A Global Human Settlement Layer From Optical HR/VHR RS Data: Concept and First Results

Martino Pesaresi, Guo Huadong, Xavier Blaes, Daniele Ehrlich, Stefano Ferri, Lionel Gueguen, Matina Halkia, Mayeul Kauffmann, Thomas Kemper, Linlin Lu, Mario A. Marin-Herrera, Georgios K. Ouzounis, Marco Scavazzon, Pierre Soille, Vasileios Syrris, and Luigi Zanchetta

Abstract—A general framework for processing high and very-high resolution imagery in support of a Global Human Settlement Layer (GHSL) is presented together with a discussion on the results of the first operational test of the production workflow. The test involved the mapping of 24.3 million km² of the Earth surface spread in four continents, corresponding to an estimated population of 1.3 billion people in 2010. The resolution of the input image data ranges from 0.5 to 10 meters, collected by a heterogeneous set of platforms including satellite SPOT (2 and 5), CBERS 2B, RapidEye (2 and 4), WorldView (1 and 2), GeoEye 1, QuickBird 2, Ikonos 2, and airborne sensors. Several imaging modes were tested including panchromatic, multispectral and pan-sharpened images. A new fully automatic image information extraction, generalization and mosaic workflow is presented that is based on multiscale textural and morphological image features extraction. New image feature compression and optimization are introduced, together with new learning and classification techniques allowing for the processing of HR/VHR image data using low-resolution thematic layers as reference. A new systematic approach for quality control and validation allowing global spatial and thematic consistency checking is proposed and applied. The quality of the results are discussed by sensor, band, resolution, and eco-regions. Critical points, lessons learned and next steps are highlighted.

Index Terms—Built-up density, CSL, global human settlement layer, linear regression, PANTEX, urban limits.

I. INTRODUCTION

THIS paper addresses the general issue of the possibility to extract global geo-information layers from High-Resolution (HR)/Very High-Resolution (VHR) input image data. More precisely, the possibility to extract a Global Human Settlement Layer (GHSL) at scale 1:50K using optical sensors ranging from 0.5 to 10 m of spatial resolution. A new multi-scale image processing paradigm based on discrete field of image descriptors (DFID) and a new inter-scale learning and classification mechanism are discussed under the hypothesis of a realistic global scenario.

The experimental setting and the results presented in this paper are extracted from the first operational test of the Image Query (IQ) system[1], tasked with the production of the first public release of the JRC GHSL during June-July 2012. The design and implementation of the system supporting the test took approximately 12 months. At this stage the focus was on the strategies for multi-scale image features compression, storage and retrieval. The material presented in this paper focuses more on learning and classification techniques and the human settlements analysis. While both, the system design and the GHSL outputs, were assessed during the test reported here, the present paper concentrates on the description of the GHSL production workflow and the GHSL results. Only some general basic characteristics of the IQ system and computational requirements will be illustrated.

The experiment discussed here included both detection and an initial characterization of built-up areas based on average size (scale) of built-up structures. The dataset covers parts of Europe, South America, Asia and Africa for a total mapped surface of more than 24,300,000 km^2 (Fig. 1). In particular, the task involved the processing and evaluation of more than 15,000 satellite scenes collected from 10 different satellite platforms and sensors having spatial resolution in the range of 0.5–10 meters (Fig. 2). The test involved the processing, indexing, and classification of around $4.07\text{E} + 12$ records containing the image elements (pixels) and their attributes (features) (Table I). The purpose of the test was two-fold: i) to study the feasibility and the system requirements necessary for the production of image-derived information layers supporting crisis management in realistic scenarios [1] and ii) to design and evaluate one specific image information retrieval task involving the production of globally consistent outputs. The realistic data, time and information use scenarios applied in this experiment are derived from: i) direct experience of the authors in supporting information needs for operational crisis management activities performed inside their institutional mandate and ii) the participation of the authors in a number of R&D and pre-operational projects exploring the

[1]The IQ system has been designed and developed within the Global Security and Crisis Management Unit of the European Commission's Joint Research Center.

Fig. 1. Geographic distribution of the HR/VHR input images processed during the experiment.
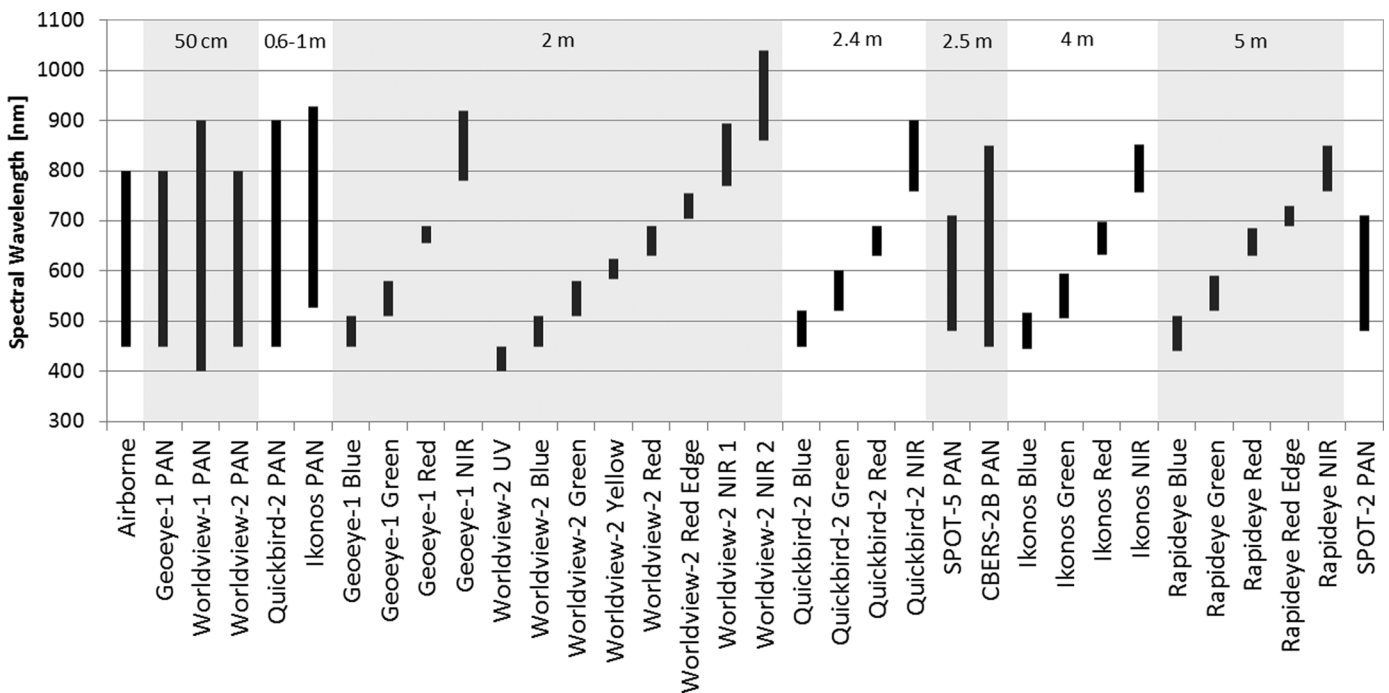

Fig. 2. Spectral coverage of sensors used in the study. The satellites cover a wide spectral range in the visible and NIR part of the spectrum. The spatial coverage includes various resolutions from 50 cm airborne to 10 m panchromatic images of SPOT 2.

exploitation of remote sensing (RS) technologies for civil security applications. Some references to these activities are briefly summarized in Section IV.

The interest of this study for the remote sensing community relies on the potential new solutions regarding two general issues of the RS community, namely: i) the capacity to derive globally/regionally-consistent information from HR/VHR input imagery, and ii) the capacity to perform such image information extraction taking into account the constraints of crisis scenarios with respect to time and data/metadata quality. Both issues are today only in parts addressed in literature and were never tested before with such a complex and representative set of HR/VHR data.

Under the proposed perspective, a processing strategy with some chances of success must address three main interconnected challenges: *input data volume, input data intrinsic inconsistency, and fully automatic processing chain*. The challenges are addressed mostly in Section III and some additional aspects are discussed in the next section of the paper.

In general, the paper is organized as follows. Section II presents the rationale behind the GHSL production including information needs and remote sensing potential. The key methodological choices regarding image information extraction are described in the Section III. Section IV summarizes the use scenarios and the derived GHSL definitions and technical specifications. Section V describes the input data, the

TABLE I
IMAGE DATASETS USED IN THE EXPERIMENT

Platform name/version	Count of image ID	Sum of input surface (km^2)	Sum of input data volume (pix)
AERIAL	6	9.21E+04	3.68E+11
CBERS 2B	9762	1.08E+07	1.72E+12
GEOEYE 1	496	8.70E+04	9.42E+10
IKONOS 2	1421	3.47E+05	9.08E+10
QUICKBIRD 2	1299	5.84E+05	3.26E+11
RAPIDEYE 1	30	1.88E+04	4.44E+08
RAPIDEYE 2	20	1.25E+04	2.96E+08
RAPIDEYE 4	100	6.25E+04	1.48E+09
SPOT 2	1	5.01E+03	5.01E+07
SPOT 5	3168	1.22E+07	1.33E+12
WORLDVIEW 1	27	7.83E+03	2.40E+10
WORLDVIEW 2	402	1.41E+05	1.07E+11
Grand Total	16732	2.43E+07	4.07E+12

workflow as well as the system managing it. Pre-processing, feature extraction, and learning/classification are detailed in Sections VI–VIII respectively. Quality control is described in Section IX. Results are then discussed in Section X. The paper concludes in Section XI with a summary, critical points, and the way forward.

II. RATIONALE

A. Information Needs

The information on human settlements is crucial for a wide range of applications including emergency response, disaster risk reduction, population estimation and analysis, or urban and regional planning, just to name a few. Urbanization plays a central role in this context. Urbanization pressures have an environmental impact, indicate population growth, and relate to risk and disaster vulnerability.

In 2011 the global population passed the level of 7.0 billion and more than half of the population is living in urban areas. Between 2011 and 2050, the world population is expected to increase by 2.3 billion and the urban population to increase to 2.6 billion, passing from 3.6 billion in 2011 to 6.3 billion 2050 [2]. The population growth expected in urban areas will be concentrated in the cities and towns of the less developed regions. Asia, in particular, is projected to see its urban population increase by 1.4 billion, Africa by 0.9 billion, and Latin America and the Caribbean by 0.2 billion. Population growth is therefore becoming largely an urban phenomenon concentrated in the developing world [3]. The figures alone are alarming enough and make us understand that we are facing major challenges to manage the urban development in a sustainable way. A central issue in this respect is the availability of up-to-date information on the extent and quality of the urban settlement. In particular in less developed countries such information is largely unavailable. Cities are often growing at a pace that cannot be controlled by the local or regional mapping agencies.

Apart from the general need for human settlement information and hence population, there is a particular lack of information on refugees or internally displaced persons (IDP) and the urban poor, which are living often in informal, sub-standard slum settlements. Despite the fact that they are neglected both by the classical topographic mapping methodologies and by the

land-cover/land-use standard classification schemata, the temporary human settlements of refugee/IDP camps are crucial for effective crisis management operations. In 2011, there were an estimated 26.4 million people displaced internally by conflict[2] and 15.2 million refugees around the world, including 4.8 million Palestinian refugees[3]. These numbers do not include IDPs related to natural disasters and other forced migration issues. The total number of refugees and IDPs was estimated in 2009 as 67 million of people[4].

The information about the quality of the (urban) settlement can provide precious input for understanding the vulnerability of population living on our planet. Effective crisis management and sustainable planning activities need consistent monitoring of BU areas, in particular of the urban poor. These sometimes large parts of the urban areas are characterized by sub-standard housing without access to water and sanitation and tenure insecurity. According to the United Nations, due to rising population and the rise especially in urban populations, the number of slum dwellers is rising. One billion people worldwide live in slums and the figure is projected to grow to 2 billion by 2030[5].

B. Remote Sensing Sources

Satellite imagery today could potentially provide information about the built environment worldwide, due to advances in computational and storage capacity, as well as data availability and cost. As demographic pressure increases at global level, our ability to monitor, quantify and characterize urbanization processes around the world is becoming paramount. Despite this potentiality of remote sensing technologies, there are few global data sets that can be used to map the human settlement. Examples include the night-time lights of the world based on the DMSP-OLS sensor [4], MODIS500 based on land use/land cover classifications [5], [6] and global population data sets like LandScan [7] or the gridded population of the world [8]. Since 2011, the Suomi National Polar-orbiting Partnership (SUOMI NPP) satellite produces night-lights at 750 m spatial resolution[6]. An overview, comparison and analysis of eight global data sets is provided by Schneider *et al.* [6]. While these data sets are useful for global analysis, they have the tendency to under-represent small, scattered rural settlements with their low spatial resolution between 500 and 2000 m. In addition, they represent a single snapshot in time that does not allow a regular monitoring. Or, if they are updated, like the LandScan data set, they are not directly comparable due to changing input sources. Although high resolution (HR, 1–10 m spatial resolution) and even very high resolution (VHR, ≤ 1 m) data with an almost global coverage are available with different sensors (e.g., SPOT, CBERS, RapidEye, IKONOS, QuickBird, WorldView 1 and 2) no consistent global coverage exists. Mapping and monitoring of urban areas at HR and VHR scales are mostly limited in terms of

[2]URL: http://www.internal-displacement.org/ publications/global-overview

[3]UNHCR Global Trends 2001, URL: http://www.unhcr.org/4fd6f87f9.html

[4]World Savvy Monitor, 2009, URL: http://worldsavvy.org/monitor/index. php? option=com_content&view=article&id=441&Itemid=847

[5]'Slum Dwellers to double by 2030', UN-HABITAT report, April 2007. URL: http://www.unhabitat.org/ downloads/docs/ 4631_46759_GC%2021%20Slum%20dwellers%20to%20double.pdf

[6]http://npp.gsfc.nasa.gov/index.html

temporal and spatial coverage and remain at the stage of case studies for individual or few cities often providing only a single time-step [9]–[11]. The largest case study analyses 54 cities all around the world [12].

The lack of a consistent global layer with HR/VHR spatial resolution can be attributed mainly to two reasons. First, the data availability of HR/VHR satellite data. Most, if not all, HR/VHR satellite missions are operated on a commercial basis and consequently a global coverage is costly. The only relevant exception is the CBERS 2B platform releasing 2.5-m-resolution panchromatic data with a very open data sharing policy in Brazil. Secondly, to date no system has demonstrated the capacity to extract automatically global information layers about human settlements from HR/VHR satellite data with the necessary accuracy and cost-effectiveness. In the global perspective, the available automatic information procedures have the following necessities: i) collecting representative training sets, ii) expensive ad hoc parameter setting and tuning, iii) maintaining costly computational infrastructures, and iv) collecting specialized input information not available globally. As a consequence, so far only time-expensive manual or semi-automatic operational procedures were available.

III. METHODS

A. Problem Setting

As already introduced, under the perspective of global/regional processing of HR/VHR image data it is mandatory to address the three main challenges: *input data volume, input data intrinsic inconsistency, and fully automatic processing chain.*

It is trivial to notice that the image data volume increases exponentially with increasing spatial resolution: to pass from 500 m to 0.5 m of spatial resolution means an increase of six orders of magnitude of input data volume. The question is, if efficiency and effectiveness of a given processing workflow is also increasing with the same order of magnitude. Furthermore, it is necessary to observe that also the input data inconsistency, such as spatial inconsistency and spectral class variability, increases proportionally to the increasing spatial resolution. The two effects have different causes, but together with increased input data volume they may lead the data processing complexity to failure.

From the thematic point of view, it is now well known that increased spatial resolution of sensors leads to increased spectral variability of the thematic classes: this is due to the changed scale of observation of the image information or *targets*. Changing scale of observation to 0.5 m of spatial resolution may reveal that just an apparently simple *building roof* may become a complex universe made of gutters, chimneys, water tanks, windows, terraces, and even trees of roof gardens. They are composed of a plethora of different materials, such as clay tiles, corrugated metal, plastic, asphalt, concrete and so on. Illumination incidence angle, surface slope, shadows, spatial pattern of the elements of these surfaces (for example tiles) and their degree of obsolescence may change dramatically their spectral reflectance/absorption characteristics. As a result, the spectral variability of the class *building roof* will also increase,

which weakens the inferential models supposed to recognize *building roofs* from image data. Attempts to separate the problem in sub-problems by recognizing the different elements separately, will easily bring to explosion of the number of target classes and their specific instances. This fact typically leads to the degradation of the model generality and applicability across different scenes and/or different geographical places, and to the explosion of the cost needed for the collection of reference data aimed at training and testing purposes.

From the spatial point of view, it is worth noting that VHR data would need high accuracy Digital Surface Model (DSM) and field-collected Ground Control Points (GCP) in order to reach a RSM displacement tolerance in the order of 2–3 m. This means that also at the best conditions with 0.5 m of input resolution, a ground displacement error of 4–6 pixels must be expected. Moreover, because of the increased agility of the VHR platforms and sensors, VHR image data can be collected virtually with any viewing (off-nadir) angle, with the effect of increasing image apparent local displacements due to parallax and panoramic distortions [13]. In practice, for normal building in the range of 5 floors and common off-nadir angles this can be translated in additional expected apparent displacement in the order of 10–20 pixels. Moreover, in the perspective of this study global available DSM (90-m-resolution SRTM) must be used, and no field GCPs are available for all the scenes under process. Consequently, expected spatial inconsistencies at least of the order of 30–50 pixels must be expected in the information extracted from the available VHR images.

As it is evident from the above observations, VHR data input are intrinsically spatially inconsistent showing always displacement errors greater than the pixel size and under the specific operational constraints discussed here, they may be often in the order of 30–50 times the pixel size. This fact has direct effect in increasing the complexity of the reference data collection and in decreasing the expected accuracy and repeatability of the image information retrieval tasks, especially in the frame of monitoring activities.

A fully automatic processing chain is required for reproducible, cost-effective and sustainable image information retrieval in the conditions addressed by this study focusing on large areas with HR/VHR spatial resolution. Nevertheless, increasing input resolution and input inconsistency/variability typically decreases the stability of the inferential models translating image data in thematic information and dramatically increases the number of free parameters to be tuned. Moreover, more detail typically calls for more expensive training and testing reference data collection, then conflicting with the necessity of minimizing human intervention in the classification process. It is worth noting that all the above mechanisms may show additional multiplicative effects on the whole computational complexity of the image information retrieval.

The solving strategy proposed in this study relies in the following basic principles: i) strong push on *fast computational capacity* through new efficient multi-scale feature extraction and learning/classification algorithms; and ii) design of image information extraction methods focused on *robustness* and *generality* issues. In particular, the following general principles were applied during the study:

- Maximization of the inferential model stability by drastic reduction of the number of free parameters to be tuned;
- Minimization of the data inconsistencies/complexity i) by systematic comparison of both input and output with stable, globally complete, lower resolution reference image data and information; and ii) by the introduction of a new hierarchical processing schema named Discrete Field of Image Descriptors (DFID), distinguishing between the input resolutions (scales) of the feature extraction and of the learning/classification phases;
- Full multi-scale approach including the definition of a new explicit inter-scale information generalization rule set;
- Integration in the workflow of available crowd source and open source repositories of geographical features.

B. Problem Solving

The IQ infrastructure supporting the experiment allows the search and retrieval of image information contents based on similarities with respect to radiometric, textural and shape (morphological) image descriptors. From this point of view it can be considered to belong to the general family of well-established content-based image retrieval (CBIR) concepts and systems [14]. Relevant applications of the CBIR paradigm to the problem of content retrieval from remotely sensed imageries can be found in the Knowledge-driven Information Mining (KIM) system proposed by Dactu and others [15] and the Geospatial Information Retrieval and Indexing System (GeoIRIS) proposed by Chi-Ren Shyu and others [16], [17]. The IQ shares with KIM and GeoIRIS the similar basic philosophy, but with some key innovations related to i) the image features processing and compression, ii) the learning and classification/ranking mechanisms and iii) the whole data representation architecture. These innovations are crucial for permitting the management of the data volume and complexity of the GHSL experiment.

In particular, the image features processing cost was radically reduced through the introduction of innovative hierarchical decomposition algorithms in particular concerning the multi-scale morphological (shape) image descriptors that are notoriously computationally intensive [18], [19]. With the new approach an increased efficiency of two-three orders of magnitude was estimated with respect to the state-of-the-art of multi-scale segmentation techniques. With the same volume of data and processing capacities, this may mean passing from some hours (or even years for global data sets) to minutes just for the feature extraction processing cost. Moreover, the indexing and compression of image features was solved by an innovative *absolute clustering* approach combined with spatial aggregation techniques [20]: this approach allows avoiding the use of traditional per-scene statistical clustering creating instabilities and inconsistencies in case of multi-scene heterogeneous input datasets.

Furthermore, while traditionally the CBIR paradigm requires human interaction for sorting the image information, the IQ system has been designed for handling (if requested) massive unsupervised learning tasks. This is obtained by substituting human-driven examples with examples extracted from broad scale classification outputs globally available. Some details about these new techniques of inter-scale learning are detailed in Section VIII. Although in the experiment discussed here we addressed only the problem of detection of human settlement, the new inter-scale learning approach can be generalized for automatic fine-scale recognition of other land use land cover classes possibly available in broad-scale reference datasets.

Finally, the standard data representation philosophy in the CBIR paradigm was radically revised in order to cope with the challenges of the GHSL production using fine-scale input image data. The CBIR paradigm was initially built around the natural entity of *image* as individual document to be sorted through the use of image-derived features organized in relational databases. Multimedia and medical initial applications areas fit well with this apparently 'natural' approach. In reality it assumes an implicit and intelligent action behind the image data collection: an intelligent focusing and zooming of the sensor to a specific subset of the real world. This is often not applicable in remote sensing, where framing and zooming of satellite scenes are largely dictated by automatic mechanisms and technological constraints including sensor platform characteristics and orbiting parameters; this decreases dramatically the amount of 'sense' that can be retrieved in the specific scene cut. KIM and GeoIRIS partially address this issue by subdividing the satellite scenes in sub-parts or *tiles* that are consequently indexed and searched instead of the original scenes. In this way, we increase the probability to frame well a hypothetical image information target to be found in satellite images. Nevertheless, the tile size parameter remains an *a priori* parameter to be decided at the time of image data ingestion in the system, influencing the size of the image information targets that can be queried. Moreover, in this approach the tile geometry (origin, resolution, and projection) is dictated by the geometry of the input image: in a multiple scene/sensor scenario as the one tested during the GHSL experiment, this makes it difficult to compare the same image information query output, in the same place, but from different sensor/time input scenes. The approach proposed in this experiment relies on a consistent hierarchical multi-scale tessellation of the globe; thus overcoming the above drawbacks (Section IV).

From the point of view of the image information representation, the methodology proposed here can be defined as based on a Discrete Fields of Image Descriptors (DFID). This is in analogy to the discrete fields approach in physics, where complex phenomena difficult to be modelled in deterministic way at the micro scale, show much more stable statistical behavior, if summarized at the macro scale. Accordingly, radiometric, textural, and morphological (shape) descriptors calculated at the geometry (resolution, coordinate of the origin, projection) of the input imagery, are aggregated to the geometry of the global discrete field by analytical projective mapping transforms. The scale of the discrete field, thus the size of the cells, it is defined by the spatial resolution of the input data and their spatial uncertainty. This includes the implicit generalization may be introduced by specific image descriptors, as for example the window size and the structuring element size in the textural and morphological descriptors, respectively. The entities that will be classified are the cells of the discrete field organized in tiles
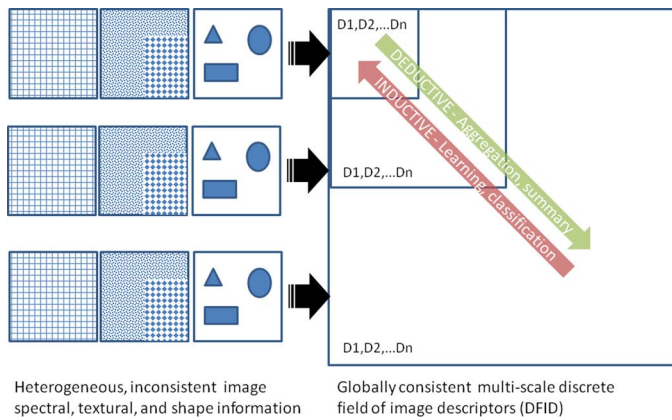
Fig. 3. The general Discrete Field of Image Descriptors (DFID) concept.

in order to optimize the I/O through raster database (DB) operations. These cells are described by the image descriptors inherited from images processed at the original resolution, but then summarized at the size of the discrete field cell. Fig. 3 depicts the proposed DFID concept. Inside and between the discrete fields, which may have different scales, various information collection and distillation processes may be discerned. In the proposed approach the most important ones are related to aggregation and summarization from detailed to broader scales and, symmetrically, learning and classification from broad to more detailed scales. The first kind of information processes can often be formalized by deductive and deterministic processing chains, as for example the generalization protocol discussed in the Section VIII-E. On the other side, the second kind of information process is typically made by inductive statistical chains, as for example the learning techniques discussed in the Section VIII.

Regarding the pixel-oriented image analysis methodology, the proposed DFID provides a more consistent approach for integrating in the same classification task heterogeneous image descriptors: in particular for integrating textural and morphological (shape) multi-scale descriptors together with traditional radiometric descriptors that have a 'natural' representation at the scale of the pixel.

With respect to other image analysis methodologies based on image segmentation and classification, as the so-called object-oriented image analysis (OBIA)[21], [22], the proposed DFID method demonstrates several key advantages: i) explicit management of spatial uncertainty, ii) stabilization of the inferential models, iii) possibility of a second-level pattern analysis of the results, and iv) reduction of memory requirements.

The DFID method allows complete and consistent management of the spatial uncertainty embedded in the image information at any scale, which is one of the main drawbacks of OBIA methods assuming spatial uncertainty always negligible respect to the image pixel size. As discussed before, these conditions are illusory in real HR/VHR image data processing scenarios and they may become completely misleading in case of thematic information targets having a size comparable with the expected input spatial uncertainty and apparent observable displacement. In the operational constraints discussed here, the image information retrieval methodologies based on preliminary image segmentation steps will face the computational problem in the post-processing trying to filter and re-aggregate spurious image segments resulting from misplacement of input data collected by different sensors and/or the same sensor in different times.

Moreover, the DFID method stabilizes the inferential information extraction model by first aggregating several image object/region instances in the same cell and then taking the classification decision based on the whole aggregated attributes. This is compatible with some well-established machine learning methodologies as for example the bag-of-the-words approach. The increase of the number of instances makes the statistical inference more stable. At the level of the cell, omission and commission recognition errors may compensate reducing the whole error rate. The OBIA paradigm instead takes the classification decision typically at the level of the single object/region that is more risky from the statistical point of view: potential errors made at this point are thus directly propagated in the subsequent inferential steps.

Furthermore, DFID allows computationally-efficient multi-scale pattern analysis of the image information retrieval results. The same mathematical tools allowing pattern analysis on lattice or raster structures can be translated to the analysis of DFID.

Finally, DFID significantly reduces the memory required for storage of image-derived information and consequently the I/O efficiency of the whole image information retrieval workflow. Internal estimations show memory requirements reduction of one to two orders of magnitude for comparable image information (detail) stored using the DFID and OBIA approaches.

The experiment presented here describes the JRC GHSL production test of December 2011, which was partially reported in [20] with the focus on the new multi-scale morphological decomposition techniques applied during the experiment. The current test significantly improves the precedent one regarding i) the capacity to handle a much larger input data complexity and volume, ii) the capacity to measure the consistency with respect to available global information layers, and iii) the capacity to validate the GHSL output.

Specific parts of the feature extraction and image classification steps included in the current IQ GHSL workflow were previously tested for automatic recognition and analysis of BU areas. In particular, the so-called PANTEX image feature [23] derived from anisotropic rotation-invariant grey level co-occurrence matrix (GLCM) textural contrast measurements and the characteristic-saliency-level (CSL) feature model [20] based on Differential Morphological Profiles (DMP) [24] and derived morphological decomposition techniques [18], [19], [25], [26].

The capacity to discriminate built-up (BU) from non-built-up (NBU) areas of the PANTEX features was previously assessed in a number of experiments including multi-temporal SPOT panchromatic data [23], [27], and a set of 56 globally-representative VHR optical scenes representing large cities [12]. Strong correlation between PANTEX image features and local density of building footprints was proven in [28], [29] using cadastral data as reference. Observing the reality at the range of scales (resolution) of this study, the physical reasons behind the correlation between PANTEX measurements and the presence of BU areas can be resumed as follows: i) the fact that BU areas are generally made by relatively small patches of heterogeneous materials and the fact that BU structures generally

cast shadows and ii) the fact that the human settlement areas show a strong dominance of objects with square corners. The PANTEX measures the local contrast then is sensible to local dissimilarity of material and presence of shadows creating high local reflectance variations. Moreover, because of his embedded anisotropic composition rule the PANTEX method can be considered as a corner detector. As a consequence, man-made objects generally receive high scores with this image-derived measurement. Similar techniques were also tested successfully using radar data input [30]. The current experiment expands those findings to a much more general scenario. Moreover, the capacity to detect and describe the characteristics of the single BU structures using DMP and DMP-derived image features were also previously tested in a number of experiments including optical data [31]–[33], hyperspectral data [34], and radar data [35]. These techniques demonstrated also effectiveness in characterization of BU areas as automatic detection of post-conflict damage assessment [36], [37] post-earthquake damage assessment, [38], [39], destroyed buildings and rubble detection [40] and more generally in image information mining tasks [16], [41]. Similar techniques integrating morphological and textural image features were also experimented in a first implementation of the JRC GHSL concept that was using ENVISAT 75-m radar input imagery [42], [43]. ENVISAT input datasets are not included in this IQ GHSL test focusing on processing input imagery with spatial resolution better than 10 m.

In the experiment discussed here only textural and morphological image descriptors were used for detection and characterization of BU patterns. This is due to the fact that at the time of the experiment only a subset of the highly heterogeneous input image data was suitable for a consistent spectral calibration with the available methods.

IV. GHSL CLASSIFICATION SCHEMA

A. Use Scenarios

Since the aftermath of the Indian Ocean tsunami in 2004, the JRC offers operational support to European Commission foreign policy services engaged in post-disaster damage and need assessment by providing remotely-sensed data interpretation. These activities were requested in a number of disaster and crisis scenarios all around the globe. The GHSL concept and requirements were developed in support to damage and need assessment and discussed with other international players such as the UN agencies and the World Bank and formalized in the post-disaster need assessment (PDNA) agreements and operational protocols. The GHSL technical specifications were also discussed with the Global Disaster Alert and Coordination System (GDACS) hosted by the JRC in order to produce more accurate exposure information layer for the automatic impact and alert modelling in GDACS. In addition, the JRC is contributing to collecting user requirements and in exploring technical feasibility of information extraction from remotely-sensed data in support to crisis management and disaster mitigation scenarios. The GHSL concept was customized inside applications involving population estimation, vulnerability, risk, exposure mapping, and damage and needs assessment in a number

of projects funded in the 6th and 7th Framework Programme for Research of the European Union. The human settlement layer specifications were discussed in particular in the projects Global Monitoring for Security and Stability (GMOSS)[44], GMES services for Management of Operations, Situation Awareness and Intelligence for regional Crises (G-MOSAIC)[7], the GMES emergency response services (SAFER)[8], and the GMES GE-OLAND[9].

The GHSL concept and technical specifications were discussed also with JRC partners having global information needs on human settlements. In particular, intensive exchange of know-how regarding the possibility of automatic analysis of human settlements using remotely sensed data was done with World Bank interested in globally consistent exposure mapping [45], with UN-Habitat, engaged in slum mapping and analysis, and with UNHCR offices interested in the estimation of population in refugees and IDP camps [46].

Finally, the broader societal impact of the GHSL concept both for technical-scientific and day-to-day users point of views were discussed in the frame of the new Digital Earth 2020 vision development [47], [48].

Definitions

The basic information contents of the current version of GHSL rely on the definition of BU structure (building) and BU areas: they are the necessary for a quantitative description of human settlement using HR and VHR remotely sensed data input [49]. BU areas are the spatial generalization of the notion of building defined as: 'areas (spatial units) where buildings can be found'. The working definition of BU structure (building) used in this experiment setting is as follows:

'buildings are enclosed constructions above ground which are intended or used for the shelter of humans, animals, things or for the production of economic goods and that refer to any structure constructed or erected on its site'.

This working definition is adapted from the data specification on buildings delivered by the Infrastructure for Spatial Information in Europe (INSPIRE)[10]. taking in to account the specific GHSL constraints and user requirements. In particular, by contrast to the INSPIRE definition, the GHSL definition does not include underground building notion for obvious limitations of the considered input data.

Moreover, GHSL notion does not impose the permanency of the BU structure on the site as instead INSPIRE does, following the classical topographic mapping tradition. The GHSL notion of BU structure is more inclusive, accepting to describe also structures belonging to temporary human settlements as refugee or internal displaced people (IDP) camps.

Finally, in a different way than INSPIRE, the GHSL repository includes also BU areas falling in the 'slum' or informal

---

[7]URL: http://www.gmes-gmosaic.eu/

[8]URL: http://www.emergencyresponse.eu/gmes/en/ref/home.html

[9]URL: http://www.gmes-geoland.info/

[10]INSPIRE Infrastructure for Spatial Information in Europe, 'D2.8.III.2 Data Specification on Building – Draft Guidelines', INSPIRE Thematic Working Group Building 2012 URL: http://inspire.jrc.ec.europa.eu/documents/ Data_Specifications/INSPIRE_DataSpecification_BU_v2.0.pdf

settlement concept: the area of a city characterized by sub-standard housing and squalor and lacking in tenure security, also called 'shanty town', 'squatter settlement' and similar.

It is worth noting that the GHSL definition is only partially fitting with other similar available definitions already popular in the RS community as the USGS *'Urban or Built-up areas'*[11], *'Impervious Surfaces'* [50], *'Urban Soil Sealing'*[12], CORINE *'Artificial Surfaces'*[13] and similar ones. Compared to these land-use/land-cover (LULC) definitions, the GHSL classification schema is more general not assuming any embedded urban/rural dichotomy (BU structures are mapped independently if they are falling in any 'rural' or 'urban' area definitions) and more focused on quantitative support to crisis management, risk and disaster mitigation activities requiring detailed mapping of buildings, population and their vulnerabilities with a multi-scale approach. Furthermore, the GHSL classification scheme with its simplification and reduction of the embedded abstraction was designed to facilitate the semantic interoperability and multi-disciplinary across-application sharing of data and results. This includes the sharing of data between different Agencies (UN, WB, EC) working in similar areas, but not necessarily sharing exactly the same abstract definitions[49].

B. Technical Specifications

This section includes a brief description of the GHSL technical specifications regarding format, scale, information production, and quality control. These specifications summarize the GHSL production guidelines used internally to coordinate the experiment.

1) Format: GHSL information is released through standard protocols defined by the Open Geospatial Consortium (OGC). In particular, Web Map Service (WMS) and Tile Map Service (TMS) platforms release the GHSL product as output of a specific query to a spatial DB, then including time and dynamic information queries. The native storage format and structure of the GHSL information is tile-based: the basic spatial unit is a surface tile, which is the representation of a given portion of the earth surface with a given size and projection. The tile entity is organized in a hierarchical multi-scale structure following the TMS standard.

While satellite-derived image features are stored and managed in the local (UTM) metric projection, the global mosaic, classification and GHSL representation are adopting a global metric projection, that is the Spherical Mercator (EPSG:900913) with the WGS84 Datum.

2) Scale: GHSL information is built and provided with three nominal scales of reference, namely local, regional and global scales. They correspond to specific parameters regarding the TMS zoom level, the spatial unit of reference and the tolerance admitted in the geo-coding of the information. Table II describes the relation between these parameters in the design of the GHSL product. The current experiment develops the output and the evaluation protocols exclusively at the regional and continental

[11]URL: http://landcover.usgs.gov/urban/umap/htmls/defs.php

[12]URL: http://www.eea.europa.eu/articles/ urban-soil-sealing-in-europe

[13]URL: http://www.eea.europa.eu/publications/ COR0-landcover

TABLE II
GHSL Scales, Spatial Units and Tolerances

Nominal Scale	Ratio	TMS zoom	Spatial Unit	Spatial RMS
local	1:10K	14	10 m	5 m
regional	1:50K	11	50 m	25 m
global	1:500K	8	500 m	250 m

TABLE III
GHSL Descriptors and Scales

Descriptor	Local 1:10K	Regional 1:50K	Global 1:500K
TileSurface	X	X	X
BuiltUpSurface	X	X	X
BuiltUpPercent	X	X	X
AverageSurfaceOfBuildings	.	X	X
NumberOfBuildings	.	.	X

scales, with the purpose to enlarge at the maximum the available universe of data under test. In fact, only a minimal part of the available image data was fitting the local-scale 1:10K quality specifications.

3) Information Contents: Assuming the collection of the two basic information from imagery: namely the built up area presence BU_{area} and the BU structure size BU_{size} for a given GHSL scale, then the derived GHSL variables are listed below:

- $TileSurface$ (m^2): It is the surface of the spatial unit calculated analytically from the projection and scale parameters;
- $BuiltUpSurface$ (m^2): It is the total surface BU in the specific spatial unit: $\sum BU_{\mathrm{area}}$
- $BuiltUpPercent$ (%): It is the percentage of BU surface in the specific spatial unit: $\sum BU_{\mathrm{area}}/TileSurface$
- AverageSurfaceOfBuildings (m^2): It is the average size of buildings expressed as average surface of building footprints candidates in the specific spatial unit: $\sum (BuiltUpPercent \times BU_{\mathrm{size}})$
- NumberOfBuildings: It is the number of BU structures estimated in the specific spatial unit: $\sum (BuiltUpSurface/AverageSurfaceOfBuildings)$

It is worth noting that because of the hierarchical characteristics of the TMS structure used to store the final geo-information layer, the information located in the broader scales can be calculated from aggregation of the same information potentially available in more detailed scales. If this information is valid the general hierarchical relation $\sum \mathtt{local} \Rightarrow \sum \mathtt{regional} \Rightarrow \mathtt{global}$ by aggregation (sum, average) is valid from detailed to general representation. Moreover, the spatial hierarchy is not only reflected in the aggregation rule, but also in the number of available descriptors. The number of abstract semantic layers available is increasing by increasing level of spatial generalization, and by passing from local to global scales. Table III summarizes the number of descriptors available in this version of the GHSL at the different scales.

V. Experimental Setup

A. General Objectives

As already introduced, the general objective of the experiment was to test the capacity to extract globally/regionally con-

sistent information layers from remotely-sensed imageries, in realistic data and use scenarios in the frame of crisis management applications. This means realistic data characteristics (resolution, volume, quality) and suitable response time. Specifically, the capacity to extract information about the presence of BU structures and their characteristics were tested.

Although several information layers at different scales were extracted during the test, only one specific output will be discussed in this paper, that is the automatic estimation of the presence of BU structures at scale 1:50K. The reason of this choice is that only for this product a suitable reference data set was available for validation purposes at the time of writing the present paper. The collection of reference data suitable for validation of the automatic characterization of BU areas at scale 1:10K is planned for the year 2013.

From the point of view of the thematic information extracted during the experiment, it is well known that urban areas and human settlements in general are one of the most challenging classes to be detected in remotely sensed image data input. This is due to the intrinsic heterogeneity of the surface materials contributing to the built environment, to the confusion with the same surface materials also available in natural areas, and to the effect of the built spatial patterns and illumination/shadow parameters in data gathered by the remote sensors. The approach in the current study was to maximize the generality and robustness of the information retrieval process, versus maximization of the accuracy based on specific sensor and/or geographical places characteristics. In practice, the main question that was addressed during the experiment was: 'is it possible to apply exactly the same image info extraction algorithm to the entire heterogeneous input data *universe* representing the complexity of the HR/VHR info extraction, without the need of manual tuning of the processing parameters?'. Other questions as for example: 'how can we optimize the recognition performance having the specific sensor X available in the region Y of the globe', even if in principle relevant were not addressed in the current study.

With the focus on global consistency issues, no intense benchmarking of specific set of parameters was applied in the experiment design. The only exception is the benchmarking of three different learning strategies as described later in this section. The general schema of the experiment is consequently defined as made by i) a heterogeneous HR/VHR input dataset, ii) the same image features calculated for the whole input, iii) alternative learning and classification strategies, and iv) the same reference data used for evaluation of the output.

B. Input Image Data Available

The satellite and airborne data used in this paper were acquired by optical sensors with a spatial resolution of 10 m or better in order to allow detection of single buildings or groups of buildings. The data are hosted in the Community Image Data Portal (CID)[14]. The CID Portal is a web portal to search and access remote sensing data and derived products hosted at JRC for a variety of applications.

In this study, we use in total 11,438 panchromatic and multispectral satellite data sets from SPOT 2 and SPOT 5,

RapidEye, CBERS-2B, QuickBird-2, GeoEye-1, WorldView 1 and WorldView 2. In addition, airborne data sets covering entire Guatemala were available as RGB imagery. The number of data sets per sensor and the area covered is detailed in Table I. The dataset under test covers parts of Europe, South America, Asia and Africa for a total mapped surface of more than 24,300,000 km^2. The input data volume is estimated in the order of 4.00+12 picture elements (pixels), stored in approximately 30 terabytes of disk space taking in to account the various number of bands, bit depth and compression formats applied in the available input scenes.

The different data sets cover a wide range of spatial resolutions from 50 cm airborne data sets to 10 m of the SPOT 2 sensor. Radiometrically the entire visible and near infrared part of the spectrum is covered with wide panchromatic bands and up to eight multispectral bands of WorldView 2. Fig. 2 depicts the spatial/spectral distribution of sensors used in this study. In addition some data sets consist of pan-sharpened multispectral images with the spatial resolution of the panchromatic band. It is important to note that no information were available about the pan-sharpening workflow applied for the production of these data, including visual enhancement filtering and the parameters necessary for eventual band calibration. Also concerning multispectral airborne data, no detailed information about the spectral characteristics of these data was available.

Around 50% of the VHR input data used in this experiment was available only in lossy data compression format: in particular JPEG, MrSid, and ECW formats were used in input during this experiment. It is worth noting that these formats introduce artifacts both in the radiometric and structural (texture, shape) image information descriptors, thus introducing robustness challenges in the whole image processing workflow.

Concerning the geo-coding, the input quality condition was highly heterogeneous: the expected RMS absolute positional accuracy was ranging from 3–5 meters of ortho-rectified data, 25 meters of raw VHR data, up to 40 kilometers in the CBERS 2B case.

The available input image data was collected in arbitrary and heterogeneous seasonal conditions, with arbitrary and heterogeneous sun/sensor elevation and azimuth parameters. In some 20% of the input data no precise information about collection parameters was available, especially in case of large mosaics of VHR input data made by third parties.

C. Global Reference Data

Several additional data sets were used in the workflow as ancillary data. For the orthorectification of some of the satellite data we used the TerraColor[15] as reference layer. This is an orthorectified global imagery base map at 15 m spatial resolution built primarily from Landsat 7 satellite imagery. The Open Street Map[16] (OSM) data were used to extract a high resolution land-sea border. During the processing low-resolution (LR) global data sets are used for reference purposes. One of the data sets is urban class of the MODIS500 Land Cover Type product

---

[14]URL: http://cidportal.jrc.ec.europa.eu/

[15]URL: http://www.terracolor.net

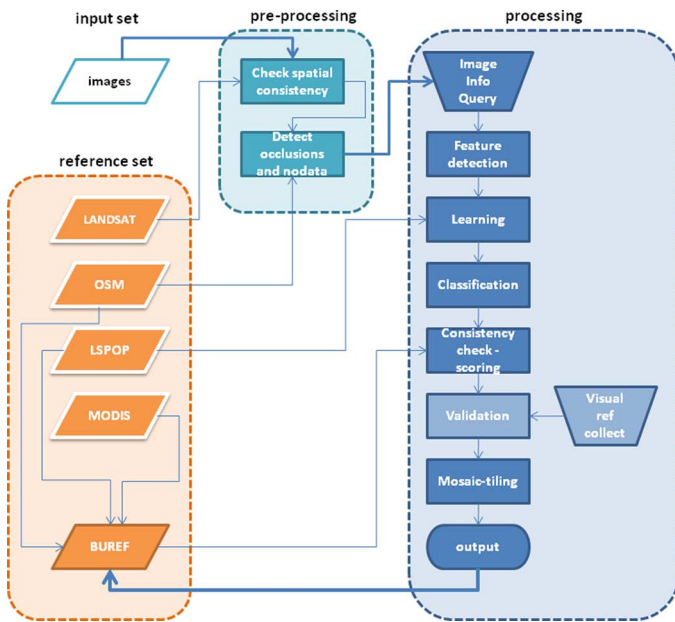[16]URL: http://www.openstreetmap.org/

Fig. 4.   The general IQ GHSL processing workflow.

[6]. In addition, the LandScan (2008 and 2010)[17] high resolution global population data sets were used. The LR reference data were used for learning and consistency checking purposes before and after the classification steps, respectively. During the learning, LR reference data substitute manual training set collection by a new inter-scale learning mechanism detailed in Section VIII. Also, LR reference information contribute to global consistency checking and optimization of several alternative outputs done at the end of the image information extraction workflow, during the mosaic and integration phase (Section V-E). During the global consistency checking, the active visual training collection loop is activated. See details in Section IX. Accordingly with the DFID paradigm exposed in Section III-B, no deterministic masking of HR GHSL information is done using LR information as input. Only statistic inferential chains are admitted from LR to HR information scales. Consequently, the GHSL output is considered information extracted from HR/VHR input image data. The only exception to this general rule is the land 'mask' applied to GHSL output that is extracted deductively from the OSM reference data. Consequently, the seawater vs. land dichotomy was not extracted from the input imagery by the proposed image information extraction workflow, but instead derived from an external source assumed as suitable for the purpose.

D.   General Workflow

The main ingredients of the IQ workflow characterizing the GHSL experiment here discussed are four: i) the input image data, ii) the reference set, iii) a pre-processing chain, and iv) a processing one (Fig. 4).

The reference set has the crucial function to support the optimization of the spatial and thematic consistency during the GHSL production. Before the first run of the workflow the

---

BUREF layer is produced by merging two globally available datasets: LSPOP and MODIS500. Details on this step are available in Section IX-C. Independently from that initial choice, the whole system is designed having an incrementally evolutionary approach: the output of any given image information extraction run/experiment, if passing validation and consistency check will contribute to improve (thematic accuracy, spatial/temporal completeness) the available BUREF layer. The expectation is that this retro-action mechanism will contribute in enhancing step-wise the overall reliability and completeness of the GHSL output.

The pre-processing module basically performs two functions: i) checking and optimization of the spatial consistency of the input image data and ii) checking and flagging eventual occlusions and no-data areas in the same images. The spatial consistency is optimized by using an available reference set having 15-m spatial resolution and an expected RMS spatial tolerance of around 20 m, while the occlusions and no-data areas are detected by an internal recognition mechanism. Details on these steps are available in Section VI. Because of the characteristics of the spatial reference layer available, the current GHSL production is not fitting the 1:10K local scale specifications, and consequently only the 1:50K regional and 1:500K global scales will be delivered.

Note that no spectral calibration steps are implemented in the current workflow. This fact means that all the subsequent processing steps work with spectrally uncalibrated satellite image data input. This choice was dictated by the fact that an important part of the input image data volume was not suitable for being radiometrically calibrated with existing tools, requiring the availability for each scene of parameters needed for top-of-atmosphere reflectance calculations. Apart from some VHR optical sensors, in particular this information was unavailable for pan-sharpened multispectral SPOT 5 image data and not possible for airborne/satellite mosaics that were instead playing an important role in the current experiment, both as volume of data involved and interest of users in the output.

The absence of radiometric calibration forced to exclude from the workflow the image descriptors derived from band combinations using sum, ratios operators as typically the NDVI indicator for example, that may significantly increase the BU/NBU discrimination in case of availability of multispectral image data. New spectral calibration methods able to process a more complete range of input data are under study and will be applied in next IQ GHSL workflow releases.

On the other side, the capacity to extract image information without using band combinations can be seen as a benefit of the proposed workflow, generalizing the same processing strategy to multi-spectral and mono-band panchromatic image data input.

E.   Parameter Sets and Task Loop

The whole processing is governed by a specific parameter set: in the experimental results discussed in this paper, three different parameter sets were put under test for benchmarking purposes. They are almost identical except for the option related to the learning and classification mechanisms activated during the workflow. More specifically, 'learning on LandScan', 'learning

on MODIS500 by area matching' and 'learning on MODIS500 by ROC optimization' were coded as parameter sets 142, 145, and 146, respectively. Details on these techniques are discussed in Section VIII.

The task loop governing the experiment processes independently with the same workflow each band of each input scene for each selected input parameter set. As a consequence, for each input scene s, n_s classification outputs will be available with $n_s = b_s * p$ and b_s equal to the number of bands available for the specific scene s and p equal to the number of tested parameter sets. Because several input scenes may contribute to the same DFID tile area with different sensor or collection parameters, the total number of outputs for each location x can be estimated as $n_x = b_s * p * ns_x$, with ns_x equal to the number of scenes contributing to the specific x place. In the general setup, all these potentially alternative outputs are sorted according to an explicit cost function including quality measurements and user-driven requirements as for example specific date or sensor constraints. This is done in the 'consistency check/scoring', 'validation' and 'mosaic/tiling' steps of the processing phase in the workflow (Fig. 4). In the current experiment, the scoring was performed exclusively according to the global quality measurements described in Section IX without additional constraints included in the cost function.

VI. PREPROCESSING

A. Positional Accuracy

The positional accuracy of each input scene and the positional consistency between any pair of overlapping scenes are fundamental quality measurements needed during the generation of the GHSL layer and for any subsequent use. This is a challenging issue given the heterogeneity of the input data sets and the absence of global reference layers matching the highest resolution of the input imagery (i.e., 50 cm). The best optical imagery reference layer with reported accuracy across the globe and freely available is the Landsat 7 nearly global coverage [51]. This freely available imagery needs to be mosaicked to ease accuracy measurements of arbitrary HR/VHR images across the globe. Although we have developed a method for mosaicking large data sets [52], [53] and have applied it at continental scale for Europe with Landsat data [54] and IRS LISSIII plus SPOT 4/SPOT 5 imagery [55], at global level we currently use Terra-Color as our imagery reference layer.

1) Positional Accuracy Against TerraColor: Except for CBERS-2 HRC imagery (see next paragraph), we have assumed that the geo-location of the input imagery as indicated by the image metadata is correct. The positional accuracy is determined using normalized cross-correlation measurements [56]. That is, given a square template cropped from the input image at a given position, normalized cross-correlations are calculated between this template and equivalent templates cropped in the reference image with center pixel falling within a search window centred on the same position as that of the original template. The vector separating this position from the position at which the maximum value of the cross-correlation occurs is used as a local estimation of the translation vector between the input and reference images at that position. If the

normalized cross-correlation is not circular enough or does not high enough, the estimated translation vector is deemed unreliable and discarded. The mean, RMSE, and standard deviation of the horizontal and vertical displacements are then reported. Sub-pixel measurements are obtained thanks to a quadratic interpolation of the correlation function [57]. Because the spatial resolution of TerraColor is 15 m, only errors above 1.5 m (the theoretical smallest displacement that can be measured with sub-pixel measurement using quadratic interpolation is a 10th of the resolution of the reference) are significant. This is enough for assessing the quality of the final GHSL layers since they are delivered at a maximum scale of 1/50000.

For the special case of CBERS-2 HRC imagery, the geo-location of the input imagery is unreliable with the scene center displaced by up to 40 km [58], [59]. In this case, we have used the procedure described above to automatically find tie points necessary to correct the geo-location of the input CBERS imagery. This is achieved by sub-sampling the CBERS images at the resolution of TerraColor. Displacement vectors are then extracted by considering as reliable only those vectors that clearly cluster in well defined region of the x-y space. The CBERS image is then warped using the affine transformation model with least square estimation of the transformation parameters given the final set of tie points.

2) Relative Positional Accuracy (Consistency): The relative positional accuracy (i.e., consistency) between any pair of overlapping images is calculated similarly to the positional accuracy. It is a consistency measurement in the sense no reference is used but merely a comparison between two data sets in their overlapping domain.

B. Cloud Detection

Precise cloud detection remains a difficult topic *per se* due to the absence of formal definition of a cloud that can be translated into an image processing chain with well defined parameter values. For instance, 'What is the threshold level to declare whether there is a thin cloud or not?' or 'Where is the precise location of the boundary of a cumulus?' are questions that will probably never be answered. Nevertheless, cloud masks indicating the image areas most affected by clouds can be generated by a wide variety of methods. For example, cloud detection in medium resolution imagery with sensors offering a wide range of spectral bands including thermal bands is usually based on the Automatical Cloud Cover Assessment (ACCA) method [60]. This type of method requires not only a wide range of spectral bands but also precise Top-Of-Atmosphere (TOA) reflectance values. The parameters necessary for TOA reflectance calculations such as sensor gain and offset parameters and solar irradiance at the given wavelengths are not always precisely known for HR/VHR satellite data. Experiments detailed in [55, ch. 3] have shown that ACCA can be adapted for sensors with green, red, near-infrared, and shortwave infrared bands such as SPOT 4 HRVIR, SPOT 5 HRG, and IRS-LISS III sensors. However, for VHR data that sometimes contain only one panchromatic channel, this path is not viable. Therefore, an ad hoc method for generating cloud masks has been developed. More precisely, rather than relying on calibrated data, we have developed a method that translates a visual definition of a cloud into an
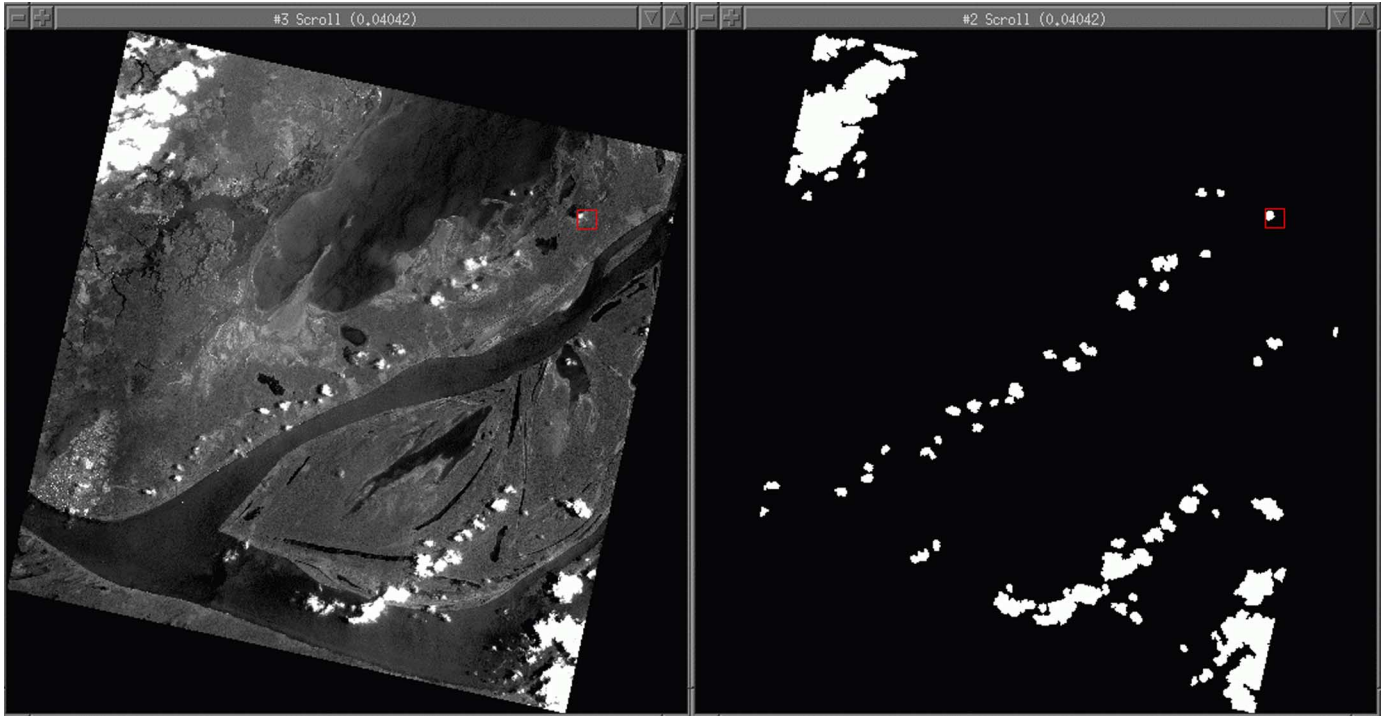
Fig. 5. Cloud detection from panchromatic data: example on a 2.5-m CBERS-2 HRC scene over Brazil.

image processing chain. There are actually two chains, the first for situations where only 1 VHR band is available, the other for all other cases (i.e., multi-band HR/VHR data).

1) Panchromatic VHR Data: This technique was originally developed for CBERS-2 HRC panchromatic images. It exploits the fact that the cloud boundaries consist of a smooth transition from bright to darker regions spanning over many pixels. Smooth transitions are detected using point-wise arithmetic differences between thick morphological gradients [61] of increasing size. The non-zero values of the resulting image delivers a mask of the cloud boundaries. The holes of this mask are than filled by a morphological reconstruction by erosion called the fillhole transformation [62]. Finally, a dilation by a disk of fixed radius (15 m) is applied to make sure that most pixels corrupted by the detected clouds are indeed covered by the cloud mask. Fig. 5 shows a CBERS-2 HRC image and its corresponding cloud mask.

2) Multispectral HR/VHR Data: For multispectral HR/VHR data (typically 4 channels), the adaptation of the ACCA method proposed in [55, ch. 3] did not provide satisfactory results in many cases due to the unavailability of precise calibration parameters. For this reason, we have developed a method taking into account the multispectral nature of the input data relying on a visual characterization translated into an image processing chain. The method relies on fact that clouds appear has regional maxima [62] in all visible and infra-ref channels of multispectral images. Therefore, the point-wise intersection of the extended regional maxima obtained for each channel is used at the basis of an image processing chain delivering a mask of potential clouds.

VII. Feature Extraction

This section gives an overview of the different types of features collected from the available imagery in order to shape a medium level semantic layer registering human presence through the evidence of built-up.

A. Textural Features

The textural image features used in this study are derived from grey-level co-occurrence matrix (GLCM) contrast textural measurements [63]. The GLCM matrix is a $n \times n$ matrix containing the relative frequencies with which two pixels linked by a spatial relation (displacement vector) occur on a local domain of the image, one with gray level i and the other with gray level j, with $i, j \in [0..n-1]$, where n is the number of gray-levels with which the image has been coded. The contrast textural measurement is formalized as follows:

$$CON = \sum_{i=1}^{n} \sum_{j=1}^{n} (i-j)^2 \cdot P_{i,j}$$

where n is the number of gray levels present in the image, and $P_{i,j}$ is the (i,j)th entry of the co-occurrence matrix. The contrast textural measures calculated using anisotropic displacement vectors are combined in a rotation-invariant image feature called PANTEX [23] by using extrema operators. PANTEX was demonstrated strongly correlated with the presence of buildings [12] and their spatial density [28]. The capacity of PANTEX to discriminate BU/NBU areas is mainly linked to the fact that it is a corner detector [64] and that the BU areas generate high local image spectral heterogeneity because

TABLE IV
PANTEX PARAMETERS USED IN THE STUDY.

Image Resolution (m)	Number of displacement vectors	Window Size (pixels)
10.0m	4	5
5.0m	6	10
2.5m	16	20
1.0m	28	50
0.5m	56	100

of heterogeneity of materials used and because of buildings are generally casting shadows. In this study, PANTEX method was improved by adding i) an *a priori* weighting mechanism substituting the usual standardization step before the integration of different displacement vectors and ii) the parametrization of the radius used for generating the displacement vector list.

The basic parameters necessary to calculate the PANTEX feature are the window size $wsize$ and the list of displacement vectors d used for generating the GLCM. $wsize$ is analytically derived from the GHSL scale of the expected output that in this experiment is set at the 1:50K GHSL technical specifications. Consequently, $wsize = 50$ meters that is corresponding to the spatial unit of the GHSL regional scale. The $wsize$ will be translated in corresponding number of image pixels by ratio with the image spatial resolution information. In this study, the displacement vectors are generated assuming a radius of 10 m that corresponds with the maximum size of the image pixel tested during the experiment, and fits well with *a priori* knowledge regarding the expected size of the majority of the BU structures. Table IV summarizes the $wsize$ and the number of displacement vectors d for the most common image resolutions used in the study.

B. Morphological Features

Radiometrical and textural features describe material properties and object patterns respectively, and are used as indicators of man presence. For each given scene both feature sets are employed as markers to validate building footprint candidates. The latter are summarized in separate information layer that is a product of a multi-scale morphological analysis protocol referred to as the "mtDAP".

The mtDAP protocol [18] computes the Differential Attribute Profile (DAP) vector field [25] from the input imagery. DAPs are nonlinear spatial signatures that are used extensively in remote sensing optical image analysis in ways analogous to spectral signatures. The DAP of a pixel is the concatenation of two response vectors. The first registers intensity differences, i.e., contrast, within a top-hat scale-scape of an anti-extensive attribute filter γ, and the second intensity differences on the bottom-hat scale-space of an extensive attribute filter ϕ. The pair (γ, ϕ) defines an adjunction with γ typically being a connected attribute opening and ϕ being a connected attribute closing.

The mtDAP can be configured with any morphological attribute filter but in this case, simple area openings and closings prove to be sufficient. The area attribute is used to order objects based on size and it is computed incrementally. The protocol

consists of three core modules; the hierarchical image representation using a Max-Tree and a Min-Tree structure [65], the attribute zone tree-polychotomy scheme and the spatial signature export module. A brief description of each module follows.

Hierarchical image representation schemes aim at organizing the image information content into meaningful structures or components and registering component transitions through their nesting properties. Examples in morphological image analysis are the Max-Tree [65], the Component-Tree [66], the Alpha-Tree [19], etc. This work makes use of the Max-Tree structure for computing anti-extensive attribute filters on gray-scale images. The Max-Tree is a rooted and uni-directed tree in which every node N corresponds to a single *peak component* that associates to a set of *flat zones* [67]. A peak component at level h is a connected component C, i.e., a connected set of maximal extent, of the binary set given by thresholding the input image at level h. A flat zone is an image region consisting of iso-tone and path-wise (or otherwise) connected image elements. If the full extent of a peak component coincides with a single flat zone, the component is referred to as a *regional maximum*. The leaves of a Max-Tree correspond to the regional maxima of the input image and every node points to its parent P that corresponds to the first superset C_P of C_N at level $h_P < h_N$. The root node corresponds to the image background and points to itself.

The Max-Tree treats bright structures as foreground information with respect to a darker background. The inverse is represented by a Min-Tree structure, i.e., foreground information are dark components resting against a brighter background. The Min-Tree is equivalent to a Max-Tree on the inverted input image.

Computing both structures on the input image offers the possibility of evaluating the significance or extent of the contribution of each connected component to the structured image information content with respect to the nesting order. This is the key idea behind the DAP decomposition which assumes size as the dominant criterion for ordering components. To obtain the DAP vector field, i.e., the set of all DAPs for a given image, the mtDAP protocol uses the Max-Tree to compute the top-hat scale-space of an area opening and the Min-Tree to compute the bottom-hat scale-space of an area closing. Instead of reiterating the same operator configured with a varying attribute threshold for each scale-space, the mtDAP employs the "one-pass" method [18] which generates an area zone polychotomy of each respective tree; each node is assigned a unique area zone identifier in a single pass through each structure.

The DAP vector field can be visualized as a 3D set in which every x-y plane, corresponding to a particular scale of the decomposition, stores in each pixel position the pixel's contrast computed in the respective top-hat or bottom-hat. An example is shown in Fig. 6(c) in which the top volume set corresponds to the top-hat scale-space and the bottom volume set to the bottom-hat scale-space. Exporting the DAP vector field requires a single pass through the image definition domain during which, for each pixel visited its node ID is retrieved. Visiting the respective node gives access to the area zone ID and contrast with respect to the zone's highest attribute boundary. The type of the tree is used to decide on which of the two volume sets the pixel is
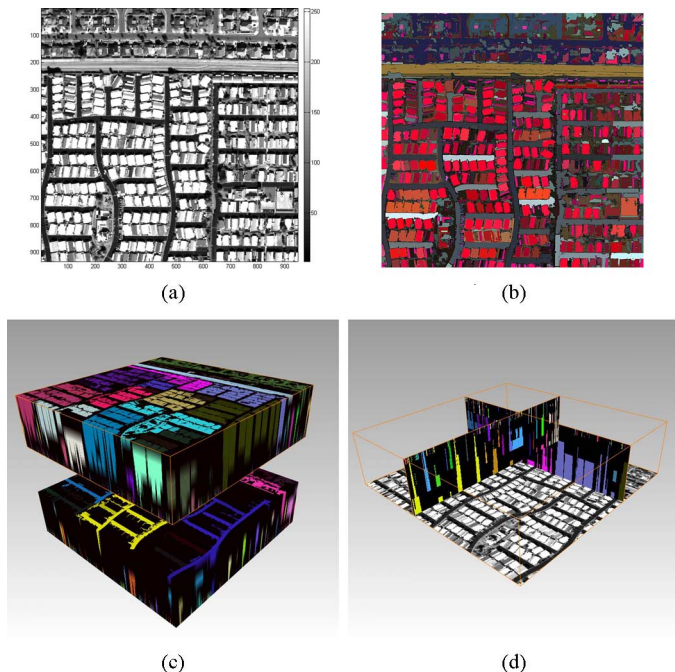
Fig. 6. Example of DAP vector fields: (a) the input image; (b) the color representation of the DAP vector field using the CSL model; (c) the DAP vector field in color-map projection in which the two volumes correspond to the opening top-hat and closing bottom-hat scale-space respectively; (d) a cross-section of the two.

be reported, the area zone ID to decide on which plane it belongs to and the contrast to assign a value on the corresponding coordinates on that plane. A cross-section through the two volume sets is shown in Fig. 6(d) where the scale-based responses of a sample set of pixels are shown.

C. Features Compression and Storage

The (scale) resolution of DAPs, i.e., λ vector length and the between-scale spacing, is a critical parameter in their utilization as feature descriptors. It is typically set either empirically or based on the outcome of some statistical learning procedure. Evidently, higher spatial input resolution offers a more detailed profile for each pixel under study. A drawback in this case is that by increasing the vector length I the number of DAP vector field planes (Fig. 6(c)) increases proportionally, i.e., $2 \times (I-1)$. When dealing with large data-sets this can be prohibitive and to counter the resulting data explosion a compression model was devised to radically reduce the dimensionality of the DAP descriptors. It is called the *Characteristic-Saliency-Level* or *CSL Model* [20] and is a medium abstraction semantic layer that can be projected on the HSV color space for the visual exploration of BU extracted from VHR satellite imagery (Fig. 6(b)).

The CSL model is a nonlinear mixture model consisting of three characteristic parameters extracted from the DAP of each pixel. That is the minimum scale at which the maximal contrast response is observed in the DAP (the characteristic), the contrast value (the saliency) and the highest peak component level from which this value is computed. The model is computed directly from the polychotomy of the two tree-based data structures and with no need for exporting the DAP vector fields. It reduces

radically the dimensionality of the DAP vector field to a three-band representation in a statistical-model free approach, i.e., it avoids clustering based on the statistical distribution of the DAP features of a given image. It does not require manual tuning and its computation is independent of the length of the DAP. This makes it suitable for user-independent applications like the one described in the population of the GHSL.

The color-mapped CSL model provides a pool of building footprint candidates like in Fig. 6(b). Warmer colors indicate higher confidence that a particular structure is a building. The selection of candidates employs the thresholded PANTEX BU indices as markers in order to draw the final high-level semantic layer containing all targeted BU.

Both the compressed multi-scale morphological (CSL) and textural (PANTEX) image descriptors are aggregated and stored with the spatial resolution corresponding to the finest scale supported by the specific image information query, that in this case is equal to a grid of 10 m, corresponding to the GHSL 10K specs. They are then ready to feed the subsequent image information queries and classification with the maximum of spatial detail. With this schema, the heterogeneous set of input sensor/platforms with different resolutions generates raw image features also heterogeneous in spatial resolution but then they are spatially standardized to the 10-m nominal grid. In many cases also this operation of spatial aggregation will introduce a compression of the data space with positive effects in the overall performances of the system (storage, indexing, and retrieval). For example, an input image feature generated with VHR image of 0.5-m resolution, if aggregated to 10 m will show a compression factor of 1:400.

The parameter set governing the CSL feature extraction was maintained constant for the whole process discussed in this experiment. A λ vector of 127 hierarchically increasing scales was set for a total of 254 scales in the γ, ϕ domains. The λ vector explored the set of image structures having a surface from 10 to 15,000 square meters that where considered enough to cover most of the target BU structures studied in the experiment. Consequently, an inter-scale step equal to 118 m was applied. Of course, during the processing of the input images these values are rescaled and rounded accordingly with the specific size and then surface of the input pixels.

VIII. LEARNING AND CLASSIFICATION

A. Learning Approach

In this experiment, a new inter-scale learning and classification paradigm was introduced with the objective of allowing a fully automatic processing chain for heterogeneous and not-calibrated input data set. The general idea behind this new approach is to move the calibration step from the input data—where it is placed classically—to the image-derived features (descriptors) before the actual classification. The general objective is to stabilize as much as possible the classification parameters against the complex input data used in this experiment.

The classical methodologies for standardization of image-derived features rely on observation of the statistical distribution of the values of the features in the specific scene. This strategy

was tested; but we rejected it during the experiment design, because it provided unstable results in multiple-scene, heterogeneous input data scenarios. In particular, scene-relative standardization approaches assume homogeneous (or at least comparable) distribution of land cover classes in each scene. This condition was largely violated in the discussed experiment set, where fully 'urban' scenes were processed together with scenes with only few buildings in some remote rural areas.

In the proposed approach, the high-resolution (HR) image-derived descriptors are rescaled through learning procedures that use low-resolution (LR) globally available information layers as reference. Of course, a correlation between the HR image descriptors and the LR global reference layers must be assumed. The role of the LR reference information layers is to increase the consistency and comparability of HR classification outputs produced from heterogeneous HR/VHR sources. It is worth noting that in the proposed approach this objective is achieved exclusively by the learning procedures described in this section; no *a priori* masking of HR data is performed with the LR reference information.

Image-derived features that are standardized with respect to an explicit objective function can be used for a fully automatic classification chain. Consequently, the advantage of the proposed methodology is that the collection of training samples can be avoided. This is particularly important for the whole experiment and in particular for testing the sustainability of global HR/VHR image information retrieval. Nevertheless, it is evident that between HR image descriptors and LR reference layers there is a scale gap that may introduce geo-spatial generalization issues. In order to mitigate the scale gap effects, three different modalities of learning and classification are implemented in the experiment: i) adaptive learning, ii) meta-learning, and iii) discovery.

In the 'adaptive learning' modality, the system optimizes the decision thresholds in the input features using a given reference layer. The 'meta-learning' modality is used to study the behavior of these decision thresholds in the set of scene processed and to detect regularities: for example, typical thresholds for a given sensor in specific regions. The output of the meta-learning is then exploited during the 'discovery' modality that can be activated in order to have the chance of recovering image information lost because of errors (such as incompleteness or inconsistencies) in the reference data, or different scale generalization of the image-derived information and in the available reference data. In practice, the adaptive learning optimizes consistencies between the image information under processing and the reference data, while the meta-learning and discovery modes take the risk of the image information recognition also in cases where reference data is not available with the necessary thematic, spatial precision.

The typical workflow combining the three modalities will be as follows: i) run adaptive learning in all available scenes and classify the outputs respect to the matching with the reference and the amount of available reference data, ii) run meta-learning in the set of successfully classified scenes with available reference data, and iii) run discovery mode in the set of scenes under a given threshold of quality detected after the adaptive learning phase.

B. Adaptive Learning on Textural Image Features

The PANTEX method [23] (see also Section VII-A) provides measurements correlated with BU density.

Nevertheless, PANTEX absolute values depend on specific image contrast and sharpness provided by a specific sensor/band in the specific scene under processing. While the measure has a high discrimination power of BU/NBU area in a variety of conditions [49], the optimal classification threshold depends on the general image contrast characteristics. To alleviate this difficulty, a normalization of the image grey-level distribution with a unique global classification threshold was tested in [49] with a rather limited success. As opposed to this signal-driven approach, we propose in this section to make use of global reference data to infer the optimal classification threshold per image.

In all the cases described below, we applied first a binary mask on the input image I in order to exclude noise or misleading information:

$$Mask_{[i,j]} = \begin{cases} 1, & \exists I_{[i,j]} \wedge Cloud_{[i,j]} \wedge OSM_{[i,j]} \\ 0, & \text{otherwise} \end{cases}$$

where $Cloud$ and OSM refer to the computed cloud mask and the OpenStreetMap land mask, respectively. All these images have the same dimensions, and the indices i and j refer to rows and columns, respectively. Then, the PANTEX X is adjusted as follows:

$$X_{[i,j]} = \begin{cases} X_{[i,j]}, & X_{[i,j]} \times Mask_{[i,j]} > T \\ 0, & \text{otherwise}. \end{cases}$$

The computation of the threshold T is differentiated in each of the following cases.

1) PANTEX Learning on LandScan: The parameter set '142' of the workflow includes the learning of image-derived PANTEX best thresholds for the discrimination of BU from NBU areas from LandScan population density reference layer. Let the image of the density of people be denoted by y and the PANTEX feature existing in the bounding box of y be denoted by x. Both datasets have different resolutions, such that one pixel q of y encompasses several pixels $q^x = \{r_i\}$ of x. Examples of both images are given in Fig. 7(B) and (C). While the spatial correlation is obvious, the links between both information must be modelled carefully. As the uncertainty exists about the information of y which can be fused into x, a statistical model is employed.

It is well admitted that the BU spatial density increases with the density of people. Also, the BU density reaches a saturation s after that the population density exceeds a break point. Indeed, the human settlements grow vertically after having filled the earth surface to accommodate more and more people. This observation led to a model between the BU density π_{BU} bounded by s and the density of population in a cell q:

$$\pi_{BU}(y(q) \mid \lambda, s) = s(1 - e^{-y(q)/\lambda})$$

where λ determines the population density breakdown point. The breakdown point λ depends from the world location, but it can be inferred from the observations x and y in the specific scene under processing.
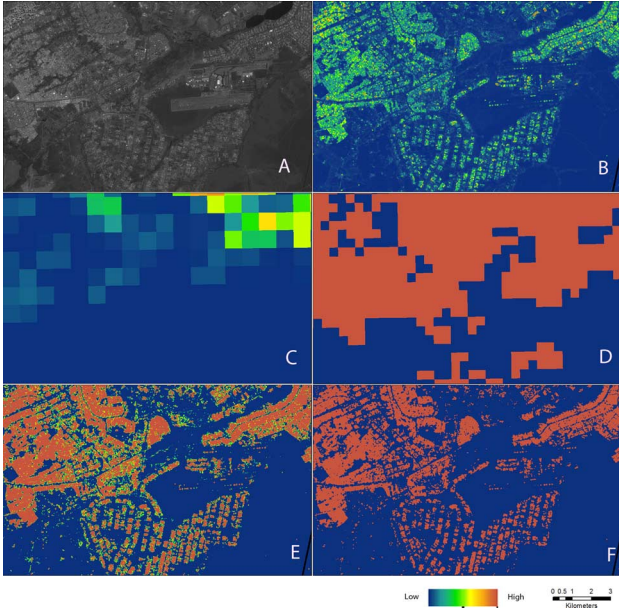
Fig. 7. Learning the best HR PANTEX rescaling from low-resolution references. (A) input image, (B) PANTEX feature at DFID 10 m resolution, (C) population LandScan data 1 Km resolution, (D) MODIS Urban data 500 m resolution, (E) PANTEX rescaled according to LandScan (param.set 142), (F) PANTEX rescaled using MODIS500 reference (param.set 146).

By analyzing the formation of the PANTEX measurement, a statistical model of the realizations of x is selected. Assuming that the input image grey levels follow an identical Gaussian distribution, and assuming a given displacement vector linking pairs of input image pixels, the differences (contrast) between the pairs of pixel values follows also a Gaussian distribution. The GLCM contrast textural measurement processed in the PANTEX feature is described in the Section VII-A. It is a linear combination of the squared local differences, such that it follows a χ^2-distribution or more generally a Gamma distribution. Finally, the PANTEX is a nonlinear merge of those measurements but does not allow for a close distribution form. Experimentally, the PANTEX values distribution is well approximated by a Gamma distribution in pure BU or pure NBU areas. Thus, the distribution of the PANTEX values can be modelled as a mixture of Gamma distributions in an unknown environment: $p(x(r)) = P_{BU} p_\Gamma(x(r) \mid \theta_{BU}) + P_{NBU} p_\Gamma(x(r) \mid \theta_{NBU})$, where $P_{BU} + P_{NBU} = 1$ are the prior probabilities of BU and NBU, $x(r)$ is the PANTEX measure at location r, and θ_{BU} and θ_{NBU} are the Gamma distribution parameters in the BU and NBU areas.

The prior P_{BU} can be assimilated as a density of BU by assuming uniform spatial distribution of buildings inside the given spatial unit under analysis, and it is approximated by π_{BU}. By injecting, the information from the density of population, the probability distribution of the pixels $q^x = \{r_i\}$ falling into q is structured as follows:

$$p(x(r_i) \mid y(q), \lambda, s, \theta_{BU}^q, \theta_{NBU}^q)$$
$$= \pi_{BU}(y(q) \mid \lambda, s) p_\Gamma(x(r_i) \mid \theta_{BU}^q)$$
$$+ (1 - \pi_{BU}(y(q) \mid \lambda, s)) p_\Gamma(x(r) \mid \theta_{NBU}^q)$$

where the Gamma distribution parameters depends on the location q.

In summary, the joint probability of both observations x and y is structured and parametrized in the following way:

$$p(x, y \mid \lambda, s, \{\theta_{BU}^q, \theta_{NBU}^q\})$$
$$= \prod_q p(y(q)) \prod_{r \in q^x} p(x(r) \mid y(q), \lambda, s, \theta_{BU}^q, \theta_{NBU}^q) \quad (1)$$

where $p(y(q))$ is the distribution of the density population values which can be approximated by histogram. The parameters λ, s, $\{\theta_{BU}^q\}$, $\{\theta_{NBU}^q\}$ of this model can be inferred by Expectation Maximization [68]. No spatial constraints are imposed in this model, while the data seem to be spatially correlated. To gain in robustness, the multi-dimensional image formed of the parameters θ_{BU}^q, θ_{NBU}^q is low-pass filtered at each iteration of the Expectation Maximization. This process benefits from the context, giving higher confidence in the estimates. While the parameters capture the link between both data x and y, it also embeds crucial information for classifying the pixels of x into BU or NBU. In particular for a pixel r belonging to a cell q, its classification is given by $\arg\max_{A \in \{BU, NBU\}} \pi_A(y(q) \mid \lambda, s) p_\Gamma(x(r) \mid \theta_A^q)$.

Finally, this method allows to make use of an additional source of information presented as a density of population in order to derive location adaptive decision for detecting BU in unknown environments. As the density of BU is not derived from the observation x only, the algorithm does not fall in unwanted local maxima.

2) PANTEX Learning on MODIS500: In this study two options were tested using the MODIS500 'urban areas' as reference. The MODIS500 was generated by automatic classification of MODIS multispectral data using regression trees techniques and dedicated models fitting with specific characteristics of the different ecological regions [6]. The source has the main advantage of being the most detailed and the most consistent global information layer available today being produced from the same sensor and the same methodology. Moreover, some work on validation of this source was also reported in [6]. Nevertheless, two main drawbacks were expected in using this source for training of the HR image descriptors used in this study: i) the resolution of the sensor and the techniques used for image information extraction would presumably underestimate the detection of scattered and/or vegetated settlement patterns, introducing a bias in favor of non-vegetated compact and large settlements, ii) during the integration phase problems in the geo-coding of the source were reported with estimated error greater than 2–3 times the pixel size (1.5 kilometers) respect to the HR,VHR data available.

The options included in the '145' and '146' parameter sets implemented two methods for learning from the MODIS500 source: respectively, 'by area matching' and 'by ROC optimization'. The option 'by area matching' minimizes the difference between the total surface of the BU areas estimated in the reference and in the PANTEX-derived layers, by exploring a range of thresholds in the PANTEX image features. The threshold in this case is computed as follows:

$$T = \min |\mathbf{c} - \mathbf{NBU}|$$

where **c** denotes the cumulative sum of a 10-bin histogram over the PANTEX values and $\mathbf{NBU} = \sum_i \sum_j Mask_{[i,j]} - \sum_i \sum_j MODIS_{[i,j]} \times f$ denotes the estimation of pixels that may indicate NBU cells; the user-defined parameter f is a factor that expresses the expected inter-scale bias in detection of BU areas originated by generalization issues. In the experiment this parameter was set as $f = 0.5$. The option 'by ROC optimization' instead selects the PANTEX threshold T that minimizes the Equal Error Rate (see Section IX-D) estimated in the ROC analysis using the MODIS500 as reference. While the option 'by area matching' requires only spatial consistency at level of scene, the option 'by ROC optimization' would theoretically require spatial consistency at least greater than half of the reference pixel size, then 250 meters. In both cases the assumption is that the majority of the information in the scene is correctly represented by the reference layer; nevertheless, the method is robust against large deviations of the reference, empirically estimated in the order of 30% of variations admitted in the LR reference data. Fig. 7(A), (B), (D), (F) show a sample of input image at 2.5-m resolution, the image-derived PANTEX feature aggregated at the DFID of 10 m resolution, the corresponding MODIS500 data, and the PANTEX feature after the learning on MODIS500.

C. Adaptive Learning on Morphological Image Features

The morphological characteristics of the image information used in the workflow are formalized through the CSL model storing in compact way multi-scale morphological image decomposition.

The 'C' or *characteristic* layer of CSL includes the scale and the (opening, closing) domain of morphological response [20]. In other words: it is stored here the best recognized estimation of the size of the image structures and their behavior (brighter, darker) respect to the relative background. It is a double scale-space decomposition in respect to the original image and their inverse. This layer is invariant to multiplicative and additive transforms of the input image data, and then doesn't require learning and/or standardization steps also if used with non-calibrated heterogeneous image data input.

The 'S' or *saliency* layer of the CSL model instead reports about the amount of contrast explained by the specific scale transform collected by the characteristic layer. In other words, it is the amount of confidence – based on the available contrast between the image structure and their background – that can be given to the specific scale inferred by the CSL model for this given structure. From another point of view, this saliency can be interpreted as a measure of the fitting of the image structures respect to a specific image information query defined in by a range of scales of the target structures. In this sense the saliency plays an important role in driving the integration between morphological descriptors and other image features for optimization of the image information discrimination.

The 'L' or *leveling* layer of the CSL model was designed in order to keep the radiometrical information of the image: in the levelling it is stored the residual contrast of the image not explained by the range of scales included in the morphological decomposition. By composition of the levelling and the other

two CSL layers, it is possible to reconstruct an approximation of the original input image without having access to the whole original multi-scale decomposition, then introducing a compression in the morphological multi-scale image features. Being the levelling designed for exploitation of the radiometrical image descriptors it is not used in this study where those descriptors are not exploited for discrimination of BU/NBU areas. In this workflow it is simply discharged

Being the saliency linked to the contrast of the image (as the levelling not used here), it is obviously very sensitive to multiplicative transforms of the input data and consequently if calculated from non-calibrated heterogeneous image data input, needs to be standardized before integration with the other information layers. This is the role the adaptive learning implemented in the workflow discussed here. Several options are under study and available in IQ GHSL workflow parameter sets, combining available reference layers at different scales and extrapolating typical saliency behaviors in BU and NBU reference areas. During this study those options were not benchmarked and consequently they are not discussed here.

During this study, the adaptive learning of the saliency layer was performed by observation of the average and standard deviation values of the saliency in the AOI of the image defined by the BU reference areas if they were available. If they were not available, as typically in the case of the 'discovery' learning and classification phase, as BU reference area it was used the one inferred by the PANTEX adaptive learning. In this case an internal retro-action mechanism between textural and morphological image information descriptors was established, while usually the approach taken here maintained two parallel independent learning paths for the two descriptors.

D. Information Fusion

During this phase, the image descriptors standardized during the learning phase are merged in order to produce the final information layer. This is done accordingly with the '*image information query*' defined by the user at the beginning of the process. In the current experiment, three options were implemented: i) intersection by closing, ii) intersection by reconstruction, and iii) simple intersection. The *ptx*, *slc* being the PANTEX and CSL saliency layers, respectively, after the learning step optimizing the BU/NBU recognition with a target resolution fitting the 1:10K GHSL specs, and $SE \in [10, 50, 500]$ being the structuring element with a scale corresponding to the spatial displacement admitted in the GHSL technical specs for the local, regional and global scales, respectively. The 'intersection by closing' is calculated as $y = \phi_{50}(ptx) \cap slc$, while the intersection by reconstruction is calculated as $y = \text{recon}(slc, \text{marker} = ptx)$. The simple intersection option is given by point-wise $y = ptx \cap slc$. In the results discussed here this option is not benchmarked and only the 'intersection by closing' was selected after some tests using the input scenes where visual reference data was available. In these tests, this option provided the best compromise between omission and commission error rates generated by, respectively, the 'simple intersection' and 'intersection by reconstruction' options.

E. Multi-Scale Generalization Protocol

The multi-scale generalization protocol is used by the IQ system in order to manage the trade-off between the precision and the computational cost of the multi-scale information representation, from local to global spatial units. While the most accurate way to summarize from local to global scales is often an aggregation based on local statistics (mean, stdev), this may have a prohibitive computational cost if applied to global high-resolution datasets served using standard WMS/TMS technologies and protocols. This fact may force to pre-calculate and store all possible image information extraction outputs before the user query, then reducing dramatically the system interactivity and increasing drastically the required information storage volume. In order to mitigate these issues, in this prototype the output geo-information at the local scale is prepared already at different successive generalization scales compacted in the same layer by the means of the multi-scale generalization protocol. In this way, this compact layer can be spatially queried at different scales, by using fast computational resampling algorithms as for example the 'nearest-neighbor' algorithm. This will provide a fast approximation of more computationally expensive and more precise aggregation options.

In Fig. 8 we show a sub-sample of the city of Sanaa, Yemen, used to test the generalization and multi-scale composition options discusses here. In Fig. 9, we show the effect in the global representation of the different generalization and multi-scale composition options. (Fig. 9(a)) the 'best' representation of the city of Sanaa, by aggregating building footprints from local to global scale using local average operator. (Fig. 9(b)) The same city as represented in the global scale by using nearest-neighbor resampling of the building footprints at scale 1:10K. Note how information is degraded especially in the small and scattered BU structures and urban fringes. (Fig. 9(c)–(e)) show the same global-scale representation of Sanaa, generated by nearest-neighbor resampling of the building footprints processed with the proposed multi-scale generalization protocol. Note how these representations better match the 'reference' one made by local average (Fig. 9(a)), while using much faster spatial resampling techniques.

Let the IQ_{BU10K} be the result of the image information query with a target resolution fitting the 1:10K GHSL specs, and let $SE \in [10, 50, 500]$ be the structuring element with a scale corresponding to the spatial displacement admitted in the GHSL technical specs for the local, regional and global scales, respectively.

Three options are implemented in the IQ system for handling spatial generalization i) by dilation, ii) by closing, and iii) by hybrid approach.

They are formalized as follows:

- by dilation:

$$IQ_{BU50K} = dilate_{SE(50K)}(IQ_{BU10K}),$$
$$IQ_{BU500K} = dilate_{SE(500K)}(IQ_{BU10K});$$

- by closing:

$$IQ_{BU50K} = \phi_{SE(50K)}(IQ_{BU10K}),$$
$$IQ_{BU500K} = \phi_{SE(500K)}(IQ_{BU10K});$$

- by hybrid approach:

$$IQ_{BU50K} = dilate_{SE(50K)}(IQ_{BU10K}),$$
$$IQ_{BU500K} = \phi_{SE(500K)}(IQ_{BU50K}).$$

In the current experiment, only the hybrid approach was tested providing a compromise optimizing the target scale of 1:50K that was the focus of the test. Because of the needs of compression of the output mosaic and simplification of the WMS infrastructure delivering the final information during the prototypal design, the three GHSL scales are fused in only one layer. The fusion is performed by the summing of the three scales so that $BU_{out} = BU_{10K} + BU_{50K} + BU_{500K}$.

IX. QUALITY CONTROL AND VALIDATION

In this section we describe the techniques applied for quality control (QC) and validation during the experiment. As general rule, QC of any given information layer should be done by systematic comparison against the same information provided by an independent source of equal or better characteristics respect to the information under test. For geo-information this means equal or better scale detail and equal or better thematic reliability, the two aspects being interlinked by the generalization issues. For the information discussed in this paper, this would mean the access to a hypothetical global topographic source having 1:50K or more detailed scale and validated thematic contents compatible with the information under test. We recall here that GHSL thematic specifications require the detection of single, even if isolated BU structures. At the time of the running of the experiment, no such source was available for different reasons including license and confidentiality issues, and technical issues mostly related to completeness and semantic consistency of the reference information with respect to the GHSL. This is typically the case of OSM sources that were evaluated and rejected at the time of the experiment design. On the other side, a standard sampling approach exploiting a pure visual interpretation validation schema would face the following difficulties in the discussed operational conditions: i) large number of samples to be collected and unacceptable estimated time cost (several person years), ii) largely unknown universe (fine-scale settlement patterns) to be stratified at the scale under analysis, including the combination of all the collection parameters (sensor, season, geographical place) potentially influencing the final output.

A. Validation Strategy

The strategy designed for solving the GHSL validation is based on active linking of two different measurement sets, namely: *accuracy measures* using visually collected reference samples, and ii) *consistency measures* using LR global reference sets. The strategy is inspired to the artificial intelligence 'active learning' approach also called 'optimal experimental design' in the statistic literature, translated to the global multi-scene evaluation case [69]. In particular, an iterative process exploiting inter-scale information systematic comparison is established in order to maximize the impact of a minimized number of visually labelled samples. The process is summarized next:

- T_0: initialization
  1) Ca_0 collection of a fist arbitrary set of visual reference data with the correct scale and semantic,
  2) Cb_0 collection of complete and consistent broader scale global reference data having similar semantic,
  3) M ranking of the best Cb-derived agreement measurements using Ca as reference
  4) A_0 Asses $M(Cb_0)$ over the whole universe under study and rank agreement and anomalies
- T_1: first release
  1) Ca_1 collection of visual reference data with the correct scale and semantic in the extreme disagreement area of A_0
  2) Cb_1 improve Cb_0 by integration of Ca_0 labelled samples
  3) A_1 Asses $M(Cb_1)$ over the whole universe under study and rank agreement and anomalies
- T_n: nth release
  1) $Ca_{(n)}$ collection of visual reference data with the correct scale and semantic in the extreme disagreement area of $A_{(n-1)}$
  2) $Cb_{(n)}$ improve $Cb_{(n-1)}$ by integration of $Ca_{(n-1)}$ labelled samples
  3) $A_{(n)}$ assess $M(Cb_{(n)})$ over the whole universe under study and rank agreement and anomalies

Assuming similar thematic contents and globally consistent behavior of the broad scale reference data, the expectation is that the low agreement areas are the most interesting for optimization of the visual labelling activities. In these areas we concentrate both the errors of the information under test and the errors or generalization effects of the broad scale reference data. Integrating the samples visually labelled at the iteration n in the global reference data used at each iteration $n + 1$, would step-wise increase the overall reliability of the reference data and then the overall reliability of the derived quality measurements. Moreover, this mechanism will decrease the probability to select the same areas as priority to be visually analyses.

At the time of the experiment reported here, only the initial iteration of the validation process was implemented and it is consequently reported in this paper.

B. Visual Reference Data

During the study the accuracy measurements are derived exclusively from comparison of the GHSL output with visually interpreted samples using HR/VHR input data sources. The visual reference data collection was done by implementation of the GHSL 50K dichotomic validation protocol [70] that includes the following steps: i) collection of spatial samples by a random systematic grid procedure and ii) interpretation of each sample by visual inspection of the corresponding part of image. The adopted sampling approach uses an absolute origin in the global projection, and then applies a systematic grid step that in this case is equal to one kilometer. For each sampling grid step, a cell with a fixed size is selected that in this case was equal to 200×200 meters. Each sample cell is subdivided in 4×4 sub-samples with the size of 50×50 meters, fitting the size and location of the GHSL 50K 'regional' spatial units. Given the

adopted dichotomic protocol, the interpreter was asked to check if the specific sub-sample was intersecting a visible building in the image with four possible coded answers : yes, no, I'm not sure, and no data available. Only the simple yes–no answers were used in this study, and only images with 2.5 meter input resolution or better were considered suitable candidate for this validation protocol. Cross-comparisons of parallel validation campaigns done by different operators on the same set of images were used in order to control the consistency and reliability of the human interpretation task.

For this experiment, approximately 95,000 built-up (BU) and 700,000 not-built-up (NBU) samples were visually collected using the adopted protocol. The samples were collected mainly from 2.5-m-resolution, pan-sharpened SPOT 5 and 0.5-m-resolution aerial RGB imagery. The total ground surface processed employing this visual interpretation protocol was over 700,000 km^2.

C. Global Reference Data

The global reference data set (BUREF) is used for ranking of the GHSL output alternatives and for optimization of a given user-defined function, that in this case maximize consistency between GHSL outputs and LR reference data. After the evaluation of several potential alternatives, the BUREF has been built by the combination of the MODIS 500 Urban Land Cover (MODIS500) and the LandScan 2010 layers. In particular, the BUREF used in this experiment was derived from LandScan population layer 2010 by rescaling it between $[\min, \max]$ thresholds expressing the amount of population for NBU and BU areas, respectively. The model assumes an increasing positive relation between the probability to be BU and the number of people living in a given area. In order to estimate these thresholds, MODIS500 was used as training set discriminating between urban and not-urban subsets in the LandScan data. The $[\min, \max]$ parameters are estimated locally with a sliding window size of 500 kilometers. In order to increase spatial consistency of the model, an overlapping parameter of 50% in the sliding window size was applied.

D. Quality Measurements

In this study, the same quality metrics are used for accuracy analysis and for global consistency checking, by changing the reference data used. In particular, accuracy analysis is done by application of the quality metrics using visually collected samples as reference, while global consistency is measured using BUREF as reference. Given continuous measurements $\{x_i\}$ and their respective classes $y_i \in \{NBU, BU\}$, their classification power is best captured by the Receiver Operational Characteristics (ROC) [71], which represents the probabilities of missed detection and false alarm for various classification thresholds:

$$p_{md}(T) = \frac{\sum_i 1(x_i < T)1(y_i = BU)}{\sum_i 1(y_i = BU)}, \qquad (2)$$

$$p_{fa}(T) = \frac{\sum_i 1(x_i \geq T)1(y_i = NBU)}{\sum_i 1(y_i = NBU)}. \qquad (3)$$
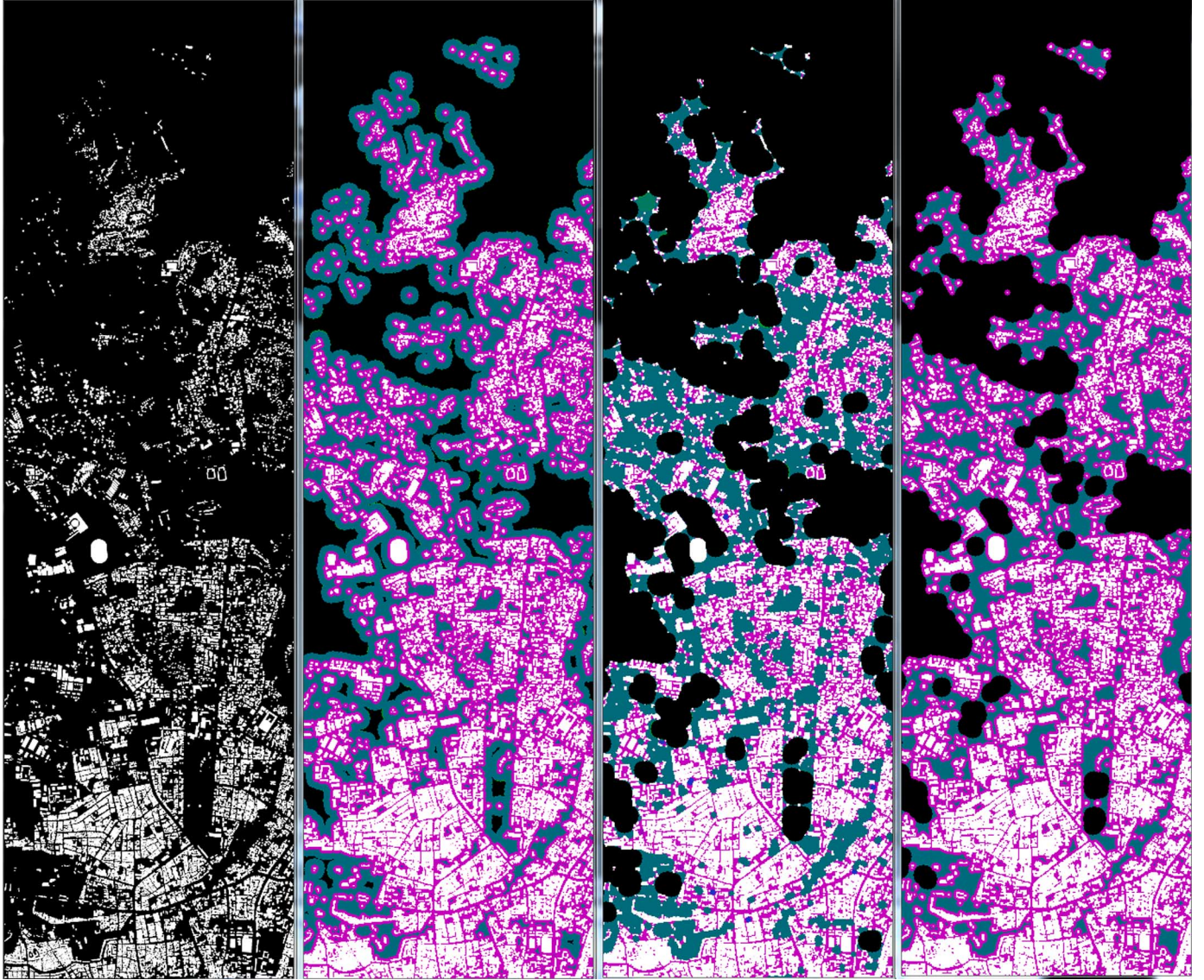
Fig. 8. Test of the generalization and multi-scale composition options in the city of Sanaa, Yemen. From left to right: i) in white building footprints at 1:10K scale, generalization by ii) dilation, iii) closing, and iv) hybrid approach. White, pink, and green show, respectively, the contribution of the 'local', 'regional', and 'global' scales to the final GHSL product.

The ROC curve is the parametric function $(p_{md}(T), p_{fa}(T)), \forall T$, which is a convex. In order to compare two types of measurement producing two ROC curves, the metrics of the Area Under the Curve (AUC), the Equal Error Rate (EER), and the Minimal Error Rate (MER) are proposed and can be interpreted as quality statistic:

$$AUC = \int_0^1 p_{fa} dp_{md}, \tag{4}$$

$$EER = p_{md}(T_e) = p_{fa}(T_e), \exists T_e \tag{5}$$

$$MER = \min \left[p_{md}(T_e) + p_{fa}(T_e) \right], \exists T_e. \tag{6}$$

The smaller these quality metrics are, the better is the classification or discrimination performance of the type of measurement.

After the evaluation of several alternatives by regression analysis, the Minimal Error Rate (MER) measurement using BUREF as reference was the one best approximating the agreement obtained using the available visual reference data [72].

Consequently, this was the measurement adopted for the global evaluation of the GHSL output during the initialization phase of the validation process reported in this paper.

X. RESULTS

In this section, we discuss the results of the proposed automatic image information retrieval process respect to two basic perspectives, namely i) accuracy measurements and ii) consistency analysis. The first is supported by visual reference data collection using HR/VHR image data input, while the second is supported by global LR reference layers. The same GHSL output discussed here has been evaluated also in other more specific tests not included in this paper. In particular, the GHSL output of 628 SPOT satellite scenes covering the major urban agglomerations in Europe was systematically compared with the high resolution European Soil Sealing Layer produced by
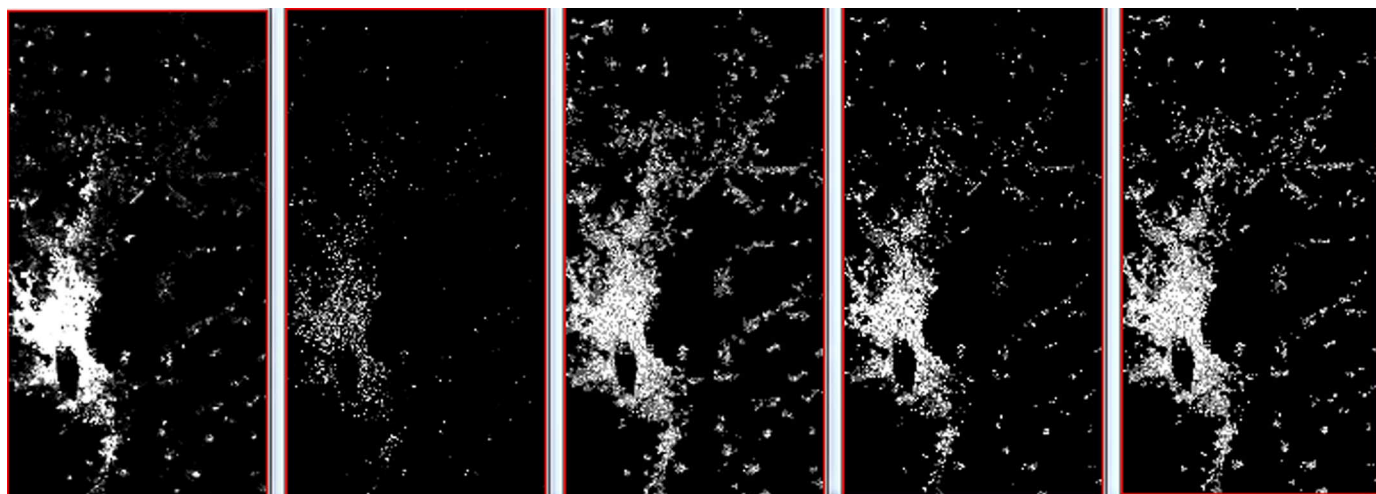
Fig. 9.   Test of the effect in the global representation of the different generalization and multi-scale composition options of the GHSL outputs in the city of Sanaa, Yemen. From left to right: 1) building footprints from cadastral maps 1:10K aggregated to the global GHSL scale using local average; 2) the same using fast 'nearest-neighbor' resampling algorithm, 3) using 'by dilation', 4) 'by opening' and 5) 'hybrid' generalization options and also fast 'nearest-neighbor' resampling algorithm.

the European Environment Agency[18]. The test provided an average 90.8%±3.9 of average agreement rate between the two sources [73]. In Brazil, a stratified random sampling procedure and visual reference data collection was applied to evaluate the GHSL output of more than 3000 input CBERS scenes [74]. The assessment provided an average agreement rate of 94%±6. Finally, a systematic comparison between the GHSL output of 2288 input CBERS scenes and the land cover of China derived from Landsat data was performed [75]. This test provided an average 98.13%±5.6 of agreement rate in the best of the benchmarked parameter sets.

Fig. 10 demonstrates a typical example of the type of information extracted from remote-sensing images in this experiment. The same scene is identified in two other available global sources that are used as a reference for visual comparison. The top-left image shows the "presence of buildings" layer of the GHSL over the city of Brasilia. It is produced from image data by the CBERS 2B sensor and the output image is shown at a 1:50K "regional" scale specification. Pixel brightness is proportional to the percentage of BU presence in the specific spatial units. Dark-green polygonal contours correspond to the footprints of satellite scenes that were used to make up the final mosaic. The final result for each spatial unit is made of the "best" information extracted from all the available processed scenes. The top-right image shows a zoom into the central region of the same city. The image shows the "average building size" GHSL information layer at a 1:10K "local" scale specification. Color-coding follows the blue-green-yellow-red order on increasing size of the BU structures. The bottom-left and right images show the same city represented in the MODIS500 and BUREF reference layers respectively, at the 1:50K "regional" scale specifications.

Comparing the output of the GHSL produced from HR image data against the other two low resolution layers, a noticeable

---

TABLE V
ERROR RATES (EER AND MER) FOR THE GHSL OUTPUT (BUFINAL) AND THE VISUAL REFERENCE DATA FOR EACH PARAMETER SET.

Data	Parameter_script_ID			Grand Total
	142	145	146	
Sum of input surface (km^2)	733,560	727,348	731,181	2,192,089
Sum of BUfinal (visual) number of positive samples	96,064	95,515	95,758	287,337
Sum of BUfinal (visual) number of negative samples	701,030	688,692	693,165	2,082,887
Average of BUfinal (visual) EER	17.26%	17.42%	18.43%	17.70%
Average of BUfinal (visual) MER	9.15%	8.57%	9.74%	9.15%

gain is observed with respect to the detail of the extracted information. Moreover, the quality of the GHSL scenes offers the option of addressing the internal characterization of BU areas by automatic analysis of the morphological characteristics of the BU structures. Recognition and characterization of BU structures is evidently not possible using low-resolution input images.

A. Accuracy Analysis

The accuracy of the GHSL output produced during the experiment was estimated by using a total of approximately 95,000 and 700,000 samples of BU and NBU classes, respectively, collected by visual interpretation of HR/VHR input images. The collection of these samples was following a random systematic grid approach as defined in the GHSL reference data collection protocol (see Section IX-B). The results of this analysis are presented in Table V. The first rows describe the surface and the number of available reference data sets. Due to some failure in the processing of specific parameter sets in specific scenes, there is a difference on the number of samples per set. The error rates
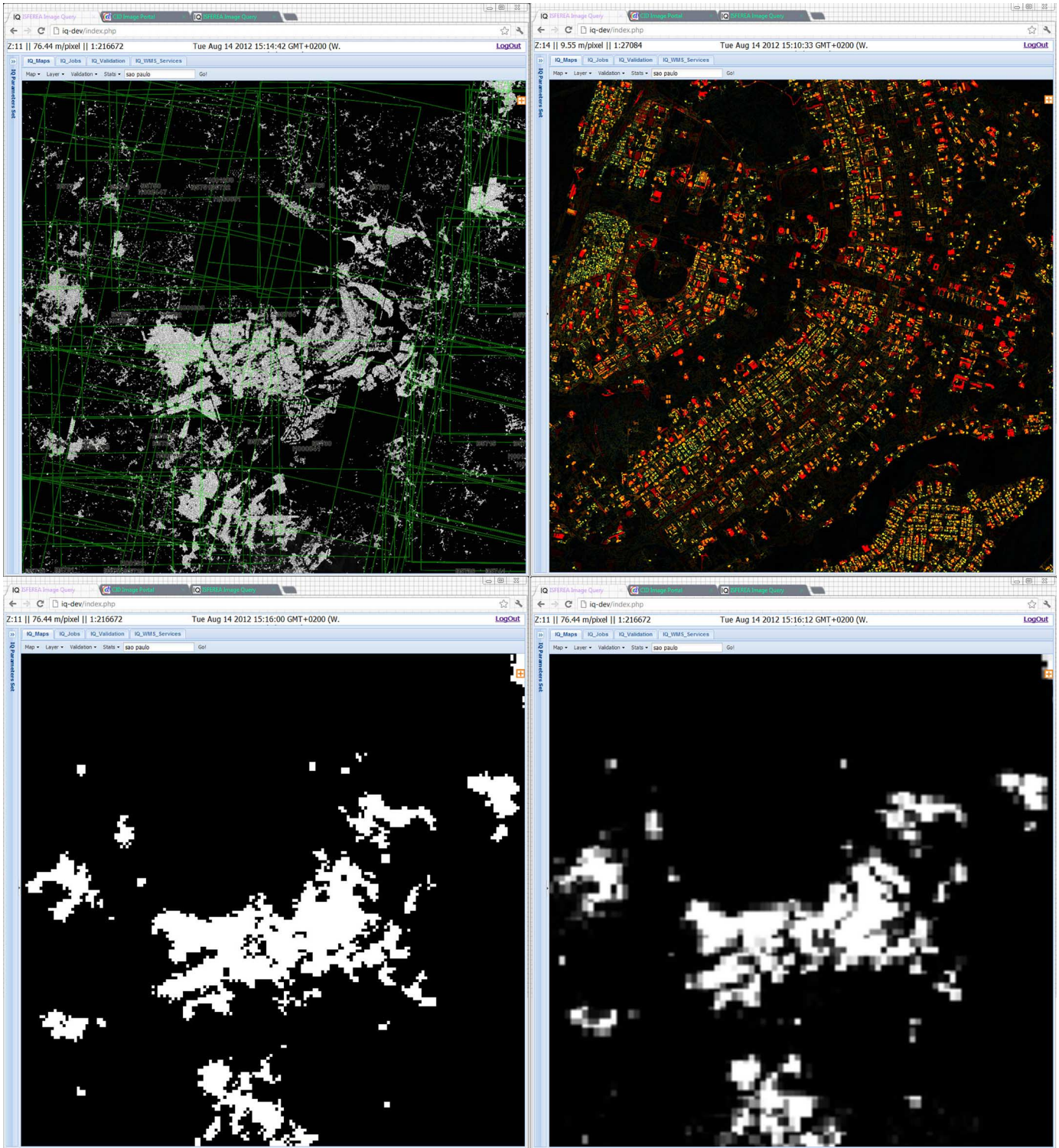
Fig. 10. City of Brasilia. top-left: the "presence of buildings" GHSL layer represented at 1:50K scale with the footprints of input scenes (CBERS 2B) in dark-green; top-right: a zoom into the city center. The image shows the "average building size" GHSL layer at 1:10K scale. Increasing BU size is mapped on the blue-green-yellow-red color map; bottom-left and right: the same city represented by the MODIS500 urban layer and BUREF respectively.

are listed with respect to three different parameter sets bench-marked during this study: namely 142, 145, and 146. Each relates to a set of options activated in the learning and classification steps as described in Section VIII.

The adopted validation protocol show an estimated EER and MER of the final GHSL built-up layer of approximately 17% and 9% respectively. This corresponds to a total accuracy of more than 90%, i.e., (1-MER). All three learning-and-classification options evaluated using the proposed protocol appear to be consistent and have only minor differences in performance. Option 142 ranks best with respect to the EER metric (17.26%), and option 145 performs best with respect

TABLE VI
ERROR RATES (EER AND MER) FOR THE GHSL OUTPUT (BUfinal) AND THE
GLOBAL REFERENCE DATA FOR EACH PARAMETER SET

Data	Parameter_script_ID			Grand Total
	142	145	146	
Average of BUfinal ROC$_{refBU}$ EER	29.84%	36.57%	37.06%	34.48%
Average of BUfinal ROC$_{MODIS500}$ EER	28.64%	30.08%	31.32%	30.01%
Average of PANTEX ROC$_{refBU}$ EER	34.23%	38.91%	38.92%	37.34%
Average of PANTEX ROC$_{MODIS500}$ EER	33.04%	33.03%	33.03%	33.03%
Average of BUfinal ROC$_{refBU}$ MER	7.93%	13.37%	13.61%	11.63%
Average of BUfinal ROC$_{MODIS500}$ MER	8.94%	9.02%	9.19%	9.05%
Average of PANTEX ROC$_{refBU}$ MER	8.17%	13.78%	13.72%	11.88%
Average of PANTEX ROC$_{MODIS500}$ MER	9.26%	9.24%	9.24%	9.25%

TABLE VII
AVERAGE BUREF AGREEMENT BY SENSOR

Sensor	Values	
	Average of $Q(x)$	StdDev of $Q(x)$
RAPIDEYE 4	97.94%	2.27%
AERIAL	95.66%	3.69%
CBERS 2B	94.37%	7.59%
SPOT 5	93.26%	7.33%
RAPIDEYE 1	90.26%	10.10%
SPOT 2	88.21%	3.10%
IKONOS 2	86.63%	12.67%
RAPIDEYE 2	86.38%	10.36%
WORLDVIEW 2	84.37%	10.52%
QUICKBIRD 2	83.58%	13.04%
GEOEYE 1	80.98%	11.58%
WORLDVIEW 1	80.03%	14.28%
Grand Total	91.50%	10.08%

to the MER metric (8.57%). Option 146 has the lowest rank in both metrics.

In order to allow an assessment of the feasibility to use the global reference layers for the quality assessment, the two quality metrics were also computed for the MODIS500 and the BUREF data sets taking into account only the areas, where also visual reference information is available for the same satellite input imagery and using the same visual validation protocol. The results are listed in Table VI.

Despite minor differences that are being investigated further, two main observations are made: i) the quality metrics computed using the visual protocol and using the low-resolution reference layers appear to be substantially consistent and ii) there is an almost systematic overestimation of the average error in both error metrics using the low-resolution references, if compared against the corresponding metrics using the visual references, which we assume to be more reliable than the low-resolution reference. The first observation supports the use of global, low-resolution reference layers for an automatic evaluation and ranking of the automatic image information extraction output. This is backed by obtaining almost the same ranking, if benchmarking the three learning-and-classification options on low-resolution reference layers, or if using the much more expensive visually collected reference data.

The second observation refers to a general issue when using LR reference layers for evaluating HR classification outputs. Any disagreement measure between HR and LR reference set always integrates two components, namely: i) error issues in both sources and ii) scale and generalization issues. This second type of mismatch can be deemed 'added value' of the HR, VHR automatic recognition procedure with respect to the available LR information layers.

The above observations lead to two different findings: i) a positive one that stems from the fact that we can expect a very conservative assessment of the global GHSL output quality using agreement measures respect to LR reference data, and ii) a negative one that can be linked to the risk of discarding GHSL outputs showing low agreement metrics, but in fact

providing a dominant 'added value' with respect to the LR references. In the 'discovery' phase of the learning and classification processing chain we try to mitigate this risk in the cases where LR references show clearly unreliable behavior; typically in cases of low-density scattered settlements, which cannot be detected with LR sensors such as MODIS. In any case, it is expected that the GHSL general validation strategy (see Section IX-A) would step-wise contribute to separate these two kinds of disagreement and consequently increase the overall reliability of the GHSL quality control procedure.

B. Global Consistency Statistics

Tables VII–IX show the estimated average consistency of the whole GHSL output produced during the experiment using the MER measure and the 1-km-resolution BUREF layer as reference. This is the quality metric that showed the best matching with the visual validation protocol figures [72]. The tables show the analysis of the results ordered by the type of sensor, by band, and by sensor resolution, respectively. The agreement metric reported in these tables is defined as the inverse of the error metric: $Q(x) = 1 - MER_{BUREF}(x)$. In all three tables the results are listed by decreasing values of the $Q(x)$ metric; best ranks appear at the top of each list. Both average and standard deviation (stdev) of the $Q(x)$ metric are shown. All values of the $Q(x)$ metric presented in these tables represent the average computed between all the available adaptive learning-classification options (3).

1) Agreement by Input Sensor: The analysis of the results by sensor (Table VII) shows a cluster of sensors providing good BUREF fitting performances with $Q(x) \geq 90\%$. These include RapidEye 1 and 5, Aerial, CBERS-2B and SPOT 5. A medium-performance cluster of sensors with $Q(x)$ between 85% and 90% includes the SPOT 2, Ikonos 2, and RapidEye 2 sensors. A lower performing cluster of sensors, i.e., with $Q(x)$ between 80% and 85%, includes the WorldView 1 and 2, QuickBird 2, and GeoEye 1 sensors. Table VII allows for an interesting observation. In the cluster of good-performing sensors, if VHR aerial image data are excluded all the rest come from traditional HR platforms such as CBERS and SPOT if compared against the more recent VHR platforms such as WorldView and GeoEye. Aerial image data was processed at 0.5 m resolution and because of this, resolution cannot explain sufficiently the low ranking of

TABLE VIII
AVERAGE BUREF AGREEMENT BY IMAGING TYPE AND BAND

ImageType and Band	Values	
	Average of $Q(x)$	StdDev of $Q(x)$
PSH GREEN	93.56%	7.53%
PSH RED	93.54%	7.56%
PSH NIR	93.53%	7.78%
PAN PAN	93.30%	8.76%
MUL SWIR	91.33%	8.57%
MUL REDEDGE	90.18%	10.25%
MUL GREEN	86.51%	12.00%
MUL RED	86.49%	12.06%
MUL COASTAL	86.33%	9.72%
MUL NIR	86.31%	12.44%
MUL YELLOW	86.16%	9.89%
MUL BLUE	85.45%	12.42%
MUL NIR2	85.17%	11.09%
PSH REDEDGE	80.03%	5.04%
PSH COASTAL	79.93%	2.35%
PSH YELLOW	79.80%	2.37%
PSH NIR2	78.08%	3.16%
PSH BLUE	75.90%	11.26%
Grand Total	91.51%	10.07%

TABLE IX
AVERAGE BUREF AGREEMENT BY CLASS OF RESOLUTION

Resolution	Values	
	Average of $Q(x)$	StdDev of $Q(x)$
$C: \{1 < x \leq 2.5\}$	92.90%	8.87%
$E: \{5 < x \leq 10\}$	92.13%	8.12%
$D: \{2.5 < x \leq 5\}$	86.83%	12.59%
$B: \{0.5 < x \leq 1\}$	83.46%	13.27%
$A: \{x \leq 0.5\}$	81.97%	11.27%
Grand Total	91.51%	10.07%

the new VHR satellite platforms. Evidently the proposed workflow for automatic image information retrieval is influenced by other sensor characteristics than remain to be further investigated for improving the overall performance.

2) Agreement by Input Bands: Analyzing the results with respect to the estimated output $Q(x)$ divided by the different bands used as input (Table VIII), three main clusters with distinct behavior are observed: i) a top ranking cluster with $Q(x) \geq 90\%$ including the PAN and GREEN, RED and NIR bands in pan-sharpening imaging mode (PSH); ii) a medium ranking cluster with $Q(x)$ ranging from 85% to 90% that includes all listed bands in multispectral imaging mode (MUL), and iii) and a low-ranking cluster with $Q(x)$ ranging from 75% to 80% that includes all their bands in PSH mode. It is interesting to note that except for the BLUE band all other bands in the lowest ranking cluster are the new bands of the WorldView-2 satellite. This behavior of the sensor needs to be analyzed in more detail. Possible explanations could be linked to the fact that there only a limited number of 20 scenes was processed that might be influenced by other characteristics. Another source of error might be linked to the pan-sharpening of these new bands. Some algorithms are not providing good results for this new band constellation.

3) Agreement by Input Spatial Resolution: The ranking of the agreement $Q(x)$ obtained by different classes of input image spatial resolution (Table IX) shows a top performance of the class C ranging from 1 to 2.5 m spatial resolution with $Q(x)$

equal to 92.9%±8.8%. Class C consists of SPOT 5 and CBERS 2B data and all the multispectral data from the VHR sensors. The second best is class E with spatial resolution ranging from 5 to 10 m and achieving $Q(x) = 92.13\% \pm 8.12\%$. Class E consists of SPOT 1,2 data and multispectral SPOT 5 and RapidEye data. Interestingly, the worst ranking achieved according to the Q metric, is class A containing all VHR data, i.e., with image resolution of 0.5 m or better. Fine-level of details seems to decrease of the signal/noise ratio that can be observed. The reliability of the GHSL automatic image information extraction workflow under test is maximized for resolutions in the range of 1 to 2.5 m and based on this observation future releases of the workflow will be configured with a standard resolution of input imagery to approximately 2.5 m before any feature extraction.

4) Agreement by Eco-Regions: It can be assumed that similar landscapes including color and pattern of the specific land covers, may introduce similarities in the behavior of any given automatic image information recognition strategy, i.e., they can introduce a dominant characteristic of the background of the image, but they can also probably contribute in the explanation of the specific materials and patterns used to make human settlements. Consequently, it is of interest to test the robustness of a specific global automatic image information retrieval task, against bias introduced by dominant landscape patterns available at a local scale. On the other side, under given constraints it could be possible to extrapolate the performances of a given automatic image information extraction task to images representing the same landscape.

Fig. 11 shows the average $Q(x)$ obtained by the current IQ GHSL workflow after extrapolation to the WWF eco-regions. According to this analysis, Brazil, Europe and China are well placed in the high confidence area with $Q(x) \geq 80\%$ together with large parts of the Sub-Saharan and Southern Africa.

Middle East, Sahara and North African areas apparently show systematic problems with the current image information extraction strategy, producing average $Q(x)$ in the range 70% to 80%. One of the known reasons behind these poor performances is the presence of scattered vegetation and very bright soil background that might create false alarms in the textural image feature and/or miss detection in the morphological image query. Similar issues were already addressed in the same areas by a method applying morphological filtering before the image textural analysis. It demonstrated a drastic increase of the performance of the automatic recognition of BU structures in arid areas having as background bright soil and scattered vegetation [76]. This method was not implemented in the current workflow. The latter is focused more on general-purpose processes of morphological and textural features, without chaining them deductively prior to the learning phase. It is expected however that such type of observations can lead to "regionally-adaptive" image information extraction workflows taking into consideration the local and regional landscape and background conditions for each scene to be processed. This is currently under investigation.

Red regions in Fig. 11 correspond to major inconsistencies between the current GHSL outputs and the available reference layers. Such outputs are rejected during the first iteration of the IQ GHSL workflow and are ignored in the compilation of the final GHSL mosaic. According to the general GHSL validation
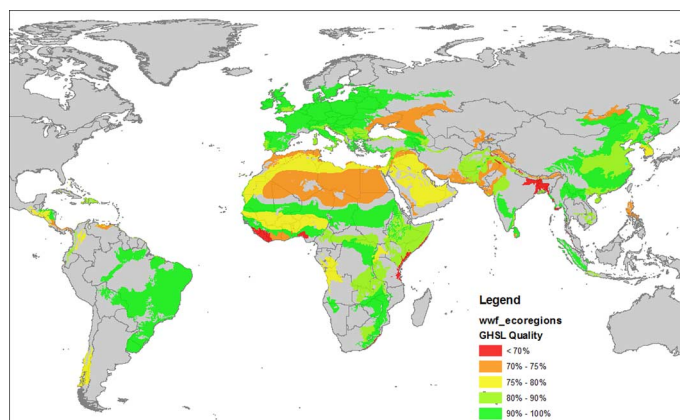
Fig. 11. Estimated BUREF Agreement of the current GHSL output by WWF eco-region.

Fig. 12. Ranking of all the scenes processed during the experiment by increasing $BUREF$ agreement optimized among all the processed options (band, learning parameters) available on the same scene. This is the process implemented for the composition of the final GHSL mosaic, taking the best of the available processed pieces of information. The blue dots represent the worst $BUREF$ agreement available on the same corresponding scenes.

strategy (Section IX-A) these areas will be included as priority for the next run of the visual reference data collection campaign and then integrated accordingly with the validation results.

C. Discovery of New Information

A total of 2895 CBERS 2B satellite scenes accounting for $3.19\,E+06\,km^2$ of ground surface where selected for testing the "discovery" modes of the IQ GHSL learning and classification workflow. For this set of scenes all adaptive learning methods failed in areas of unavailable reference information on the presence of BU areas. That is because most of these scenes are cover remote rural areas away from towns or cities. In the first iteration of the IQ workflow the BUREF layer consisted of a combination of low-resolution globally available layers (LandScan, MODIS500) in which low density rural areas are poorly represented. During the "meta-learning" phase, the statistical distributions of the best thresholds on image features were analyzed in the satellite scenes where the adaptive learning phase was providing high reliability. The satellite scenes chosen for the "meta-learning" phase where selected by a query listing all the CBERS 2B scenes where at least 20% of surface was understood as BU in the available BUREF, and producing an output with BUREF MER less than 10%. The thresholds learned from this phase were then applied blindly to the satellite scenes suitable for the "discovery" phase.

The sum of the new BU surface discovered in the HR input images during this phase was estimated as $1.00\,E+05\,km^2$. This was with an average and Stdev fitting of 98.57% and 2.42% with respect to the BUREF respectively. Fitting was measured as the inverse of the MER using BUREF as reference.

D. Mosaic Quality Assessment

According to the general approach applied in this experiment, the output of the image information extraction task is the mosaic of all the DFID cells ranked by the cost function or 'query' defined by the user (Section III-B). In this schema, multiple input image data may contribute to the same DFID cell: this is because different data collections may be available coming from different combinations of platform/sensor, date, and pre-processing parameters. Moreover, with the workflow applied here several output alternatives may be generated for each input scene according the number of bands and the
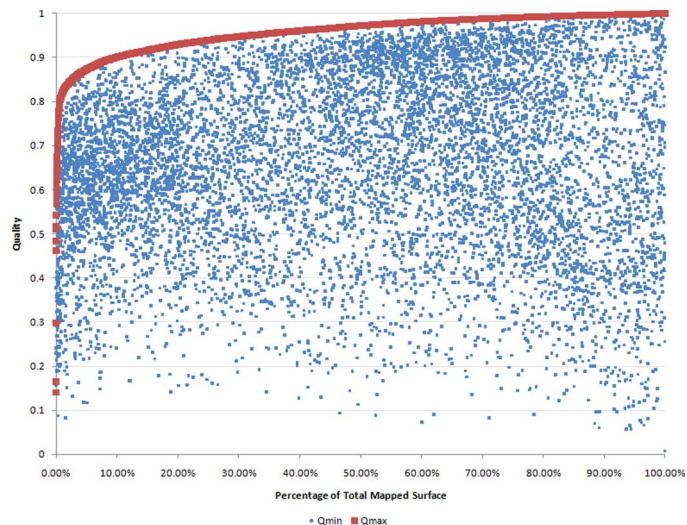
number of parameter sets benchmarked during the experiment (see Section V-D). All these output alternatives are evaluated accordingly to a global cost function expressing the general objective of the experiment. In this case, the general objective being the test of the capacity to generate globally consistent information layers from heterogeneous HR/VHR input data, a cost function maximizing the global consistency was implemented. In particular, the disagreement of the GHSL output respect to the BUREF was adopted for this purpose during the experiment (see Section IX). For each DFID cell of each output option the global cost function is evaluated, all the cell-options belonging to the same place are ranked and only the best one minimizing the cost function is retained.

Fig. 12 shows a simulation of the effect of this mechanism in the data processed during the experiment. All the input images processed during the study are represented with two dots red and blue, showing respectively the best and the worst scoring of all the available output options, according to the adopted global cost function. In the Fig. 12 they are ranked from left to right by increasing cost of the best option. The spread and density of the blue dots below the optimized red line shows the decision space explored by the system during the composition of the final image information layer. Thanks to this optimization mechanism, the final mosaic is expected to have a rejection of 10% of the mapped surface, selecting $Q(x) = 90\%$ as the minimal quality threshold.

XI. CONCLUSIONS

A. Summary of the Results

A proof-of-concept of the possibility to build a new Global Human Settlement Layer (GHSL) derived from HR and VHR optical remotely-sensed data was presented. The test involved 24.3 millions of square kilometers of test areas spread in four continents, automatically mapped with the image data collected by a variety of optical satellite and airborne sensors with a

spatial resolution ranging from 0.5 to 10 m. In this mapped area the total number of people living in 2010 was estimated to be 1,268,448,973 (LandScan). It is the largest known test of automatic image classification involving such kind of image input. Several imaging modes were tested including panchromatic, multispectral and pan-sharpened images. A new multi-scale framework was introduced, integrating the automatic image information retrieval with global available geo-information layers derived from other satellite sensors or GIS modelling. For the first time was demonstrated the capacity of automatic information extraction from remotely sensed data at detailed scale in global realistic scenarios, and the capacity to control the global consistency of the output both spatially and thematically. The robustness of the adopted image features was tested globally with a high variety of input data quality including extremely challenging "worst-case" scenarios. New multi-scale morphological and textural image feature compression and optimization methods were introduced, together with new learning and classification techniques allowing the processing of HR, VHR image data using low-resolution reference data.

The validation of the automatic results by a visual inspection protocol provided an accuracy rate of more than 90%. These results are consistent with other independent validation campaigns testing the same classification output with comprehensive reference data available in Europe [73], Brazil [74], and China [75]. The average agreement between the automatic high-resolution output generated by the experiment and the available low-resolution representation of the urban areas was estimated at 91.5%. Because of the comprehensive and systematic approach of the experiment, a comparative study across HR/VHR sensors, bands and across different geographic areas can be made using precisely the same image information extraction methodology and a consistent global reference layer. The observation of the anomalies in the global agreement ratio will focus the attention on specific sensors and specific geographical areas for further analysis, validation campaigns and methodological improvements.

B. Critical Points

Two main critical points where highlighted during the experiment, namely: i) the incapacity to exploit radiometrical descriptors – in particular band radios – as input of the classification process, and ii) the general structure of the processing flow, that was largely based on the 'scene' data granularity.

It is already known by several internal tests that the inclusion of radiometric image descriptors in the adopted classification process would increase the overall reliability of the automatic image information retrieval. In particular, the inclusion of a vegetation presence index made by band ratios would discriminate image structures falling in the textural and morphological criteria defined by the GHSL query, but not belonging to built-up structures. Typical examples could be small (less than 50-m wide) quadrangular agricultural fields showing vegetated cover: these targets are most probably contributing to the commission error of the GHSL release discussed here.

Some additional drawbacks were identified and can be related to the adopted 'per scene' processing flow structure. In particular, problems are related to the large heterogeneity in 'size' of

the scenes processed during the experiment. They were largely different both as geographical surface covered and as number of available image elements (pixels) and bands. This fact created potential instability in the learning and classification processes using the LR reference data and sub-optimal I/O performances and memory setting in the processing units. It is worth noting that more than 80% of the computational cost of the experiment was estimated as belonging to I/O operations between RAM and disk storage. Consequently the experiment highlighted that I/O operations are of foremost importance and need to be more optimized.

C. Next Steps

Next steps include the mitigation measures of the critical points listed above and the design of more refined experimental and benchmarking sets. In particular, it will be evaluated the possibility to extend the multi-scale DFID paradigm designed for learning and classification purposes to radiometric calibration purposes. This will take place in the frame of the input data scenarios discussed in the present experiment. The plan of next experimental and benchmarking activities includes i) sensitivity analysis of the GHSL inferential model against some key parameters as input image resolution and image collection parameters (sun, sensor geometry) and different learning reference set ii) testing with suitable reference data the automatic characterization of the settlement components as number and average size of buildings that were not evaluated in the current experiment, iii) testing automatic change detection techniques able to release global consistent multi-scale outputs, and iv) testing of new spatial generalization techniques able to summarize the different fine-scale human settlement patterns at the global and continental level.

Furthermore, it is expected to extend the standard GHSL processing to a wider range of input image data including i) a wall-to-wall complete European coverage of SPOT data 2012 through the GMES instrument, ii) enlargement of the available areas mapped with CBERS data in Brazil and China, and iii) activation of a process-on-demand mechanism allowing generic users to process suitable geo-coded image data using the IQ GHSL workflow. Finally, the integration of the GHSL with geospatial information extracted from other satellite platforms and active/passive sensors will be studied in the frame of cross-platform synergy.

REFERENCES

[1] M. Pesaresi, T. Kemper, L. Gueguen, and P. Soille, "Automatic information retrieval from meter and sub-meter resolution satellite image data in support to crisis management," in *Proc. 2010 IEEE Int. Geosci. and Remote Sensing Symp. (IGARSS)*, Jul. 2010, pp. 1792–1795.

[2] World Urbanization Prospects: The 2011 Revision, U.N. Dept. Economic & Social Affairs Population Div., 2012.

[3] D. Satterthwaite, "The Transition to a Predominantly Urban World and its Underpinnings," Int. Inst. Environment and Development, Human Settlements Working Paper, 2007.

[4] C. Elvidge, M. Imhoff, K. Baugh, V. Hobson, I. Nelson, and J. Safran, "Nighttime lights of the world: 1994–95," *ISPRS J. Photogramm. Remote Sens.*, vol. 56, pp. 81–99, 2001.

[5] E. Bartholome and A. Belward, "GLC2000: A new approach to global land cover mapping from earth observation data," *Int. J. Remote Sens.*, vol. 26, pp. 1959–1977, 2005.

[6] A. Schneider, M. Friedl, and D. Potere, "Monitoring urban areas globally using MODIS 500 m data: New methods based on urban ecoregions," *Remote Sens. Environ.*, vol. 114, no. 8, pp. 1733–1746, Aug. 2010.

[7] J. Dobson, E. Bright, P. Coleman, R. Durfee, and B. Worley, "LandScan: A global population database for estimating populations at risk," *Photogramm. Eng. Remote Sens.*, vol. 66, no. 7, pp. 849–857, Jul. 2000.

[8] "Gridded Population of the World,"3rd ed. Center for International Earth Science Information Network (C.I.E.S.I.N.), Columbia University, and Centro Internacional de Agricultura Tropical (C.I.A.T.), 2012.

[9] I. Baud, M. Kuffer, K. Pfeffer, and R. Sliuzas, "Understanding heterogeneity in metropolitan India: The added value of remote sensing data for analyzing substandard residential areas," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 12, no. 5, pp. 359–374, 2010.

[10] S. Niebergall, A. Loew, and W. Mauser, "Integrative assessment of informal settlements using VHR remote sensing data: The Delhi case study," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 1, no. 3, pp. 193–205, 2008.

[11] D. Ehrlich and C. Bielski, "Texture based change detection of built-up on spot panchromatic imagery using PCA," in *Proc. Joint Urban Remote Sensing Event (JURSE)*, 2011, pp. 81–84.

[12] M. Pesaresi, D. Ehrlich, I. Caravaggi, M. Kauffmann, and C. Louvrier, "Towards global automatic built-up area recognition using optical VHR imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 4, pp. 923–934, Dec. 2011.

[13] T. Toutin, "Geometric processing of remote sensing images: Models, algorithms and methods," *Int. J. Remote Sens.*, vol. 25, no. 10, pp. 1893–1924, 2004.

[14] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Analy. Machine Intell.*, vol. 22, no. 12, pp. 1349–1380, Dec. 2000.

[15] M. Datcu, H. Daschiel, A. Pelizzari, M. Quartulli, A. Galoppo, A. Colapicchioni, M. Pastori, K. Seidel, P. Marchetti, and S. D'Elia, "Information mining in remote sensing image archives: System concepts," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 12, pp. 2923–2936, Dec. 2003.

[16] C.-R. Shyu, M. Klaric, G. Scott, and W. Mahamaneerat, "Knowledge discovery by mining association rules and temporal-spatial information from large-scale geospatial image databases," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp., IGARSS 2006*, Aug. 31–4, 2006, pp. 17–20.

[17] C.-R. Shyu, M. Klaric, G. Scott, A. Barb, C. Davis, and K. Palaniappan, "Geoiris: Geospatial information retrieval and indexing system mdash;content mining, semantics modeling, and complex queries," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 839–852, Apr. 2007.

[18] G. K. Ouzounis, M. Pesaresi, and P. Soille, "Differential area profiles: Decomposition properties and efficient computation," *IEEE Trans. Pattern Analy. Machine Intell.*, vol. 34, no. 8, pp. 1533–1548, Aug. 2012.

[19] G. K. Ouzounis and P. Soille, "The Alpha-Tree Algorithm," JRC Scientific and Policy Report, European Commission, Joint Research Centre, 2012.

[20] M. Pesaresi, G. K. Ouzounis, and L. Gueguen, S. S. Shen and P. E. Lewis, Eds., "A new compact representation of morphological profiles: Report on first massive VHR image processing at the JRC," in *Proc. SPIE 8390*, Baltimore, MD, USA, Apr. 2012, vol. 8390, p. 8390025.

[21] L. Bruzzone and L. Carlin, "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, 2006.

[22] S. Blaschke, T. Lang, and G. J. Hay, Eds., *Object-Based Image Analysis: Spatial Concepts for Knowledge-Driven Remote Sensing Applications*1st ed. Berlin Heidelberg New York, Springer-Verlag, 2008.

[23] M. Pesaresi, A. Gerhardinger, and F. Kayitakire, "A robust built-up area presence index by anisotropic rotation-invariant textural measure," *J. Earth Observ. Applicat.*, vol. 1, no. 3, pp. 180–192, Sep. 2008.

[24] M. Pesaresi and J. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 309–320, Feb. 2001.

[25] M. D. Mura, J. Benediktsson, B. Waske, and L. Brunzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, Oct. 2010.

[26] L. Gueguen, P. Soille, and M. Pesaresi, "Structure extraction and characterization from differential morphological decomposition," in *7th Conf. Image Information Mining: Geospatial Intelligence from Earth Observation (IIM), 2011 ESA EUSC JRC*, Jun. 2011, pp. 53–58.

[27] L. Gueguen, P. Soille, and M. Pesaresi, "Change detection based on information measure," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pt. 2, pp. 4503–4515, 2011.

[28] M. Pesaresi, M. Halkia, and G. Ouzounis, "Quantitative estimation of settlement density and limits based on textural measurements," in *Proc. 2011 Joint Urban Remote Sensing Event, JURSE 2011*, 2011, pp. 89–92.

[29] G. K. Ouzounis, V. Syrris, and M. Pesaresi, "Multi-scale evaluation of Global Human Settlement scenes against reference data using statistical learning," *Pattern Recognition Lett.*, 2013, doi: 10.1016/j.patrec.2013.04.004.

[30] P. Gamba, M. Pesaresi, K. Molch, A. Gerhardinger, and G. Lisini, "Anisotropic rotation invariant built-up presence index: Applications to sar data," in *Proc. 2008 IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, 2008, vol. 5, pp. V338–V341, 1.

[31] L. Lu, Q. Li, L. Jing, H. Guo, and M. Pesaresi, "Classification of CBERS-02b high resolution image using morphological features for urban areas," in *Proc. 2nd Int. Workshop on Earth Observation and Remote Sensing Applicat. (EORSA 2012)*, Jun. 2012, pp. 16–20.

[32] J. Benediktsson, M. Pesaresi, and K. Arnason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pt. I, pp. 1940–1949, 2003.

[33] H. Akcay and S. Aksoy, "Automatic detection of geospatial objects using multiple hierarchical segmentations," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2097–2111, 2008.

[34] J. Benediktsson, J. Palmason, and J. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, 2005.

[35] P. Marpu, K.-S. Chen, C.-Y. Chu, and J. Benediktsson, "Spectral-spatial classification of polarimetric sar data using morphological profiles," in *Proc. 3rd Int. Asia-Pacific Conf. Synthetic Aperture Radar, APSAR 2011*, 2011, pp. 324–326.

[36] E. Pagot and M. Pesaresi, "Systematic study of the urban postconflict change classification performance using spectral and structural features in a support vector machine," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 1, no. 2, pp. 120–128, Jun. 2008.

[37] L. Gueguen, M. Pesaresi, A. Gerhardinger, and P. Soille, "Characterizing and counting roofless buildings in very high resolution optical images," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 1, pp. 114–118, Jan. 2012.

[38] L. Li, Z. Li, R. Zhang, J. Ma, and L. Lei, "Collapsed buildings extraction using morphological profiles and texture statistics—A case study in the 5.12 Wenchuan earthquake," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS 2010)*, Jul. 2010, pp. 2000–2002.

[39] H. Guo, L. Lu, J. Ma, M. Pesaresi, and F. Yuan, "An improved automatic detection method for earthquake-collapsed buildings from ADS40 image," *Chinese Sci. Bulletin*, 2009.

[40] G. Ouzounis, P. Soille, and M. Pesaresi, "Rubble detection from VHR aerial imagery data using differential morphological profiles," in *34th Int. Symp. Remote Sensing of the Environment*, 2011.

[41] L. Gueguen, M. Pesaresi, and P. Soille, "An interactive image mining tool handling gigapixel images," in *Proc. IEEE IGARSS 2011*, Jul. 2011, pp. 1581–1584.

[42] G. Trianni, E. Angiuli, and M. Pesaresi, "Statistical analysis of anisotropic rotation-invariant textural measurements of human settlements from multitemporal SAR data," in *Proc. 2011 Joint Urban Remote Sensing Event (JURSE)*, Apr. 2011, pp. 117–120.

[43] G. Trianni, E. Angiuli, and M. Pesaresi, "Improved anisotropic rotation invariant built-up presence index for a global human settlements layer (GHSL from medium resolution SAR data)," in *7th Conf. Image Information Mining: Geospatial Intelligence From Earth Observation (IIM), 2011 ESA EUSC JRC*, Jun. 2011.

[44] , B. Jasani, M. Pesaresi, S. Schneiderbauer, and G. Zeug, Eds.*, Remote Sensing From Space – Supporting International Peace and Security*. New York, NY, USA: Springer-Verlag, 2009.

[45] U. Deichmann, D. Ehrlich, C. Small, and G. Zeug, "Using High Resolution Satellite Data for the Identification of Urban Natural Disaster Risk," World Bank, Washington, DC, USA, World Bank European Union Joint Report, 81 pp., 2011.

[46] J. Crisp, "Who has Counted the Refugees?" UNHCR and the politics of numbers, Geneva, Switzerland, Working Paper No. 12. Policy Research Unit, UNHCR, CP 2500, CH-1211, 1999.

[47] M. Goodchild, H. Guo, A. Annoni, L. Bian, K. De Bie, F. Campbell, M. Craglia, M. Ehlers, J. Van Genderen, D. Jackson, A. Lewis, M. Pesaresi, G. Remetey, R. Simpson, A. Skidmore, C. Wang, and P. Woodgate, "Next-generation digital earth," in *Proc. National Academy of Science of the USA*, 2012, vol. 109, no. 28, pp. 11 088–11 094.

[48] M. Craglia, K. de Bie, D. Jackson, M. Pesaresi, G. Remetey-F, C. Wang, A. Annoni, L. Bian, F. Campbell, J. Ehlers, M. van Genderen, M. Goodchild, H. Guo, A. Lewis, R. Simpson, A. Skidmore, and P. Woodgate, "Digital earth 2020: Towards the vision for the next decade," *Int. J. Digital Earth*, vol. 5, no. 1, pp. 4–21, 2012.

[49] M. Pesaresi and D. Ehrlich, "A methodology to quantify built-up structures from optical VHR imagery," in *Global Mapping of Human Settlement Experiences, Datasets, and Prospects*, P. Gamba and M. Herold, Eds. Boca Raton, FL, USA: CRC Press, 2009, ch. 3, pp. 27–58.

[50] D. Lu and Q. Weng, "Use of impervious surface in urban land-use classification," *Remote Sens. Environ.*, vol. 102, no. 1–2, pp. 146–160, 2006.

[51] C. Tucker, D. Grant, and J. Dykstra, "NASA's Global orthorectified Landsat data set," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 3, pp. 313–322, Mar. 2004.

[52] P. Soille, "Morphological image compositing," *IEEE Trans. Pattern Analy. Machine Intell.*, vol. 28, no. 5, pp. 673–683, May 2006.

[53] C. Bielski and P. Soille, "Order independent image compositing," *Lecture Notes in Computer Science*, vol. 3617, pp. 1076–1083, Sep. 2005.

[54] C. Bielski, J. Grazzini, and P. Soille, "Automated morphological image composition for mosaicing large image data sets," in *Proc. IEEE IGARSS 2008*, Barcelona, Spain, Jul. 2008, pp. 4068–4071.

[55] P. Soille, Ed., "The IMAGE-2006 Mosaic Project," Joint Research Centre of the European Commission, 2008 [Online]. Available: http://dx.doi.org/10.2788/25572

[56] D. Barnea and H. Silverman, "A class of algorithms for fast digital registration," *IEEE Trans. Comput.*, vol. C-21, no. 2, pp. 179–186, Nov. 1972.

[57] Q. Tian and M. Huhns, "Algorithms for subpixel registration," *Comput. Vision, Graphics, Image Process.*, vol. 35, no. 2, pp. 220–233, 1986.

[58] J. d'Alge, R. Cartaxo, and G. Erthal, "Geometric Quality Assessment of CBERS-2," Tech. Rep. Instituto Nacional de Pesquisas Espaciais, 2004 [Online]. Available: http://mtc-m18.sid.inpe.br/col/dpi.inpe.br/banon/2006/08.03.19.33/doc/appl_13_2004.pdf

[59] E. Castejon, C. Forster, L. Fonseca, and T. Korting, "Orbital image correction for multiple models of geometric transformations," in *Proc. of Simp. Brasileiro de Sensoriamento Remoto, 15. (SBSR)*, São José dos Campos, Brazil, 2011.

[60] R. Irish, J. Barker, S. Goward, and T. Arvidson, "Characterization of the Landsat-7 ETM+ automatic cloud cover assessment (ACCA) algorithm," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 10, pp. 1179–1188, Oct. 2006.

[61] J.-F. Rivest, P. Soille, and S. Beucher, "Morphological gradients," *J. Electron. Imag.*, vol. 2, no. 4, pp. 326–336, Oct. 1993.

[62] P. Soille, *Morphological Image Analysis: Principles and Applications*, 2nd ed. Berlin Heidelberg New York: Springer-Verlag, 2003.

[63] R. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. 3, no. 6, pp. 610–621, Nov. 1973.

[64] L. Gueguen, P. Soille, and M. Pesaresi, "A new built-up presence index based on density of corners," in *IEEE IGARSS*, Munich, Germany, 2012.

[65] P. Salembier, A. Oliveras, and L. Garrido, "Anti-extensive connected operators for image and sequence processing," *IEEE Trans. Image Process.*, vol. 7, no. 4, pp. 555–570, 1998.

[66] R. Jones, "Connected filtering and segmentation using component trees," *Comput. Vision Image Understand.*, vol. 75, no. 3, pp. 215–228, 1999.

[67] P. Salembier and J. Serra, "Flat zones filtering, connected operators, and filters by reconstruction," *IEEE Trans. Image Process.*, vol. 4, no. 8, pp. 1153–1160, 1995.

[68] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Process. Mag.*, vol. 13, no. 6, pp. 47–60, 1996.

[69] B. Settles, "Active Learning Literature Survey," Computer Sciences 1648, University of Wisconsin–Madison, USA, 2009.

[70] D. Ehrlich, M. Marin-Herrera, M. Pesaresi, T. Kemper, and M. Halkia, "Validation of Regional Built-Up Maps Derived from SPOT 5 Imagery," Tech. Rep. Joint Research Centre of the European Commission, 2012.

[71] S. Kullback, *Information Theory and Statistics*. New York, NY, USA: Dover, 1968.

[72] V. Syrris and M. Pesaresi, "On the assessment of automatically processing HR/VHR imagery using low-resolution global reference data," in *Joint Urban Remote Sensing Event*, Sao Paulo, Brazil, Apr. 2013.

[73] M. Pesaresi and M. Halkia, "Global Human Settlement Layer and Urban Atlas Integration: Feasibility Report," European Commission, Joint Research Centre, JRC Scientific and Policy Report EUR 25328 EN, 2012.

[74] T. Kemper, X. Blaes, D. Ehrlich, F. Haag, and M. Pesaresi, "On the feasibility to map the settlements of Brazil with the CBERS-2B satellite," in *Joint Urban Remote Sensing Event*, Sao Paulo, Brazil, Apr. 2013.

[75] L. Lu, H. Guo, M. Pesaresi, P. Soille, and S. Ferri, "Recognition of built-up areas in China using CBERS-2B HR data," in *Joint Urban Remote Sensing Event*, Sao Paulo, Brazil, Apr. 2013.

[76] M. Pesaresi and A. Gerhardinger, "Improved textural built-up presence index for automatic recognition of human settlements in arid regions with scattered vegetation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 1, pp. 16–26, 2011.

Martino Pesaresi graduated in Town and Regional Planning from the IUA Venice, Italy, in 1992 with an overall first class mark of 110/110.

He has been independent consultant in remote sensing and geo-information analysis applied to urban and territorial studies in the 1990s. In 2001–2004 he was working in the digital map production industry as responsible of the R&D department. In 2004, he joined the EC JRC, Institute for the Protection and Security of the Citizen (IPSC), Global Security and Crisis Management Unit. In this framework he was contributing to various European Programs dealing with the use of remote sensing technologies in support to regional planning, disaster mitigation, crisis management, damage and needs assessment, illicit crop monitoring and refugee camps analysis, among others. In 2005–2007 he was the scientific coordinator of the Global Monitoring for Security and Stability (GMOSS) Network of Excellence in the aeronautics and space priority of the 6th FP funded by the EC DG Enterprise & Industry. Since 2007, he is the team leader of the Geo-Spatial Information Analysis for Global Security and Stability action at the JRC. In this frame he has been scientific responsible and manager of a variety of projects belonging to several European Programs. All these projects involved remote sensing data analysis and automatic image information retrieval in real-world and operational scenarios. His main scientific contributions are in pattern recognition, image information mining and automatic target detection in remote sensing applications: special focus on automatic analysis of meter and sub-meter resolution imageries using textural and morphological multi-scale image descriptors.

Guo Huadong is Director-General of the Chinese Academy of Sciences (CAS) Institute of Remote Sensing and Digital Earth (RADI), an Academician of CAS, and a Fellow of the Academy of Sciences for the Developing World (TWAS). He presently serves as President of the International Council for Science (ICSU) Committee on Data for Science and Technology (CODATA), Secretary-General of the International Society for Digital Earth (ISDE), and Editor-in-Chief of the *International Journal of Digital Earth* (IJDE) published by Taylor & Francis.

He has over thirty years of experience in remote sensing, specializing in radar for Earth observation and remote sensing applications, and involving research on digital Earth since the end of the last century. He has been Principle Investigator for over twenty major national projects or programs in China, and Principle Investigator for seven international radar remote sensing projects. He also serves as Director of the International Center on Space Technologies for Natural and Cultural Heritage under the Auspices of UNESCO.

Xavier Blaes received the Ph.D. degree in agronomic sciences from the Catholic University of Louvain (UCL), Belgium.

He joined the Environmetrics and Geomatics unit of UCL in 1998 to work on crop discrimination and biophysical parameters retrieval using optical, interferometric and polarimetric SAR signals. From 2007 to 2010, he was GIS coordinator for the UNDP in Congo Democratic Republic. He is involved remote sensing and GIS research activities for the JRC since 2010.

Daniele Ehrlich received the B.S. degree in forestry in 1984 from the University of Padova, Italy, and the Ph.D. degree in 1992 from the University of California at Santa Barbara.

He is a senior staff member of the Joint Research Centre of the European Commission, based in Ispra, Italy. He has over 20 years of experience in remote sensing and GIS applied to a variety of disciplines including crop area estimation, tropical forest mapping, crisis management with focus on damage assessment and humanitarian assistance. His current research focuses on quantifying the extent and the dynamics of settlements using high-resolution satellite imagery. He uses the derived settlement layers for a systematic analysis of the global built-up environment, for population estimations and for generating physical exposure databases for global disaster risk assessments.

Stefano Ferri received the M.S. degree in environmental engineering from Politecnico di Milano, Italy, in 2004.

His main activities have been focused on geotechnic, urban and territorial analysis, environmental impact evaluation and territory planning. Since 2009 he has collaborated with the Joint Research Centre of Ispra, Italy, as a GIS expert. He has been involved in various international projects supported by European Commission. His research activities include processing of remote sensing images, urban analysis, development of systems and applications for very large data management.

Lionel Gueguen received the engineering degree in telecommunications and the M.S. degree in signal and image processing from the Ecole Nationale Superieure des Telecommunications de Bretagne, Brest, France, in 2004. He received the Ph.D. degree in signal and image processing from Ecole Nationale Superieure des Telecommunications Paris, in 2007.

From July 2009 to June 2012, he worked with Joint Research Centre of the European Commission in the Support to External Security Unit. In September 2012, he joined DigitalGlobe Inc. as an R&D scientist to study and develop image information mining techniques applied to electro-optical satellite images.

Matina Halkia holds a Diploma of Architect Engineer, and a Master of Arts from Tufts University, Medford, MA, USA.

She trained at post-graduate level at MIT's Media Laboratory in IT-applied research in architecture. From 1998 to 2001 she was principal researcher at Starlab N.V. in Brussels, Belgium, where she led a research group in interactive architecture. As of 2001, she joined the European Commission in Ispra, Italy, as a scientific officer. Since 2009 she has been working in the analysis of human settlements, through remote sensing of the environment, in particular, built-up analysis and characterization.

Mayeul Kauffmann received the Master's degree in international relations and security and the Ph.D. in economics.

He was Research Director at the School of Peace (Ecole de la paix), Grenoble, France, and taught conflict analysis and quantitative methods in France and Switzerland. He is now working as a consultant at the IPSC (EC JRC).

Thomas Kemper received the Ph.D. degree in geosciences.

He is scientific officer at the Joint Research Centre (JRC) of the European Commission in Ispra, Italy. From 2004 to 2007 he worked for the German Aerospace Center (DLR), where he helped in setting up the Center for Satellite Based Crisis Information (ZKI), which provides rapid mapping information after natural disasters. Since 2007, he has been working on the analysis of human settlements, in particular informal settlements such as slums and IDP/refugee dwellings.

Linlin Lu received the Ph.D. degree from the Institute of Remote Sensing Applications, Chinese Academy of Sciences (CAS), Beijing, in 2009.

Since then, she has been working as a researcher in the land surface dynamic group at the Key Laboratory of Digital Earth Science of RADI, CAS. She is currently collaborating with the EC JRC IPSC in support of the Global Human Settlement Layer (GHSL) initiative. Her research interests include image information detection and time series analysis applied to human settlement and exposure mapping, urban development and vegetation phenology.

Mario A. Marin Herrera is a skilled and experienced GIS professional in the field of Geographic Information Systems, Remote Sensing, Environmental Sciences and Urban Planning. He received his M.Sc. degree in Geographic Information Systems at the University of Edinburgh. His undergraduate studies in Environmental Sciences at the University of Pablo de Olavide along with his expert course in Urban Planning, have given him the opportunity to obtain an in-depth appreciation of the various fields of spatial planning and environmental management. His current research interests focus on the areas of nature conservation, land use, sustainability and GIS.

Georgios K. Ouzounis received the Bachelor degree in electrical and electronic engineering (B.Eng.) and the M.Phil. degree in computer science from the Department of Computing and Electrical Engineering, Heriot Watt University, Edinburgh, Scotland, U.K., in 1998 and 2001, respectively. He received the Ph.D. in image analysis and connected morphological filters from the Johann Bernoulli Institute ofMathematics and Computing Science, University of Groningen, The Netherlands, in 2009.

From September 2009 to September 2012 he was appointed as a researcher at the Geospatial Image Analysis Group of the Global Security and Crisis Management Unit, Joint Research Centre, European Commission, Ispra, Italy, where he worked on the development of advanced VHR satellite and aerial image analysis algorithms. Since October 2012, he has been with DigitalGlobe working as a research scientist on related application areas.
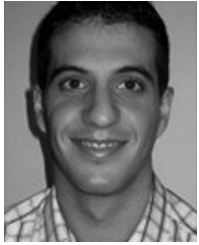
Marco Scavazzon received the degree in computer science from the University of Venice, Italy, in 2005.

Since April 2011, he has been working as a software engineer at the Joint Research Centre of the European Commission, Ispra, Italy. He is involved in the implementation and development of a grid computing environment for large scale processing of satellite images. His interests are focused in systems integration and performance optimization.

Vasileios Syrris received the B.Math degree in mathematics from the Aristotle University of Thessaloniki, Greece, and the M.Sc. degree in artificial intelligence: knowledge-based systems from the University of Edinburgh, U.K. He received the Ph.D. degree in computational intelligence from the Aristotle University of Thessaloniki, Greece.

He is currently working as a scientific researcher for the Geo-Spatial Information Analysis for Global Security and Stability action, Institute for the Protection and Security of the Citizen, Joint Research Centre of European Commission, Ispra, Italy. His research interests are in machine learning, robotics, automation, computer vision, remote sensing, control engineering, e-learning, statistics, data analysis and representation.

Pierre Soille received the engineering degree in 1988 from the University of Louvain, Louvain-la-Neuve, Belgium, the doctorate degree in 1992 from the same university in collaboration with the Ecole des Mines de Paris (France), and the habilitation degree in 1997 from the University of Montpellier II, Montpellier, France.

He has been with the EC Joint Research Centre, Ispra, Italy, since December 1999. His current research activities are dedicated to the automatic information retrieval from high to very high resolution satellite image data for the generation of new global geospatial data sets.

Luigi Zanchetta is an IT Project Manager with more than 20 years of experience in the field of information technology. During this period, he worked on a variety of topics including software development, database and system administration, networking and security. Currently he is working for the European Commission at the Joint Research Centre of Ispra, Italy, where he is involved in the design and implementation of complex IT infrastructures for image elaboration using parallel computing.