, can be investigated and optimized independently of application logic. In addition, the data cube model has been proven to be scalable and reliable in operational applications (Baumann et. al, 2015). Based on these considerations it was selected as storage backend in the EO-IU4SQ system prototypical implementation.

Once users successfully create and store semantic queries in the web-based query interface, these queries are published as OGC compliant Web Processing Services (WPSs) and become available within the client-server architecture, to be executed by any WPS client remotely on a single server or server cluster, see Figure 11. This client-server architecture, providing EO data processing capabilities together with the “ready-to-analyze” data, guarantees fast response to queries.

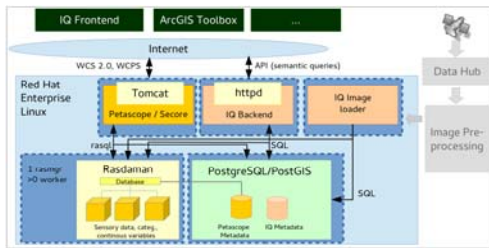


Figure 11. Web-based Image Querying (IQ) system. IQ can be intended as a synonym of EO-IU4SQ (Sudmanns et. al, 2016).

2.6 Hybrid inference engine

The EO-IU4SQ’s hybrid inference engine combines the rule base and the fact base to infer new information products. It consists of four inference subsystems.

(i) Transformation of space domain knowledge into image domain knowledge through the sensor transfer functions. It provides: (a) an estimate of the EO imaging sensor’s physical properties, including spatial, spectral and temporal resolutions, considered necessary to intercept the target LC phenomenon. (b) An estimate of the 2D visual attributes, such as categorical color names, numeric spectral index values, planar shape descriptors, texture values and spatial relationships, mapped from each target LC class described in the 4D world model, see Figure 10.

(ii) Automated EO image pre-processing (enhancement), e.g., absolute radiometric *Cal*, stratified atmospheric correction, stratified TOC, etc., refer to Section 2.3.

(iii) Low-level (pre-attentional) EO image processing, either raw primal sketch, e.g., texels detected as ZX segments, or full primal sketch, e.g., texture segmentation as perceptual grouping of texels (Marr, 1982; Sonka et al., 1994), refer to Section 2.3.

(iv) High-level (attentional) EO image interpretation (classification). It is split into two: a preliminary Level 2 cloud/cloud-shadow detection (Baraldi and Tiede, 2015) plus application-independent general-purpose 8-class LCCS-DP classification, see Figure 12, followed by an LCCS-MHP classification phase, consisting of a hierarchical battery of application- and user-specific one-class LC classifiers. In the two-phase LCCS classification, each final LC classifier adopts a nested (hierarchical, stratified) convergence-of-visual-evidence

OBIA approach, where independent sources of visual evidence are pixel-based color properties in addition to object-based spatial properties, specifically, local shape, texture and spatial relationships, either topological or non-topological (Matsuyama and Hwang, 1990).

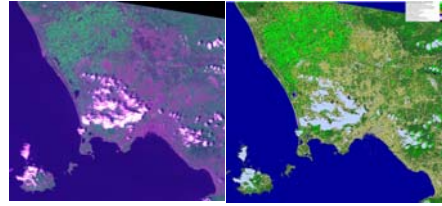


Figure 12. Left: 4-band (B, G, R, NIR) ALOS AVNIR-2 image of Campania, Italy, radiometrically calibrated into TOARF values and depicted in false colors (R = MIR, G = NIR, B = Visible Blue), 10 m resolution. No histogram stretching for visualization purposes. Right: Level 2 SCM product, whose legend comprises the 8-class LCCS-DP taxonomy, plus Cloud/Cloud-shadow quality layers.

3. RESULTS

In the EO-IU4SQ system prototype implemented as a proof-of-concept, a semantic spatiotemporal query is formulated by a user through the EO-SQ subsystem’s GUI to the fact base of EO images provided with derived products by the EO-IU subsystem. In this GUI the user selects (1) the AOI, (2) the target time period and creates or selects via graphical elements the decision-rule pipeline for semantic querying the fact base. Once it is executed, the query results are shown in the image domain and/or as summary statistics (3). The query output product can be downloaded as a geoTiff or re-used for further queries. In Figure 13 the GUI shows an AOI overlapping with the Hohe Tauern National Park (Austria), where a glacier extent invariant within a target time interval was inferred in near real-time from the fact base to exclude seasonal snow.

4. DISCUSSION

The QA4EO *Val* guidelines require each step in an EO data processing pipeline to be provided with community-agreed OP-Q²Is, featuring a degree of uncertainty in measurement, to assess the propagation of errors in comparison with reference standards. According to the GEO, in Stage 3 *Val* “spatial and temporal consistency of the product with similar products are evaluated by independent means over multiple locations and time periods representing global conditions. In Stage 4 validation, results for Stage 3 are systematically updated when new product versions are released and as the time-series expands” (Group on Earth Observation, 2010).

To comply with the QA4EO Stage 4 *Val* requirements, the EO-IU and EO-SQ subsystems must be provided with OP-Q²Is estimated by independent means on large-scale EO image time-series. This is a huge validation process still ongoing (Baraldi et al., 2016). Although it has not been subject to a Stage 4 *Val* process yet, the proposed EO-IU4SQ system implementation works as a proof-of-concept, where automated Level 2 product generation is accomplished as a pre-condition for semantic querying.

In future works, within the proposed EO-IU4SQ system architecture (Figure 3) any information processing block prototype will be considered eligible for permanent optimization, modification or replacement to improve the EO-IU or EO-SQ subsystem capabilities in operating mode.

Furthermore, the semantic network formalism required to graphically represent the world model will be augmented (Grove, 1999) and integrated with an algebra capable of describing spatiotemporal data types and operations in a language-independent and formal way (Reis Ferreira et al., 2014).

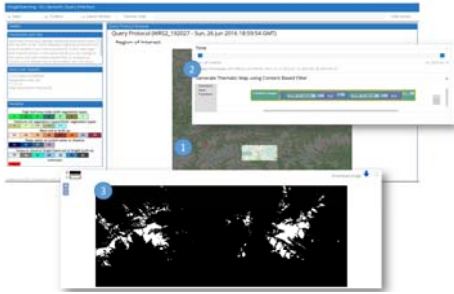


Figure 13. Semantic querying to infer new information layers from the fact base. In this example a glacier extent invariant within a time interval was inferred in near real-time from the fact base to exclude seasonal snow.

ACKNOWLEDGEMENTS

The test implementation of the EO-IU4SQ system, called Image Querying (IQ) system, was awarded 1st place in the Copernicus Masters 2015 - T-Systems Big Data Challenge. This study was supported by the Austrian Research Promotion Agency (FFG), in the frame of project AutoSentinel2/3, ID 848009.

REFERENCES

- Baraldi, A., Boschetti, L., 2012. Operational automatic remote sensing image understanding systems: Beyond geographic object-based and object-oriented image analysis - Part 2, *Remote Sens.*, 4, pp. 2768-2817.
- Baraldi, A., Tiede, D., 2015. *AutoCloud+*, ESA Invitation to tender - Next Generation EO-based Information Services, ESA/AO/1-8373/15/I-NB.
- Baraldi, A., Girona, M., Simonetti, D., 2010. Operational three-stage stratified topographic correction of spaceborne multi-spectral imagery, *IEEE Trans. Geosci. Remote Sensing*, 48(1), pp. 112-146.
- Baraldi, A., Humber, M., Tiede, D., Lang, S., 2016. Stage 4 validation of the Satellite Image Automatic Mapper software process and Level 2 products, *Int. J. of RS*, submitted.
- Baraldi, A., Parmiggiani, F., 1996. Combined detection of intensity and chromatic contours in color images, *Optical Eng.*, 35(5), pp. 1413-1439.
- Baumann, P., Mazzetti, P., Ungar, J., Barbera, R., Barboni, D., Beccati, A., Bigagli, L., 2015. Big data analytics for earth sciences, *Int. J. Digital Earth*, pp. 1-27.
- Blaschke, T., Hay, G., Kelly, M., Lang, S., Hofmann, P., Addink, E., Queiroz Feitosa, R., van der Meer, F., van der Werff, H., van Coillie, F., Tiede, D., 2014. Geographic object-based image analysis - towards a new paradigm," *ISPRS J. Photogram. Remote Sens.*, 87, pp. 180-191.
- Cherkassky, V., Mulier, F., 1998. *Learning from Data: Concepts, Theory, and Methods*. New York, NY: Wiley.
- Cimpoi M., Maji S., Kokkinos I., Vedaldi A., 2014. Deep filter banks for texture recognition, description, and segmentation, *CoRR*, abs/1411.6836.
- Di Gregorio, A., Jansen, L., 2000. *Land Cover Classification System (LCCS)*, FAO Corporate Document Repository.
- Dumitru, C., Cui, S., Schwarz, G., Datcu, M., 2015. Information content of very-high-resolution SAR images: Semantics, geospatial context, and ontologies," *IEEE J. Selected Topics Applied Earth Obs. Remote Sens.*, 8(4), pp. 1635-1650.
- European Space Agency, 2015. *Sentinel-2 User Handbook*, Standard Document, Issue 1 Rev 2.
- Reis Ferreira, K., Camara, G., Vieira Monteiro, A., 2014. An algebra for spatiotemporal data: From observations to events, *Trans. in GIS*, 1(2), pp. 253-269.
- Gijssenij, A., Gevers, T., van de Weijer, J., 2010. Computational color constancy: Survey and experiments, *IEEE Trans. Image Proc.*, 20(9), pp. 2475-2489.
- Griffin, L., 2006. Optimality of the basic color categories for classification, *J. R. Soc. Interface*, 3, pp. 71-85.
- Group on Earth Observation/Committee on Earth Observation Satellites (GEO/CEOS), 2010. *A Quality Assurance Framework for Earth Observation, version 4.0*.
- Grove, S., 1999. Knowledge-based interpretation of multisensor and multitemporal remote sensing images, *Int. Archives of Photogram. Remote Sens.*, 32, Part 7-4-3 W6.
- Laurini, R., Thompson, D. 1992. *Fundamentals of Spatial Information Systems*. London, UK: Academic Press.
- Marr, D., 1982. *Vision*. New York, NY: Freeman and C.
- Matsuyama, T., Hwang, V., 1990. *SIGMA - A Knowledge-based Aerial Image Understanding System*. New York, NY: Plenum Press.
- Mayo, M., 2003. Symbol grounding and its implications for artificial intelligence, 26th Australian Computer Science Conference, Adelaide, Australia, 16.
- Nagao, M., Matsuyama, T., 1980. *A Structural Analysis of Complex Aerial Photographs*. Plenum Press, New York.
- Open Geospatial Consortium (OGC) Inc., 2015. OpenGIS® Implementation Standard for Geographic information - Part 1: Common architecture.
- Pessoa, L., 1996. Mach Bands: How many models are possible? *Vision Res.*, 36(19), pp. 3205-3227.
- Smeulders, A., Worring, M., Santini, S., Gupta, A., Jain, R., 2000. Content-based image retrieval at the end of the early years, *IEEE Trans. Pattern Anal. Machine Intell.*, 22(12), pp. 1349-1380.
- Soares, J., Baraldi, A., Jacobs, D., 2014. Segment-based simple-connectivity measure design and implementation, *Tech. Rep.*, University of Maryland, College Park.
- Sonka, M., Hlavac, V., Boyle, R., 1994. *Image Processing, Analysis and Machine Vision*, London, U.K.: Chapman & Hall.
- Vermote, E., Saleous, N., 2007. *LEDAPS surface reflectance product description - Version 2.0*, University of Maryland, College Park, Dept Geography, Code 614.5.
- Sudmanns, M., Tiede, D., Augsten, N., Baraldi, A., Belgiu, M., Lang, S., 2016. Array-Datenbanken für semantische inhaltsbasierte Suche und Analyse in Satellitenbildarchiven. Dreiländertagung der SGPF, DFPF und OVG. Bern, Switzerland.
- Vo, A., Truong-Hong, L., Laefer, D., Tiede, D., d'Oleire-Oltmanns, S., Baraldi, A., Shimoni, M., 2016. Outcome of the 2015 IEEE GRSS Data Fusion Contest. Part-B: 3D contest," *IEEE J. Selected Topics Applied Earth Obs. Remote Sens.*, X, pp. YYY-ZZZ.