In this paper a new very high resolution satellite and aerial image analysis system is presented. It relies on a polychotomy scheme of a tree-based representation of the image space that separates the image content into a rich set of lower level semantic layers. Aiming at minimizing the manual fine tuning of the decomposition, the system supports a user interaction module with which, positive and negative samples of the targeted features can be marked. The set of samples is used to train a statistical learning algorithm from which a classification rule is derived. Enforcing the rule on the underlying image representation structure results in a drastic simplification of the input scene which is set to contain all positively identified structures. The process can be re-iterated in near real-time for the analysis of massive data-sets. This makes it a valuable instrument for crisis management and examples are given.

I. INTRODUCTION

Disaster risk detection and monitoring constitute a key element in the cycle of operations to support crisis management. With time being a parameter of critical importance in most crisis scenarios, early and accurate detection of factors that may lead to disasters contributes in risk reduction and improved preparedness and response in the unfolding event horizon. In the post-disaster time scale, monitoring the progression of events allows the dynamic reconfiguration of emergency countering strategies, and the registration of data to be used in operation planning, for conducting post-disaster needs assessments (PDNA) in support of recovery and reconstruction planning, and other relevant activities.

Advances in remote sensing technology provide the means for both accurate detection and monitoring of events. In particular, increased sensor resolution and quality on-board satellite and airborne systems offers access to sensitive details characterizing structures of interest and other targeted assets. A challenge that remains however is the effective analysis, processing and management of the bulk of information generated. In this field, two critical elements can be identified; the design and utilization of efficient algorithms, and the capacity of existing computer systems to handle the vast amount of information. The use of dedicated computer systems allows for a certain degree of confidence in the ability to facilitate mass processing modules. Yet, it is the algorithmic architecture that guarantees their effectiveness in compliance with the time requirement.

Algorithms for rapid and robust image and signal analysis often rely on some structured input representation. Operational examples are the Max-Tree, the Component-Tree, the Alpha-Tree and others. Utilizing them in

---

1 All authors have the same affiliation
the context of crisis management requires some key adaptations, primarily concerning the user interaction. Any such algorithm typically produces a dichotomy on the input data based on some threshold value on the selected descriptors. This is an iterative procedure introducing two major bottlenecks; the stepwise selection of the most appropriate attribute, describing best the targeted features, and the fine-tuning of the respective threshold parameters. In the approach presented, both issues are resolved by introducing a polychotomy scheme that relies on a set of generic parameters from which a large pool of feature attributes may be computed. Aiming at minimizing the extent of user interaction, the proposed system operates a set of self-calibrating routines based on experts’ knowledge. The latter is captured through regular hardware input devices, e.g. mouse, or more advanced human-machine interfaces such as visual trackers, and others. The learned patterns are projected on some compact multi-dimensional feature space from which decision rules are derived automatically. Harvesting these rules allows for rapid and automated target extraction and interpretation to higher level semantics.

This modular architecture allows for switching between the core processing modules in order to manage each scenario in a custom and mostly effective manner. It is referred to as the process engine, and constitutes a key element of the Image Query (IQ) system developed and operated by the Joint Research Centre (JRC) of the European Commission. The process engine work-flow is demonstrated in the computation of sample tiles of an advanced, medium-resolution global human settlement layer based on SAR data, and target extraction in real crisis scenarios. This is followed by a discussion on the system performance.

II. STRUCTURED IMAGE REPRESENTATION

The process engine operates on images sourced in by the embedded database of the IQ system. Issues related to geo-coding, ortho-rectification, pan-sharpening, and in general to data preparation are handled by other pre-processing modules that are not dealt with in this article.

Given a data-ready image, the process engine initiates a structured image representation protocol with which the input data are projected onto a region adjacency graph (RAG). Utilizing RAGs allows for rapid switching between the most commonly used dendrograms, examples of which are the Max-Tree [Salembier98] and the Alpha-Tree [Ouzounis11b]. The type of the tree structure and the pool of associated node attributes are set by the user. Following is a brief description of the two trees and the purpose that each one serves.

The Max-Tree

The Max-Tree [Salembier98] is a hierarchical image representation structure that encodes the partition of flat zones of the input image. A flat zone is a connected component of the level set, i.e. a region consisting of path-connected iso-intensity image elements. A peak component is a connected component of the threshold set. If a peak component defines a single flat zone of the same extent, it is referred to as a regional maximum. The Max-Tree is a rooted, uni-directed tree with its leaves corresponding to regional maxima and its root corresponding to the single connected component defining the background. Each node associates to a set of flat zones for which there exists a unique mapping to a peak component. Every node points to its parent which is the node associated to the first ancestor below the given level. The root points itself. This child-to-parent linkage reflects the nesting of level components along the gray scale. Nodes are addressed by their gray level and the node-at-level index. The node structure typically consists of four members; the gray level \( h \), the gray level after node processing \( h' \), the parent address and a pointer to an auxiliary data structure. The latter is variable in size and it is set and initialized upon selecting the types of attributes to be supported by the tree for further processing.
The Max-Tree algorithm runs a three stage process cycle in which the tree construction is separate from image filtering and restitution. This allows for interacting with the tree structure without the need for re-computing it. An example of a Max-Tree on a 1D signal is shown in Fig.1. Reversing the nesting order, i.e. having the brightest component defining the background, and dark components defining the leaves of the tree, yields the dual structure that is referred to as the Min-Tree. It was shown (Urbach07) that the Max-Tree of the inverted input image with respect to its gray-level range is equivalent to a Min-Tree.

The Alpha-Tree

The Alpha-Tree (Ouzounis11b, Ouzounis11c) is a hierarchical partition representation structure of the input image. It encodes the stack of nested partitions of $\alpha$-connected components or $\alpha$-CCs (Nagao79), with $\alpha$ being a threshold on some pre-specified dissimilarity metric between elements of the image space. In its simplest form, the dissimilarity is set to be the intensity difference or slope between adjacent pixels.

The Alpha-Tree nodes correspond to unique $\alpha$-CCs that are addressed with respect to the $\alpha$-threshold and a node-at-$\alpha$-level index. The leaves of the tree correspond to regular flat zones, and its root to the single-cell partition whose extent covers the image definition domain in its entirety. It is a rooted and uni-directed tree and like with Max-Trees, each node points to its parent, and the parent points to itself. Each tree node has a pointer to the auxiliary data structure discussed earlier, which allows for filtering and segmentation based on a pool of different attributes, during the processing stage. This is detached from the tree construction, and can be iterated in real time. Filtering the Alpha-Tree yields a segmentation of the image based on the concept of attribute constrained connectivity (Soille07, Soille08, Ouzounis11b). This is a recently introduced method in remote-sensing optical image analysis and several applications have been demonstrated.

III. IMAGE POLYCHOTOMY AND DATA ORGANIZATION

The projection of an image onto a hierarchical representation structure allows for its decomposition based on a wide range of structural, intensity and other statistically derived features. This has been typically approached by repetitive attribute filtering using attribute thresholds from a totally ordered set. This approach however leads to efficiency bottlenecks and instead, a number of different methods have been proposed that calculate the decomposition directly from the underlying tree structure. Examples are the connected pattern spectra (PS) (Urbach07) and the differential attribute profiles (DAP) (D.Murra10).
Connected pattern spectra (PS) are multi-dimensional attribute histograms that register the contribution of the image information content (connected components) with respect to a bounded attribute value range that is discretized to a pre-specified number of bins. Connected pattern spectra can be computed from both the Max-Tree and the Alpha-Tree structure, and an example is shown in Fig.3.

Differential attribute profiles have been implemented on the Max-Tree structure (Ouzounis11a). The DAP of a given image element \( p \) is the concatenation of two, equal in size, 1D vectors that are both perpendicular to the image plane and point at opposite directions. They are often referred to as the positive and negative vector of the DAP of \( p \). The set of all DAPs corresponding to the entire image definition domain is called the DAP vector field (Ouzounis11a). The latter is separated to its positive and negative instance, which are essentially volume sets like the example shown in Fig.4 (b). The positive instance is a hierarchical ordering of connected components with respect to a given attribute, assuming that foreground components are bright and rest against a dark background. The negative instance assumes the inverse intensity ordering. Each vector element corresponds to an intensity difference. Consider the positive instance for example, and assume an attribute threshold vector of \( I \) elements/scales. Given an attribute filter that reduces the bright information, assume that it is iterated for two consecutive thresholds addressed by \( t \) and \( t - 1 \) such that \( t \) belongs to \( I \) and \( t \geq 1 \). Subtracting the two results yields a DAP vector field plane whose intensity at point \( p \) defines the value of the element \( t \) of the positive vector. Iterating the attribute filter for all threshold values and performing the same operation yields the complete positive vector for each image element.
Both decomposition methods can be labeled as polychotomy schemes in the sense that they allow the organization of the input data into multiple layers, according to one or more criteria. The term polychotomy is used to differentiate between operations directly on the image plane that yield a decomposition and those on the respective tree that yield multiple cuts on the structure. Moreover, the two schemes are inter-related; each instance of the DAP vector field corresponds to the respective 1D connected pattern spectrum, i.e. each plane consists of all the peak components explicitly contained at one pattern spectrum bin.

**IV. INTERACTIVE IMAGE INFORMATION MINING AND APPLICATION EXAMPLES**

The polychotomy resolution is application dependant and can be readjusted interactively or automatically through some optimization procedure. The DAP plane or PS bin splitting or merging (Ouzounis11a, Urbach07) is an operation on the underlying tree-based image representation structure and comes at a minimal computational cost. The polychotomy is controlled through a switchboard that provides the enable/disable functionality on each PS bin. The DAP vector filed is manipulated through a switch vector and the border values of each bin can be adjusted interactively. Bins are color-coded to signify the amount of image information content they associate to.

Image information mining based on the switchboard interface is bidirectional. The contents of selected image regions can be used to highlight the respective bins, and enabling switches on the switchboard can be used to retrieve the respective connected components from the image. In the first case, image regions of interest are represented by some closed polygonal entity that allows for the extraction of all connected components fully contained within it (Ouzounis11d). The component attributes are retrieved from the corresponding tree-based representation and the respective bins are identified. Polygons can be marked manually using standard input devices like mouse/joystick or by advanced human-computer interfaces like the visual tracker (Castoldi11). The latter has been demonstrated in the visual search for rubble on Google aerial images of Port-au-Prince, Haiti, made available following the devastating earthquake of January 12th, 2010. Fig.5 shows examples of image tiles containing rubble (top row) the user response (bottom row) in terms of fixation points generalized by spatial aggregation, observation persistence and manual confirmation (mouse clicks).

The second case in which, the image information content is fully projected on the switchboard, enabling individual or groups of switches lets connected components of attributes within the ranges of the selected bins to be marked on the image. This method can be used for verifying the suitability of the applied attribute.
threshold discretization and for visual exploration of the image content based on a given set of attribute
criteria. An example concerning the same crisis scenario of Haiti, is shown in Fig.6. The input image tile is
shown in (a) and the $\alpha$-connected pattern spectrum at the top of (b). The latter is configured with variance and
area attributes, along the $x$ and $y$ axis respectively. The images are reproduced from (Ouzounis11c). Selecting
bins that describe best the rubble clusters with respect to the given attributes (bottom of image (b)), produces
a rubble layer shown in (c) that subject to a spatial aggregation by a local average leads to the rubble map,
shown in (d). This process chain allows for rapid mapping of disaster indicators and has been used to evaluate
the damage extent in the particular event.

threshold discretization and for visual exploration of the image content based on a given set of attribute
criteria. An example concerning the same crisis scenario of Haiti, is shown in Fig.6. The input image tile is
shown in (a) and the $\alpha$-connected pattern spectrum at the top of (b). The latter is configured with variance and
area attributes, along the $x$ and $y$ axis respectively. The images are reproduced from (Ouzounis11c). Selecting
bins that describe best the rubble clusters with respect to the given attributes (bottom of image (b)), produces
a rubble layer shown in (c) that subject to a spatial aggregation by a local average leads to the rubble map,
shown in (d). This process chain allows for rapid mapping of disaster indicators and has been used to evaluate
the damage extent in the particular event.

threshold discretization and for visual exploration of the image content based on a given set of attribute
criteria. An example concerning the same crisis scenario of Haiti, is shown in Fig.6. The input image tile is
shown in (a) and the $\alpha$-connected pattern spectrum at the top of (b). The latter is configured with variance and
area attributes, along the $x$ and $y$ axis respectively. The images are reproduced from (Ouzounis11c). Selecting
bins that describe best the rubble clusters with respect to the given attributes (bottom of image (b)), produces
a rubble layer shown in (c) that subject to a spatial aggregation by a local average leads to the rubble map,
shown in (d). This process chain allows for rapid mapping of disaster indicators and has been used to evaluate
the damage extent in the particular event.
Selected bins on the pattern spectrum can also be used as feature vectors to train statistical learning algorithms. This method is used to fine-tune the attribute ranges of the targeted features on an image tile basis, when major modality variations influence the radiometric properties of the total area of interest. This can be due to tile acquisitions from different sensors, changes occurring due to seasonal / weather conditions, human intervention, natural phenomena and others. Fig.7 shows an example of learning the targeted patterns from sets of positive and negative examples. Image (a) shows an input tile of urban settlements. The marked area is a zoom in section shown in (b). Positive and negative samples are collected by manual setting of polygons on targeted regions. The connected components fully contained within the polygonal selections (d) are projected onto the switchboard (f), where bins are color-labeled depending on the class membership and confidence level. Training an SVM classifier with the labeled bins in the form of a feature vector produces a decision rule that is used to filter the respective tree. Images (b) and (e) show the filter outputs for the input image and the zoom-in section respectively.

The last example demonstrates the use the DAP decomposition scheme towards an ongoing project concerning the production of a new medium-resolution (75m) global human settlement layer (GHSL) for disaster mitigation and planning purposes in support of crisis management. The product is built using
Fig. 8 – Sample tiles of various medium and medium resolution global human settlement layers. Image (a) shows scattered image tiles corresponding to major urban realities, overlaid on Google Earth. The night lights layer by NOOA is shown in (b). The NASA MODIS layer at 500m resolution is shown in (c), and the JRC product built with ENVISAT-ASAR data at 75 m resolution in (d).

ENVISAT-ASAR data, provided by the European Space Agency (ESA). An example tile is shown in Fig. 8 (d). DAP vector field planes, are set to contain urban realities grouped into classes based on several attribute criteria. This multi-scale representation of the morphological image information content is fused together with the PANTEX output (Pesaresi08) and a backscattering data layer. PANTEX is a built-up index computed based on a set of rotation-invariant, anisotropic contrast, texture measurements that are calculated from the co-occurrence matrix of the gray-level input image. The GHSL output for the image tiles overlaid on Google Earth (image (a)) is shown in image (d). Images (b) and (c) show earlier urban layers produced by NOAA (DMS sensor) – the nighttime lights time series, with resolution approximated from a 30 arcs/sec grid, and by (Schneider10) using NASA MODIS data at 500m resolution, respectively. Compared to the JRC product a substantial difference is noticed in the accuracy of urban regions detection.

V. CONCLUSIONS

In this paper the core modules of the process engine of the JRC IQ system were presented. The system relies on a hierarchical image representation module that makes use of the Max-Tree and Alpha-Tree algorithms. Both tree-based image representations are used to compute high-resolution pattern spectra that serve as
control modules through the switch functionality (switchboard) described. The switchboard further to the
direct control over the pattern spectrum is used to adjust the DAP vector field plane separation, which
implicitly affects their contents. Examples of interactive image information mining were given in real crisis
scenarios and in the context of crisis management. On-going work includes the exploration of further
attributes custom to built-up structures and the porting of existing algorithm on advanced computational
platform like distributed memory machine and cloud computing environments. A new VHR (2-0.5m) GHSL
based on optical data is currently being developed.

VI. REFERENCES

(Castoldi11): R. Castoldi, M. Broglia, M. Pesaresi. Image interpreters and interpreted images: an eye tracking
study applied to damage assessment. Proceedings of the 3rd Int. Workshop on Validation of Geo-information

3762, 2010.

(Nagao79): M. Nagao, T. Matsuyama, Y. Ikeda. Region extraction and shape analysis in aerial photographs.

(Ouzounis11a): G.K. Ouzounis, M. Pesaresi, P. Soille. Differential area profiles: properties and efficient
computation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011 [accepted].

(Ouzounis11b): G. K. Ouzounis, P. Soille. Attribute-Constrained Connectivity And Alpha-Tree Representation

(Ouzounis11c): G.K. Ouzounis, P. Soille. Pattern spectra from partition pyramids and hierarchies.
Proceedings of the 10th Int. Symposium on Mathematical Morphology, Intra, Italy, LNCS, Springer, pp. 108-
119, July 6-8, 2011.

(Ouzounis11d): G.K. Ouzounis, L. Gueguen. Interactive collection of training samples from the Max-Tree
structure. Proceedings of the 18th IEEE International Conference on Image Processing, Brussels, pp. 1481-
1484, September 11-14, 2011.

(Pesaresi08): M. Pesaresi, A. Gerhardinger, and F. Kayitakire. A robust built-up area presence index by
anisotropic rotation-invariant textural measure. IEEE Journal of Selected Topics in Applied Earth

(Pesaresi10): M. Pesaresi. Toward global automatic built-up area recognition using VHR imagery. IEEE

(Schneider10): Schneider, A., Friedl, M., Potere, D., 2010, Monitoring urban areas globally using MODIS
1733-1746.

