

Detecting Man-Made Structures and Changes in Satellite Imagery With a Content-Based Information Retrieval System Built on Self-Organizing Maps

Matthieu Molinier, Jorma Laaksonen, *Senior Member, IEEE*, and Tuomas Häme

Abstract—The increasing amount and resolution of satellite sensors demand new techniques for browsing remote sensing image archives. Content-based querying allows an efficient retrieval of images based on the information they contain, rather than their acquisition date or geographical extent. Self-organizing maps (SOMs) have been successfully applied in the PicSOM system to content-based image retrieval in databases of conventional images. In this paper, we investigate and extend the potential of PicSOM for the analysis of remote sensing data. We propose methods for detecting man-made structures, as well as supervised and unsupervised change detection, based on the same framework. In this paper, a database was artificially created by splitting each satellite image to be analyzed into small images. After training the PicSOM on this imagelet database, both interactive and off-line queries were made to detect man-made structures, as well as changes between two very high resolution images from different years. Experimental results were both evaluated quantitatively and discussed qualitatively, and suggest that this new approach is suitable for analyzing very high resolution optical satellite imagery. Possible applications of this work include interactive detection of man-made structures or supervised monitoring of sensitive sites.

Index Terms—Change detection, content-based information retrieval, high-resolution optical satellite images, man-made structure detection, self-organizing maps (SOMs).

I. INTRODUCTION

EARTH observation (EO) data volumes are growing rapidly, with an increase in both the number of satellite sensors and in their resolutions. Yet, it is estimated that only 5% of all EO data collected up to now have been used. Therefore, traditional remote sensing archiving systems—with queries made typically on sensor type, geographical extents, or acquisition date—could become obsolete as the amount of data to be stored, accessed, and processed explodes. Using image content would allow a more efficient use of these databases. This has led to the emergence of content-based image retrieval (CBIR) systems for archive management of remote sensing images [1]–[9], and for annotation or interpretation of satellite images [1], [9]–[11].

Manuscript received March 18, 2006; revised July 25, 2006. This work was supported in part by the Academy of Finland and in part by the National Technology Agency of Finland (TEKES).

M. Molinier and T. Häme are with the VTT Technical Research Centre of Finland, 02044 Espoo, Finland (e-mail: matthieu.molinier@vtt.fi).

J. Laaksonen is with the Helsinki University of Technology, 02015 Espoo, Finland (e-mail: jorma.laaksonen@hut.fi).

Digital Object Identifier 10.1109/TGRS.2006.890580

Previous work has been made on databases of relatively small images acquired from medium-resolution sensors. A visually oriented query method was experimented in [3] on a small test image archive, containing 484 windows extracted from Landsat Thematic Mapper (TM) images. An intuitive method for semantic labeling of images, suited for query by image content, was described in [4] and tested on the same image archive. A stochastic representation of image content was used in [6] and [11] for interactive learning within a database of about a thousand 1024×1024 Landsat TM scenes—however, queries were made by marking the training areas.

A recent work [12] combining support vector machines (SVMs) with active relevance feedback was successfully tested on a database of 3600 small satellite images with ground truth for six classes. But, the database was built manually, and class labels reflected the majority content of each small image. Relevant feature selection for satellite image indexing was discussed in [13], on a similar database of 600 small extracts of SPOT5 images (64×64 pixels). A rigorous framework for remote sensing image information mining has been presented in [8], integrating spectral information (through land cover and land use identified by the SVM classification) and spatial information (Gabor wavelet coefficients as a textural feature). Performance evaluation was run on a database of 128×128 pixels extracts of Landsat TM images, with promising results. Other works [2], [5], [7] seemed to focus more on managing large databases of full remote sensing scenes.

Very few works describe the utilization of CBIR techniques for the purpose of a single satellite scene interpretation, let alone change detection. Techniques inspired from knowledge discovery in database were however used in [14] for segmenting images of natural scenes. A method for multisource remote sensing data fusion with multiple self-organizing maps (SOMs) [15] has been introduced in [16]. Spectral, geographical, temporal, and spatial features were merged in a data vector. But, the SOMs were used in [16] for classification not for information retrieval from the remote sensing image archives nor for detection of changes.

Change detection in remote sensing imagery is involved in an ever-growing number of applications, allowing to monitor the impact of natural disasters (e.g., floods [17], earthquakes [18]), forest changes [19], or even compliance to nuclear nonproliferation treaties [20]. Classical change detection techniques rely, for example, on image difference or ratio, postclassification comparison, classification of multitemporal datasets, or change

vector analysis [21]. Pixel-based methods require a subpixel registration accuracy, as misregistrations greater than one pixel produce numerous errors when comparing images [21].

We present an original utilization and improvement of a CBIR system, PicSOM, for the analysis of remote sensing images. In the PicSOM image database browsing system [22], several thousands of images are mapped on SOMs through the extraction of image descriptors, including textural and color features. After the SOMs are trained, the user can visually query the database, and the system automatically finds images similar to those selected. This approach has been successfully applied to databases of conventional images [23], [24]. We aim at using the potential of this neurally motivated CBIR system for detecting man-made structures or changes between two or more satellite images, with a special interest in the changes involving human activity (e.g., buildings, structures). The approach presented in this paper could be applied to various types of sensors, even to synthetic aperture radar (SAR) with appropriate feature descriptors. In this paper, we focused on very high resolution optical images, as the change detection is challenging in such imagery—due to the high level of details in the images, and sensitivity to illumination conditions or misregistration effects.

The key idea of this paper is to artificially create an “image database” from each satellite image to be analyzed, by clipping it into thousands of small images or imagelets. PicSOM can be trained on that virtual database, then visually queried for finding objects of interest like man-made structures. Previous work [25] showed promising results on very high resolution satellite images, even with features that had been developed for generic images, and did not take into account the spectral or spatial properties of the EO data.

In this paper, we extended the PicSOM system so that it handles multispectral imagery and includes features adapted to the detection of man-made structures in high resolution optical satellite images. Fusion of panchromatic and multispectral information was done naturally within the PicSOM, in which several SOMs are trained in parallel (one SOM per feature). Qualitative and quantitative evaluations of the methods were carried out for man-made structure detection and change detection, using partially labeled datasets. Possible applications of this work are high resolution satellite image annotation or monitoring of sensitive areas for undeclared human activity, both in an interactive way.

This paper is organized as follows. The next section explains how we built a database from one or two satellite images. The SOM and PicSOM are introduced in Section III, and the features extracted for the study are introduced in Section IV. Section V explains the use of PicSOM for retrieving man-made structures and detecting changes. Experimental results are presented in Section VI and discussed in Section VII. Conclusions and future work are given in Section VIII.

II. DATABASE PREPARATION

CBIR is a technique for finding relevant images in a database. We present here how we created a database of images from one or two satellite images.

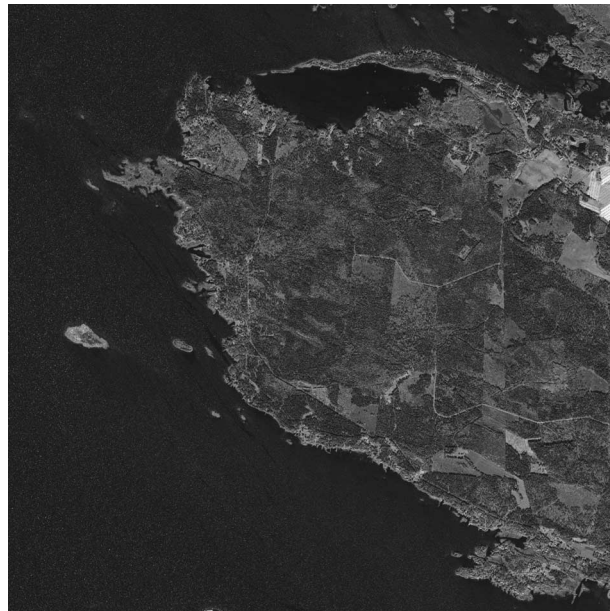


Fig. 1. True-color pan-sharpened 2005 QuickBird study scene.

A. Satellite Imagery

Two QuickBird scenes were acquired in the beginning of September 2002 and in the middle of June 2005, covering the same coastal area in Finland. They include four multispectral channels with a 2.4-m ground resolution—blue (450–520 nm), green (520–600 nm), red (630–690 nm), and near-infrared (NIR) (760–900 nm), and a panchromatic channel (450–900 nm) with a ground resolution of 0.6 m. Both scenes were remarkably cloud free and were coregistered from one to another. Slight misregistration effects remained due to a lack of digital elevation model of the monitored area.

Panchromatic and multispectral images were kept as separate data sources for later feature extraction. For the sole purpose of displaying the images in the PicSOM, a fusion of panchromatic and multispectral images was done by pan sharpening—no feature was calculated on the pan-sharpened images. Pan sharpening was also used to visually check the correct registration between multispectral and panchromatic channels. The same study area of size about 4×4 km, as shown in Fig. 1, was extracted from both scenes.

B. Database Generation

An image retrieval system, such as PicSOM, typically requires several thousands of images in a database in order to produce a relevant indexing. We artificially created databases from the two study scenes to be analyzed. Each scene was cut into $71 \times 71 = 5041$ nonoverlapping small images or imagelets, of size 100×100 pixels for panchromatic data and 25×25 pixels for multispectral data.

By this operation, we expected to reduce the amount of contents in each image. The study scene in Fig. 1 is a rich detailed image with lots of target classes (water, forest, arable land, built-up areas, etc.). On the contrary, imagelets, as shown in Fig. 2, usually contain only two or three distinct target classes. The pan-sharpened images were cut in the same manner

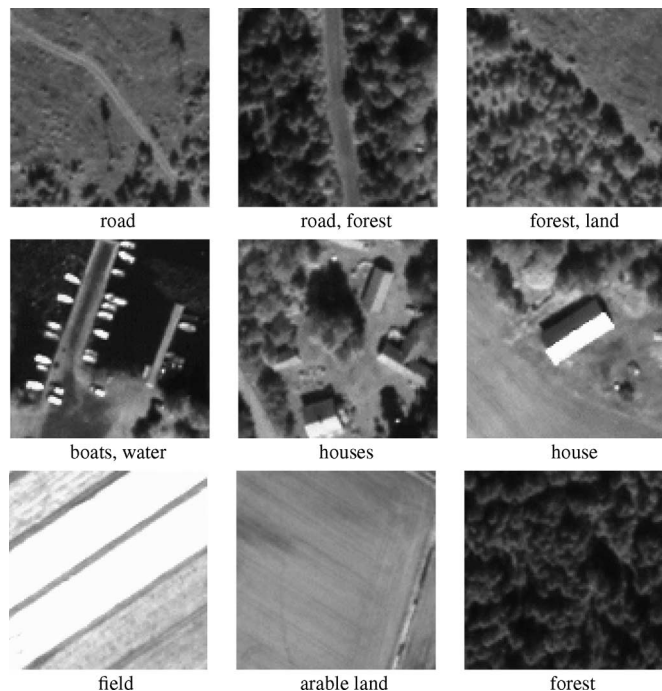


Fig. 2. Samples of imagelets automatically extracted from 2005 study area.

(into 100×100 pixels imagelets) to be included in the PicSOM for visualization.

In this paper, imagelets were named in such a way that it tells their location within the study area—row and column of the original image—and year of acquisition. A triplet (row, column, and acquisition date) designated unambiguously an imagelet, regardless of the data type it had been extracted from. We denote by database the reunion of imagelets coming from panchromatic, multispectral, and pan-sharpened data sources for a given geographical extent—here, the study scene. Each acquisition date generate a subdatabase of the parent database. Given a row and column index, the corresponding two imagelets of the 2002 and 2005 subdatabases covered the same geographical extents within the coregistration accuracy.

C. Creating Ground Truth for Method Evaluation

The 2002 study area was labeled into seven classes—{agricultural field, arable land, buildings, clearcuts, forest, roads, and water}. Because the classification of very high spatial resolution images is challenging, it was highly assisted by a visual inspection. Water and forest classes were automatically labeled with the AutoChange software [26], developed at VTT. Designed for automatic change detection between two images, it can be used for classifying a single image, as it relies on a modified version of k-means clustering [26]. Other classes were obtained by a maximum-likelihood classification after visually marking the training areas. Small buildings and narrow roads were not retrieved correctly by the maximum-likelihood classification. Man-made structures being the main target of this paper, all buildings and roads were manually delineated in the pan-sharpened image by a visual inspection. Aiming at detecting appearing or disappearing constructions, buildings

were also manually delineated in the 2005 imagelets along the same process.

The full-size classification layer was cut as were the satellite images. Multiple labels were then assigned to each imagelet, depending on the classes they contained. The lists of imagelets containing pixels of each class were built and saved as seven text files handled by the PicSOM. The labels were used in the PicSOM to ease querying or selecting imagelets of interest during system development and testing. It also allowed a quantitative evaluation of the methods as described in Section VI.

III. SOMS FOR CBIR

In this section, we describe the principle of the PicSOM content-based retrieval and browsing system based on SOMs. PicSOM has been developed as a framework for studies on content-based information retrieval, and it has been applied in divergent domains of both visual and nonvisual data. A central design principle of the system has been its good scalability to large database sizes by efficient off-line indexing of the objects. A demonstration version of the PicSOM is displayed in [22].

A. SOMs

The SOM [15] is a neurally motivated unsupervised learning technique which has been used in many data-analysis tasks. A genuine feature of the SOM is its ability to form a nonlinear mapping of a high-dimensional input space to a typically 2-D grid of artificial neural units. During the training phase of a SOM, the model vectors in its neurons get values which form a topographic or topology-preserving mapping. Through this mapping, the vectors that reside near each other in the input space are mapped into the nearby map units in the SOM grid. Patterns that are mutually similar with respect to the given feature extraction scheme are thus located near each other on the SOM surface.

The training of a SOM starts from the situation where the model vectors \mathbf{m}_i of each map unit i are initialized with random values. For each input feature vector or sample $\mathbf{x}(t)$, the “winner” or the best matching map unit (BMU) $c(\mathbf{x})$ is identified on the map by the condition

$$\forall i : \quad \|\mathbf{x}(t) - \mathbf{m}_{c(\mathbf{x})}(t)\| \leq \|\mathbf{x}(t) - \mathbf{m}_i(t)\| \quad (1)$$

where $\|\cdot\|$ is commonly the Euclidean norm. After finding the BMU, a subset of the model vectors constituting a neighborhood that centered around node $c(\mathbf{x})$ is updated as

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) + h(t; c(\mathbf{x}), i) (\mathbf{x}(t) - \mathbf{m}_i(t)). \quad (2)$$

Here, $h(t; c(\mathbf{x}), i)$ is the “neighborhood function,” a decreasing function of the distance between the i th and $c(\mathbf{x})$ th nodes on the map grid. The training is reiterated over the available samples, and the value of $h(t; c(\mathbf{x}), i)$ is allowed to decrease in time to guarantee the convergence of the prototype vectors \mathbf{m}_i . Large values of the neighborhood function $h(t; c(\mathbf{x}), i)$ in the beginning of the training initialize the network, and small values on later iterations are needed in fine-tuning. The outcome of

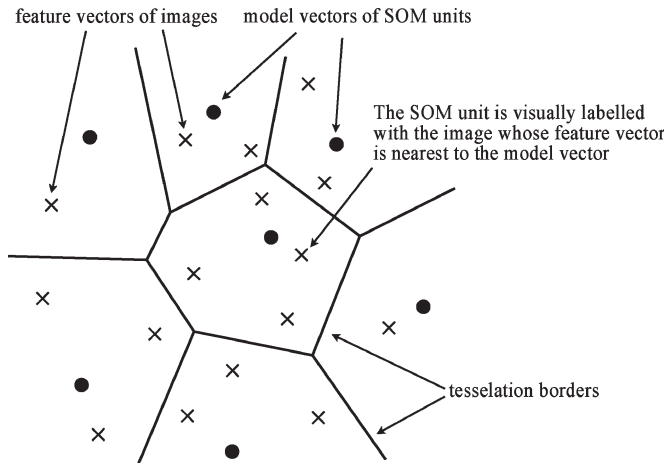


Fig. 3. Visualization of the distributions of the feature vectors of images (crosses) and the SOM model vectors (circles) in a 2-D space.

training a SOM is essentially the topographic ordering of the training data samples on the surface of the map grid.

B. PicSOM System

The PicSOM system [22] used in this paper has originally been developed for CBIR research [23], [24], [27]. It implements two essential CBIR techniques, query by examples (QBE) and relevance feedback. These methods can be used for iterative retrieval of any type of visual or nonvisual content. In iterative QBE, the system presents in a visual interface some images or other objects to the user, who then marks a subset of them as relevant to the present query. This relevance information is fed back to the system, which then tries to find more similar objects and returns them in the next query round. In Section V, we show how the same technique can also be applied in the semiautomated interactive analysis of satellite images—for example, detecting man-made structures.

In PicSOM, multiple SOMs are used in parallel, each created with different low-level visual features. Each imagelet is therefore used multiple times as input to the different feature extraction techniques, and the resulting feature vectors are used in training different SOMs. When an image SOM has been trained, all the feature vectors x are once more mapped to it, each one in its BMU. Every SOM unit and its model vector m is then assigned a visual label from the imagelet whose feature vector was the nearest to the model vector. Fig. 3 demonstrates how the feature vectors, BMUs, and visual labels can situate in an imagined 2-D feature space. The tessellation regions determine which imagelets will be represented by each SOM unit in the index defined by that specific feature extraction scheme.

The different feature extraction schemes and the SOMs that result from the training impose different similarity functions on the images. As a consequence, two imagelets whose contents are similar to each other with respect to some feature will be mapped to nearby units on that SOM's surface, whereas they will be located far apart on another feature's SOM if that feature regards them as dissimilar. Every interactive image query is unique, and each user of a CBIR system has his or

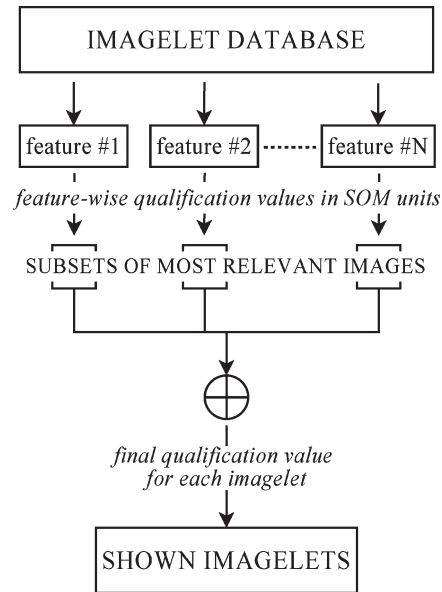


Fig. 4. Principle of PicSOM. Features are extracted from a database of images (imagelets in our case). Subsets of most relevant imagelets for a given feature are formed according to the featurewise qualification values in SOM units. Those subsets are then combined to select the images to be shown.

her own transient view of images' similarity and relevance to the specific query. Therefore, a system structure capable of holding many simultaneous similarity representations can adapt to different kinds of retrieval tasks. Also, the studies made by other researchers showed that joint multiple SOMs have a better discrimination capability than a single SOM [16].

Relevance feedback has been implemented in the PicSOM by using the parallel SOMs. Each image presented in the interactive on-line use of the system is graded by the user as either relevant or nonrelevant. All these relevance grades are then projected to the best matching SOM units of the graded images on all the different SOM surfaces. Maps where many relevant images are mapped in the same or nearby SOM units agree well with the user's conception on the relevance and semantic similarity of the images. When we assume that similar images are located near each other on the SOM surfaces, we are motivated to spread the relevance information placed in the SOM units also to the neighboring units.

This relevance spreading is implemented in the PicSOM as follows. All images marked as relevant are first given an equal positive value that is inversely proportional to their number. Likewise, the nonrelevant images receive negative values that are inversely proportional to the number of nonrelevant images. The overall sum of these relevance values is thus zero. The values are then summed in the BMUs of the images, and the resulting sparse value fields are low-pass filtered to produce the final score or qualification value for each SOM unit.

As the final step, each image is given a total qualification value that is obtained as a sum of the qualification values from its best matching or index units from the SOM surfaces of the different features. Those yet unseen images, which have the highest qualification values, will then be shown to the user on the next query round (Fig. 4). In the PicSOM, features that fail to coincide with the user's conceptions always produce lower

qualification values than those descriptors that match the user's expectations. As a consequence, the different features do not need to be explicitly weighted as the system automatically takes care of weighting their opinions.

Qualification values can similarly be calculated off-line for different semantic content classes, if such have been defined for the database. Suppose that all images in a specific ground truth class are marked as relevant, with a positive value in their respective BMUs on the different SOMs. After a low-pass filtering, we then obtain class distributions on the SOM surfaces, as can be seen in Fig. 7. The better a feature is in mapping the semantic class in a dense region, the sharper the resulting distribution on the SOM surface will be.

IV. FEATURES

Feature extraction and selection is a central task in any CBIR system. The features are usually low level, and describe general properties of the images, like color distribution, texture, or shapes. In [12], weighted color histograms, Hough transform as a shape descriptor, and textural information based on the Fourier transform were used as features. In [13], features like Haralick descriptors, Gabor features, and Gauss–Markov random fields parameters were included, as well as geometric features resulting from a local statistical analysis of the edges contained in images.

In this paper, we used six low-level features automatically extracted from the imagelets. Some were part of the PicSOM; three were added in this paper for the specific purpose of detecting man-made structures or changes in the satellite imagery.

A. Features Included in the Standard PicSOM

The PicSOM system was originally developed for databases of conventional images (those found, for example, in common web image databases). Generic and low-level image features are integrated within the PicSOM system. We use here the term imagelet to equivalently describe any data source (panchromatic or multispectral) covering the geographical extent of that imagelet. The features were the following.

- 1) Average red, green, and blue (RGB). This 3-D feature was calculated as the average values of the RGB channels of the pixels in an imagelet.
- 2) Color moments. The RGB color coordinates of the pixels were first transformed to the hue-saturation-value (HSV) coordinate system. Then, the three first moments (mean, variance, and skewness) of the HSV values were calculated and stored in a 9-D feature.
- 3) Texture. This feature was formed by studying the eight-neighbors of each imagelet pixel. For each eight-neighbor of a given pixel, a counter was incremented when the illumination in that neighbor was larger than in the center pixel. When all the imagelet pixels had been scanned, the final counts were divided by the total number of pixels in the imagelet. The resulting 8-D feature vector describes the local illumination differences and, thus, the small-scale texture of the imagelet.

TABLE I
FEATURES EXTRACTED FROM THE IMAGELETS

| Feature | Data source | Dimensionality |
|-------------------|-----------------------------|----------------|
| average RGB | Multispectral (RGB only) | 3 |
| color moments | Multispectral (RGB only) | 9 |
| texture | Panchromatic | 8 |
| xy -coordinates | Index of the imagelet | 2 |
| NDVI histogram | Multispectral (Red and NIR) | 100 |
| edges histogram | Panchromatic | 180 |

In a preliminary work [25], the above three low-level image features were extracted from the pan-sharpened RGB imagery. Pan sharpening was used in an attempt to keep the spatial and spectral information in the same data source. In addition, the radiometry had been reduced from 11 to 8 bits, so that the images could be displayed in the PicSOM interface. Despite the extraction of features from the pan-sharpened and radiometrically reduced images, the results were encouraging.

In this paper, the full 11-bit radiometry was conserved and the data sources were kept separated. Pan sharpening was still used to display imagelets in the PicSOM, but strictly not for any feature extraction. Those generic features were extracted on the most relevant input data source (see Table I).

B. Additional Features for Satellite Imagery

Other features can as well be calculated in another environment then loaded into the PicSOM as plain text files. In order to help retrieving the man-made structures and detecting the changes in satellite imagery, three additional features were included.

- 1) xy coordinates. This feature used the spatial location of the imagelet in the original scene as a 2-D feature (row and column indexes). As the images from the two years were coregistered, the value of this feature was always the same for any imagelet position, regardless of the year, image content, or data source.
- 2) NDVI histogram. The normalized difference vegetation index, or

$$\text{NDVI} = \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}} \quad (3)$$

in which NIR means the NIR spectral wavelength and Red means the visible red light, is a widely used feature in remote sensing to characterize the amount of green vegetation. The bare soil gives low NDVI values, whereas the values of green vegetation are high. In land surfaces, the index value is usually positive. The NDVI has also a good correlation with the green biomass at low biomass levels, but it saturates when the bare soil fraction is close to 0%.

In this paper, the NDVI was utilized to separate the vegetated surfaces from the nonvegetated, such as buildings and roads. Imagelets are likely to have often heterogeneous content; thus, the mean or standard deviation of NDVI on the imagelets would not effectively characterize them. Therefore, an NDVI histogram was computed on imagelets, with 100 bins ranging from minimum to

maximum NDVI values of the uncut study scene. Histograms of different imagelets could easily be compared, since the bin positions were independent of imagelets. Negative values of NDVI, usually associated to water or deep shadows, were left unchanged because the study scene contains a significant proportion of water areas, that were part of the zone to monitor. The allure of NDVI histograms was hypothesized to help in discriminating imagelets containing, e.g., buildings from the completely vegetated imagelets. The counts of the NDVI histogram constituted a 100-D feature vector.

- 3) Edges histogram. It was forecasted in [28] that, in addition to spectral features, structural features (like edges or shapes) would be suited to the CBIR in high-resolution satellite images. Pattern directionality [29] has been used in automatic video indexing [30], and recently in systems to manage EO image archives [31]. It is based on the histogram of a local gradient magnitude against directional angle. We used a slightly modified version of that feature.

Let I be an imagelet, and S_x and S_y as the 3×3 Sobel operators extracting vertical and horizontal edges. Noting by $*$ the convolution operation, the imagelet gradient magnitude $\|G\|$ and angle θ are defined at each pixel by

$$\|G\| = \sqrt{G_x^2 + G_y^2}, \quad \text{with } G_x = I * S_x \quad \text{and} \quad G_y = I * S_y \quad (4)$$

$$\theta = \tan^{-1} \left(\frac{G_y}{G_x} \right) + \frac{\pi}{2}, \quad \theta \in [0, \pi]. \quad (5)$$

The angle θ is counted counterclockwise and equals 0 when the gradient G is in the horizontal direction. In [30], a histogram is obtained by quantizing θ and counting how many pixels have a gradient magnitude $\|G\|$ higher than a noncritical threshold.

Because shadows or details can generate in very high resolution satellite images strong edges that do not correspond to structures, we avoided choosing a threshold in this paper. Instead, for each angle, the histogram count was weighted by the mean of the two extremal gradient magnitudes (minimum and maximum values). Typically, simple buildings are rectangular, with corresponding edges appearing in directions θ_0 (direction of the edge with the strongest magnitude) and $\theta_0 + (\pi/2) \bmod \pi$. The histogram was then centered around θ_0 by a circular shifting. After conversion to degrees, the result was a histogram of gradient magnitude between angles -90° and $+89^\circ$, normalized in angular position. If an imagelet contains a rectangular building, the histogram should show a primary peak in the central position and secondary peaks at both ends. Imagelets containing only forest should show a flat, close to a uniform histogram—no direction of the gradient would be clearly dominating.

Fig. 5 shows sample imagelets from the 2002 image, with their NDVI and edges histograms. A single building (first imagelet) generates enough gradient magnitude to leave a signature at the center of the edges histogram and two secondary lobes in the outermost parts. The NDVI histogram shows high

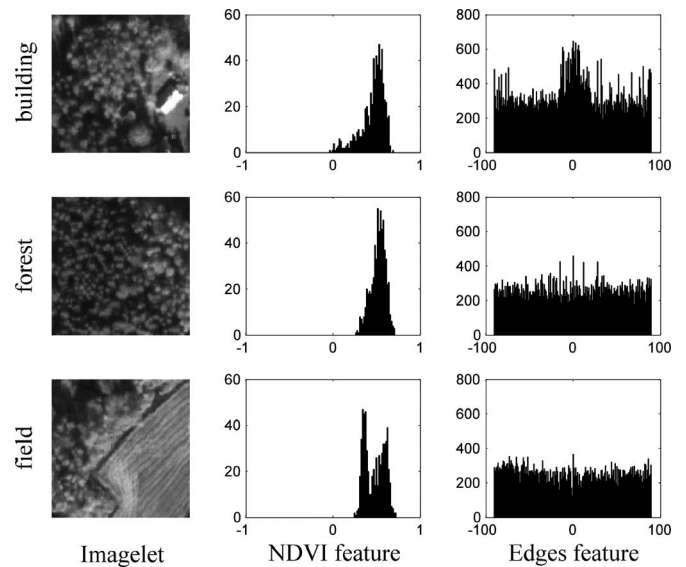


Fig. 5. Three sample imagelets from 2002 representative of classes building, forest, field, and their features.

values for vegetation and values closer to zero for the building. The NDVI histogram of the second imagelet, containing forest only, does not include as low NDVI values as the first imagelet with the building. Also, the NDVI histogram is symmetrical around the modal value, whereas the edges histogram is rather uniform and reveals no man-made structure in the imagelet. The third imagelet has a clear bimodal NDVI histogram because it contains forest and a ploughed field with bare soil in approximately equal proportions. The edges histogram does not show any clear structure here either.

C. Feature Extraction and SOM Training

Spectral information was extracted from the multispectral imagery: Average RGB and color moment features were calculated discarding the NIR channel, whereas the NDVI histogram feature was computed using red and NIR bands. Textural and spatial features (edges histogram and texture feature, respectively) were extracted from the higher spatial resolution panchromatic data. Table I sums up features dimensionality and the data source they were extracted from.

The map sizes were set to 64×64 units for the five visual features SOMs, and 71×71 for the coordinate SOM. There were, therefore, on average, $5041/4096 \approx 1.23$ imagelets mapped in each unit of the visual SOMs and exactly one image per unit on the coordinate map. Each feature vector was used 100 times in training the SOM according to (1) and (2). Fig. 6 illustrates the visual labels, i.e., the most representative imagelets, on a SOM calculated for the textural feature. It is interesting to notice that water regions were mapped into two separate areas. This was because the sea was calm in 2002 (on the left of the map), but quite wavy in 2005 (in the top-right corner of the map): images of calm and wavy water areas have rather different textural properties.

The distributions of the ground truth classes on the six different feature SOMs are shown in Fig. 7. The more localized the distributions of feature vectors are on the map, the better it

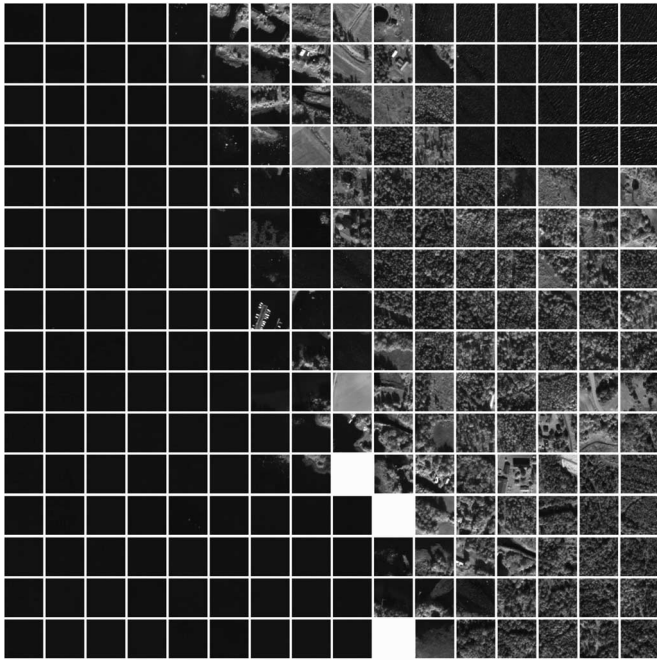


Fig. 6. Organization of imagelets by their texture on a 16×16 SOM surface.

is for detection. In that sense, color moment and edges features seem to be good candidates for building detection.

V. USING PICSOM FOR DETECTING MAN-MADE STRUCTURES AND CHANGES

The PicSOM system was originally developed for browsing databases of images. We explain here how we used and extended the PicSOM for detecting man-made structures, or changes—both in supervised and unsupervised manners.

A. Interactive Detection of Man-Made Structures With CBIR

CBIR techniques have already been used for the remote sensing image archive management. In [12], the user is asked to label several pictures as relevant or nonrelevant at each query round. A study reported in [31] also includes query by pictorial examples.

In this paper, we have used the PicSOM CBIR system to interactively find imagelets containing man-made objects such as buildings or roads. The system first displays a random selection of imagelets in a web browser. The user then selects all imagelets containing man-made objects—or anything else but water and forest—and sends this information back to the system by pressing the “Continue query” button. In the forthcoming query rounds, the user can then focus the query with their selections more precisely to more specific semantic targets, such as buildings, roads, or clearcuts. Fig. 8 shows the user interface of the system in the middle of an interactive query session. The user has selected some man-made objects shown in the middle of the browser window. On the top part, the distribution of those imagelets is shown on the six different SOMs. In the bottom of the interface, some of the new imagelets returned by the system are shown to the user.

B. Automated Detection of Man-Made Structures

As described in the end of Section III-B, models for semantic imagelet classes, such as buildings or other man-made structures, can be created with different features from the ground-truth-labeled training data. Imagelets from different areas or different acquisition dates can then be classified by the system to detect the presence of similar structures. In order to achieve this, the qualification value assigned to each imagelet by the PicSOM system is simply interpreted as a discrimination value, which indicates the likeliness that the specific imagelet belongs to that semantic class.

After calculating the discrimination value, the PicSOM will also order the test data by their qualification values in the decreasing order of similarity to the class model. If one, instead of the ordered list, desires the system to give a decisive output for each imagelet (either belonging or not to the studied semantic class), a threshold value has to be selected. Standard pattern classification techniques for selecting the threshold and evaluating the performance of the classification system can then be used in assessing and tuning the detection accuracy. Such techniques include, e.g., receiver operating characteristic (ROC) curves [32] and recall-precision graphs.

One may note that the above-described detection procedure is analogous to the interactive CBIR operation, in the sense that the labeled training data have now replaced the imagelets graded by the user as relevant. Similarly, the test data are in the role of the yet unseen database images.

C. Supervised Content-Based Change Detection

We outlined above how the PicSOM system can be used to create a multifeatureal discrimination model for a semantic class of imagelets. If the model is applied to a pair of imagelets, both depicting the same geographical location, but acquired at different dates, one can perform the change detection.

Assuming that the system produces larger discrimination values for imagelets that portray man-made structures than those that do not, then a temporal increase in this value indicates that a new man-made structure has probably appeared. Vice versa, if the discrimination value decreases between the two acquisition dates, then it is likely that a building or other unnatural structure has disappeared.

Again, all the imagelet pairs can be sorted by a decreasing order of change in this discrimination value. In that ordering, the first imagelets are those that most probably depict areas where a new building or other structure has appeared. By selecting a proper threshold value or taking a preset number of the top-ranked imagelets, one can check if new structures really have appeared. Fig. 9 presents our change detection approach based on SOMs. The basic framework—decomposition into imagelets, feature extraction, and mapping on the SOM—is common to a content-based supervised and unsupervised detection of man-made structures or changes.

It must be noted that this kind of change detection is supervised, in the sense that one needs to have labeled the training data in order to create the class models. The models can be used for change detection in the data temporally and

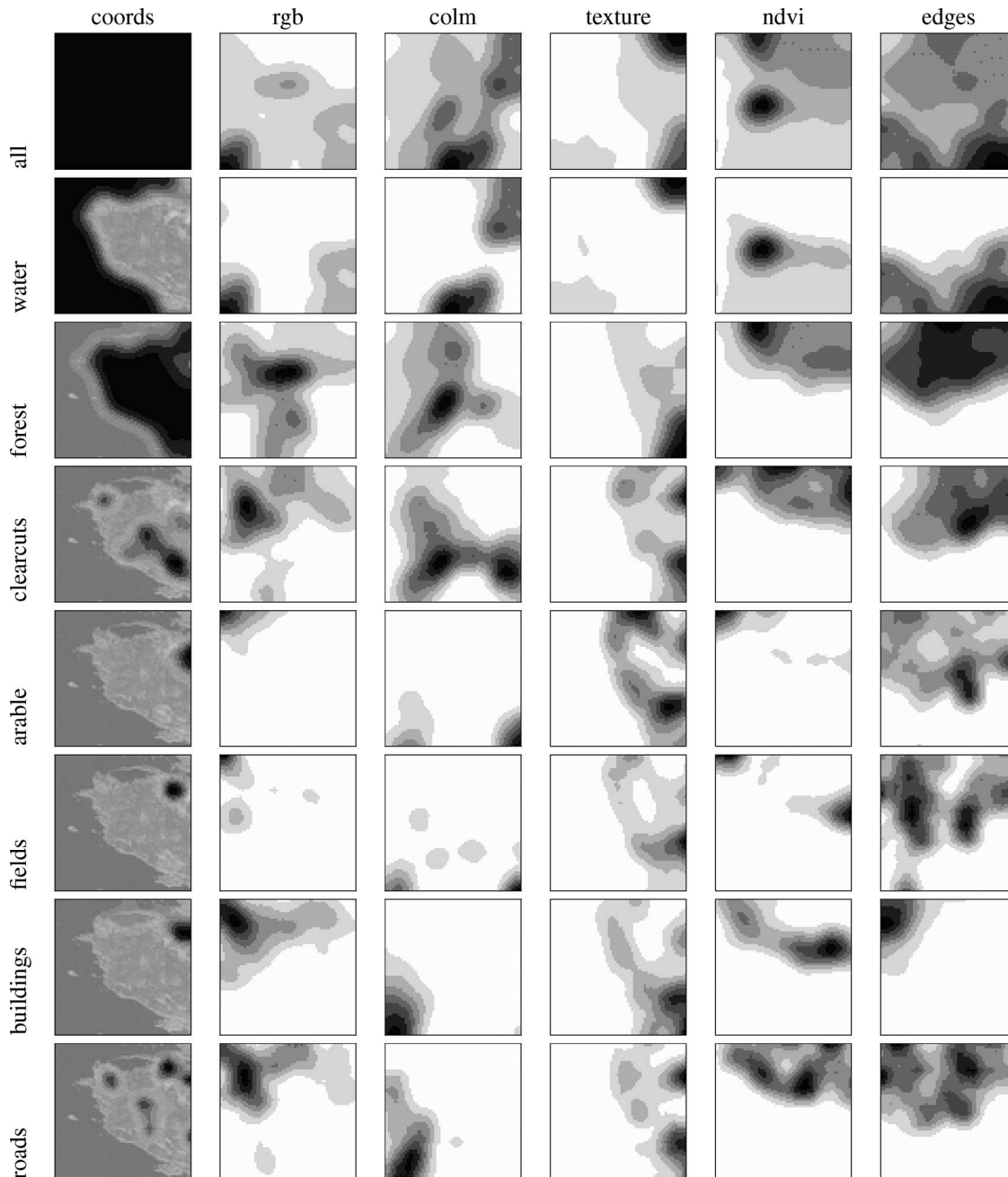


Fig. 7. Distributions of the ground truth classes on the different feature SOMs. Darker shades present areas of denser distribution.

geographically independent from the training data, as long as the training samples from that semantic class still match the analyzed imagelets accurately enough. Therefore, the models created for detecting buildings in Northern Europe might not be applicable, for example, to Mediterranean regions.

D. Unsupervised Content-Based Change Detection

We also devised an unsupervised method for finding pairs of imagelets—one from year 2002 and the other from 2005—which differed the most in the sense of some of the extracted features. The dissimilarity between imagelets was again defined solely on the SOMs. This means that we did not calculate any pixelwise differences between the imagelets, but

defined the changes relatively to a particular feature extraction scheme. In this manner, the change detection is less dependent on small variations in the absolute image coordinates due to inaccurate coregistration.

Some variations in the imagelets are due to various forms of noise (e.g., varying illumination conditions, overlapping shadows), and do not correspond to true changes in the land cover. We assumed that the differences caused by a noise lead to situations where the BMU for the calculated feature vector remained the same, or was moved to a neighboring SOM unit. Only true changes in the imagelet's content should then give rise to such a substantial change in the feature vector's value, that its projection on the SOM surface would be moved to a significantly different location. The substantiality of the change

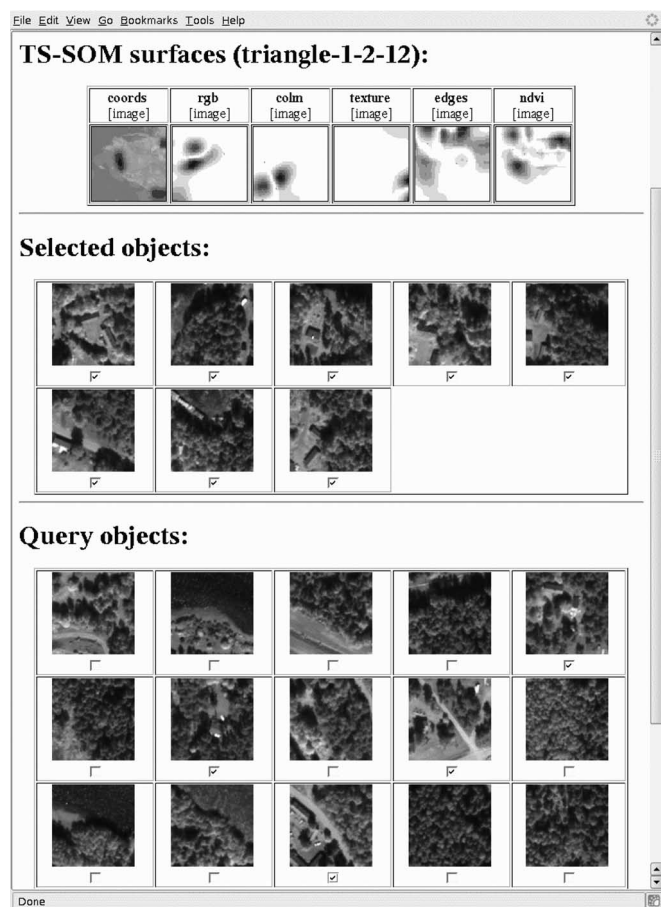


Fig. 8. Part of the web user interface of the PicSOM system in an interactive query for man-made objects.

can therefore be measured as the distance on a SOM between the BMUs of the imagelet's feature vectors from different years.

The proposed unsupervised change detection technique was as follows. For each imagelet pair from years 2002 and 2005—excluding the water areas—we solved the two BMUs on one particular feature's SOM. The Euclidean distances between the BMUs were then calculated, and the imagelet pairs ordered by descending pairwise BMU distance. A number of imagelet pairs were then regarded as the locations where the most substantial changes had taken place.

The same procedure can be repeated for all the five visual features defined in Section IV. A feature selection stage can also determine the combination of features that performs best for a given change detection task. In case more than one feature was involved in the change detection framework, the between-BMU distances on different feature SOMs had to be combined. The distances were first normalized on each feature map separately, into the range $[0, 1]$, by histogram equalization. That process aimed to compensate for cases when the average distances were larger on one map than on the others. These equalized distances are then summed to produce the final value indicating the amount of change.

One must note that this unsupervised detection scheme treats all kinds of changes equally. As a result, a good detection performance cannot be expected for any specific type or semantic class of changes. For such needs, provided that the labeled

training data are available, the supervised detection technique described in Section V-C will be more appropriate.

VI. RESULTS

The proposed methods were evaluated quantitatively and qualitatively for different tasks: interactive and off-line detection of built-up areas, and supervised or unsupervised change detection. The proposed change detection methods were also quantitatively compared to other change detection methods.

A. PicSOM for Detection of Buildings

The performance of the proposed methods for detecting buildings in very high resolution images was evaluated quantitatively by the ROC curve analysis, and both quantitatively and qualitatively in an interactive experiment on a validation set.

1) *Using the 2002 Training Set to Detect Buildings in the 2005 Set:* Considering imagelets from the study scene and their associated labeling in 2002 and 2005, we searched for a good combination of features to detect buildings. A training set was made up of the 5041 imagelets of 2002 from the study scene, while the testing set contained the imagelets of the 2005 study scene. Those sets included 244 and 266 buildings, respectively. Twofold cross validation on the training set was used with a sequential forward selection (SFS) [33] to select those features which maximize the area under the ROC curve [32], denoted here as auc. NDVI, edges, and color moment features were found to form the optimal feature set, whose combination gave the highest area under curve $\text{auc} = 0.94$. Some ROC curves for the testing set are shown in Fig. 10. The color moment feature alone performed well too, but it did not have as good a sensitivity in the beginning (left part) of the ROC curve as the combination of those three features.

2) *Interactive Validation:* In this experiment, the feature SOMs were trained on the imagelets of both years from the study scene. A validation set was taken from the same original QuickBird images, but in different geographical extents. This set was cut into $60 \times 90 = 5400$ imagelets for each year, in the same manner as the training set in Section II. PicSOM interface was then used to retrieve, from the 2005 validation set, imagelets that contain buildings.

Two subexperiments were conducted and repeated by four persons; two of them being external to the present paper. In experiment A, the plain CBIR approach with query by pictorial example and relevance feedback was evaluated. In the first query round, the PicSOM presented to the user 20 imagelets that were picked randomly from the validation set. The user would then select those imagelets containing man-made structures other than roads (e.g., buildings or boats). These imagelets marked as relevant were fed back to the PicSOM, that returned in the next query round 20 other imagelets similar to those already selected. The experiment was stopped after ten query rounds. In experiment B, the feedback was brought to the PicSOM both by the user (relevance feedback as in A) and by the class model (imagelets containing buildings in 2002 training set), in equal weight. The initial selection of 20 imagelets presented to the same user was identical in both experiments, but different between the four experimenters.

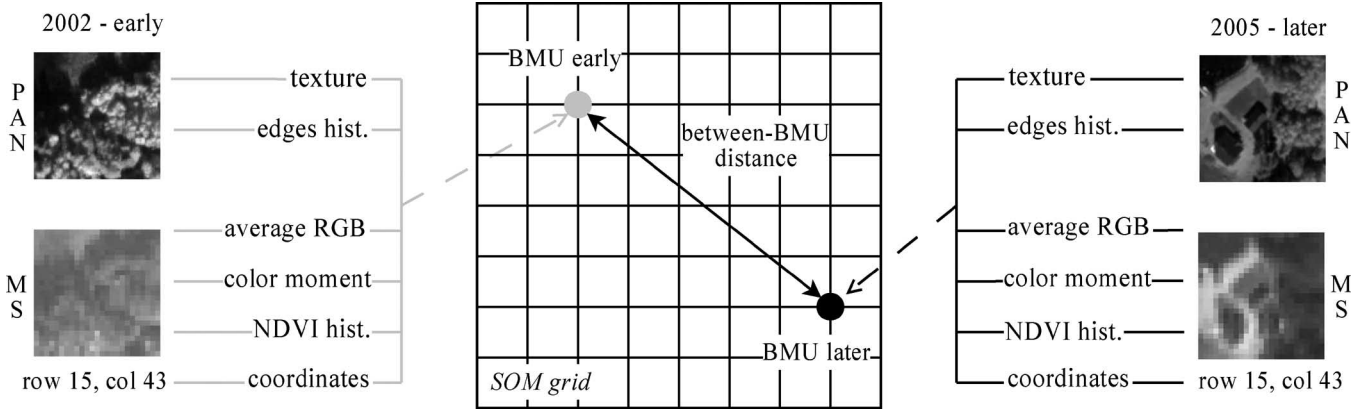


Fig. 9. Overview of the SOM-based change detection framework. Features are extracted from appropriate data sources—panchromatic or multispectral—in the earlier (2002) and later (2005) imagelet. The imagelets are then mapped on the SOM grid, where the distance between the BMUs reflects the change magnitude between the two dates.

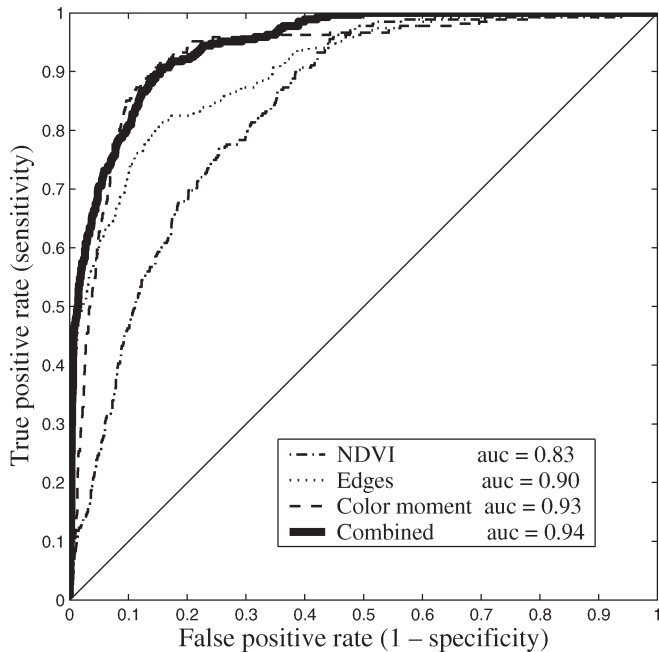


Fig. 10. ROC curves of different features for building detection in 2005 set.

The number of relevant images returned versus query round is shown in Fig. 11. The lines are the average results over the four experimenters, with a standard deviation $\pm\sigma$ marked by “*” and “o.” On the average, only 1.5 imagelets from the initial random set of 20 imagelets contained a building. There were probably only a few imagelets containing houses in the validation set, as there were in the training set (only about 200). In both experiments, the PicSOM returned a comfortable majority of relevant images already after three query rounds. The system seems to perform slightly better when the class model is integrated into the feedback mechanism. However, the result of experiment A suggests that the plain CBIR system is also performing well in the interactive building detection.

Similar results were achieved when visually selecting clearcuts or arable land as a target. This already shows a used case of the PicSOM system with the remote sensing data, as a supervised, general purpose, and interactive tool for satellite image annotation, by visual and intuitive querying. A proposed

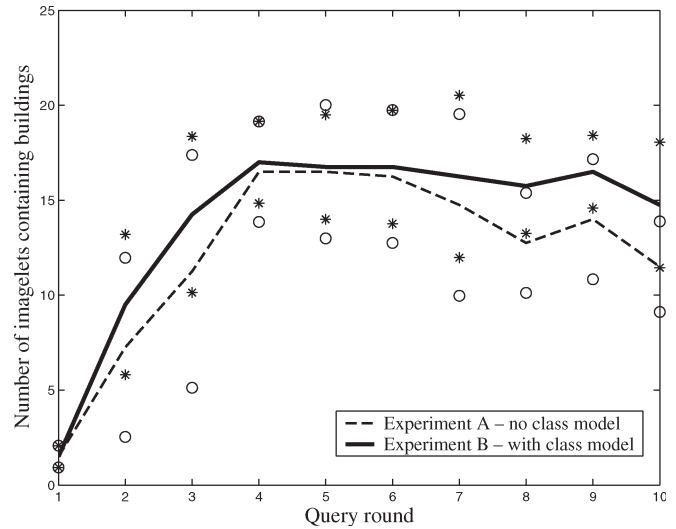


Fig. 11. Average number of relevant imagelets versus query round, (A) without and (B) with the class model.

visual output consists of “lighting up,” in the studied satellite image, only the imagelets in which objects of interest have been detected—e.g., buildings in Fig. 12.

B. Supervised Change Detection

Our proposed supervised change detection method was evaluated in its ability to find buildings that have appeared between years 2002 and 2005. The labeled database contained 40 imagelet pairs where this change really took place. As an independent testing set was not available, we selected the optimal set of features for change detection by optimizing the ROC auc value for the training set of 2002–2005 imagelet pairs with the SFS algorithm. A baseline change detection method was also used in a quantitative comparison.

1) *Baseline:* As a baseline for supervised change detection method, we used a postclassification comparison [21]. Panchromatic and multispectral data were merged and classified by a standard maximum likelihood. Prior probabilities of classes were set to 0.3 for water and forest, and 0.08 for the other five classes. Training areas for each class were manually delineated

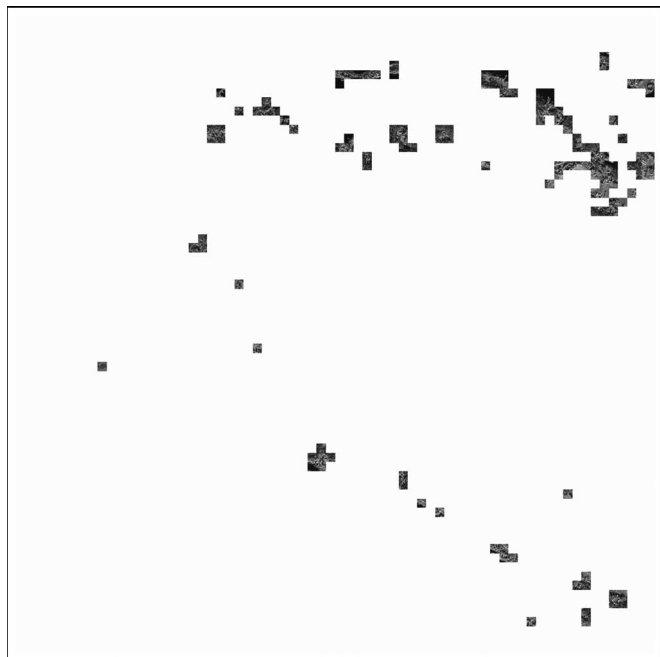


Fig. 12. Imagelets containing buildings (2002 study scene), mapped back to the original image coordinates.

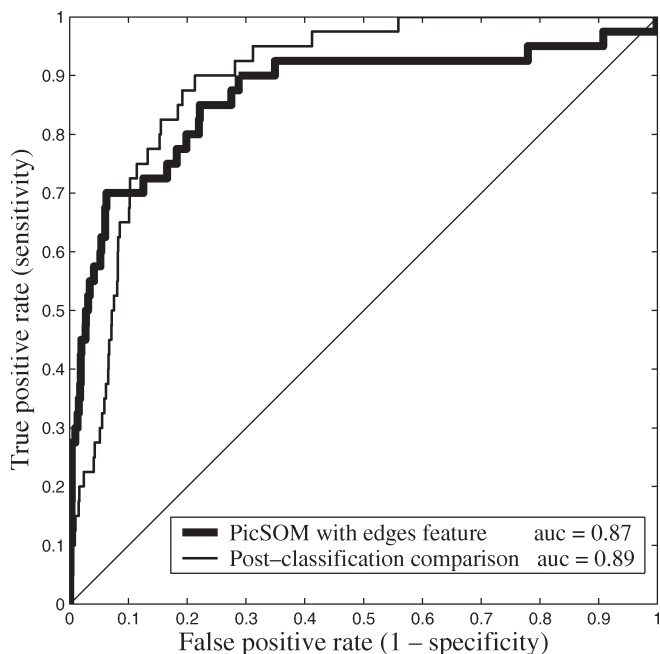


Fig. 13. ROC curve of the supervised detection of buildings appeared between 2002 and 2005—PicSOM and baseline method.

on the 2002 and 2005 full scenes by a visual inspection. For each year, exactly half of the buildings from the ground truth built in Section II-C were used in the training.

In order to allow a comparison with our method, the two classification images were cut into imagelets. For each imagelet pair, the number of pixels classified as buildings in 2005 and that were not buildings in 2002 was counted. By focusing on the detection of appeared buildings, imagelet pairs were sorted by this decreasing value and compared to the ground truth, to build the ROC curve in Fig. 13.

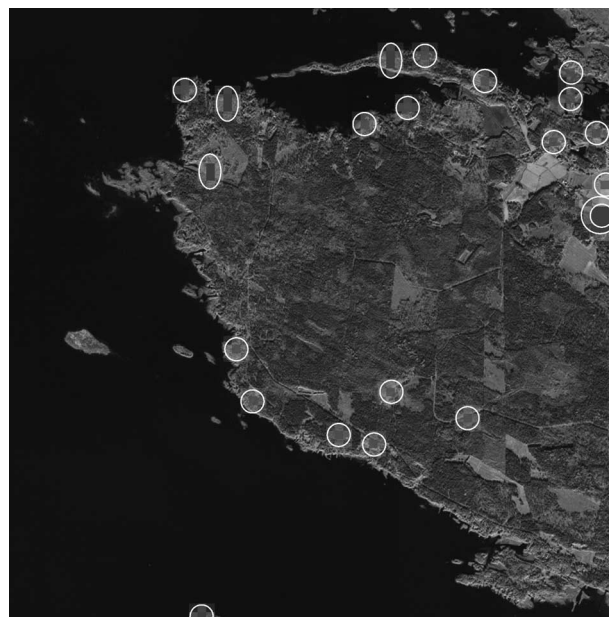


Fig. 14. Thirty most prominent changes between 2002 and 2005 detected in the edges SOM are marked on the map—these areas match well with true changes in buildings.

2) *Proposed Method*: It turned out that the best detection accuracy was obtained by using the edges feature alone. Its ROC curve with $auc = 0.87$ is shown in Fig. 13. Other features and feature combinations performed rather poorly in this task, with an auc at best close to 0.6. Our content-based supervised change detection method is more accurate than the baseline method at the beginning of the ROC curve, with less false positives. Later, the baseline outperforms the proposed method. However, in the baseline method, half of the buildings in the 2005 were manually delineated during training, whereas our method did not use any knowledge about buildings in the 2005 image—only in 2002.

Setting the detection threshold in our method at 71 imagelet pairs, 12 true positives were returned (effective changes in built-up areas) and 59 false positives. A higher threshold value returned 11 true positives and 19 false positives. These 30 locations where changes were detected are shown with circles and red color over the test area in Fig. 14.

C. Unsupervised Change Detection

We evaluated our unsupervised change detection method with the ground truth available for buildings in 2002 and 2005. Imagelets containing only water were discarded—they would otherwise have contributed too much in the set of the “most changed” imagelets. The best feature combination was searched with SFS, maximizing the area under the ROC curve.

1) *Baseline*: We considered some variations of the image difference [21] as baseline methods for unsupervised change detection. Difference images were computed for panchromatic and for each multispectral channel. The square root of the sum of square differences between each multispectral channel was also computed, resulting in one change magnitude image.

No threshold was applied to the change magnitude images. They were cut into imagelets, that were then sorted by a

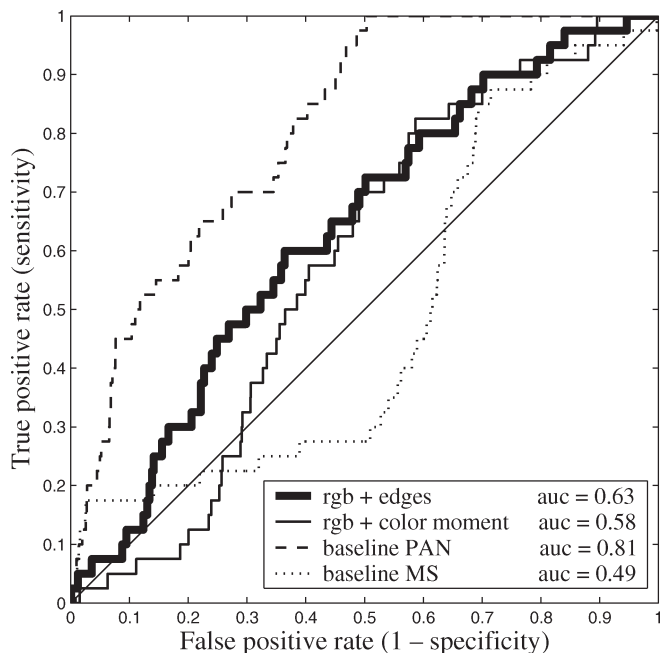


Fig. 15. ROC curves for unsupervised change detection in built-up areas.

decreasing order of the sum of the change magnitude over all imagelet pixels. By a comparison to the ground truth, ROC curves were produced—Fig. 15.

2) *Proposed Method*: The best result was obtained with the combination of average RGB and edges features, which gave $\text{auc} = 0.63$, as shown in Fig. 15. The result is quite reasonable as the procedure detects all the changes, not only changes in buildings. Therefore, false positives appear quite often, but not so often as with random picking. However, if we would like to really demonstrate this method's potential, we would need a ground truth that includes all changes, not only in buildings.

Image differencing on panchromatic channel performed better than our method. Similar results were obtained with the NIR channel only, although slightly less accurate. The three optical channels considered independently in the image differencing performed quite poorly, with an ROC curve below the diagonal, similar to that obtained with the combination of all multispectral channels—Fig. 15.

The choice of the threshold (how many imagelet pairs are considered to exhibit significant changes) is problem dependent both in the supervised and unsupervised methods. For instance, when monitoring a strategic site, missing an appearing target is worse than returning false positives. In earlier work [25], we heuristically proposed to set the threshold at the square root of the number of imagelet pairs in the database. We empirically observed that areas where color moment feature had substantial differences matched well with clearcuts—i.e., forested areas that had been cut in between 2002 and 2005.

VII. DISCUSSION

A. Detection of Man-Made Structures

The set of features that performed best for detecting buildings in 2005 after training the PicSOM on the 2002 database

(Section VI-A) was formed by NDVI, color moment, and edges features. In those cases, the surrounding or the context of a building is a useful hint, because buildings usually appear in groups and in mutually similar environments. Yet, this feature combination did not work well for supervised change detection. This may be because color moments and NDVI very seldom change when a new building is built, as they tend to appear in the vicinity of the existing buildings. Therefore, feature combinations that are good for detecting buildings are not necessarily good for detecting changes in buildings. In other words, the context may be important for detecting a building, but it is often nonchanging when such a building appears or disappears in a short time frame, like three years in our study.

The targets of interest were small-sized buildings, which made their detection even more challenging. Small buildings can easily remain undetected by a feature extraction if masked by high trees or shadows. Detection may also fail when the building is located on an imagelet border; thus, only a part of it is present in the imagelet. Moreover, the features should account for different target sizes and adapt to various spatial resolutions, depending on the sensor. The importance of scale for satellite image description has been emphasized in [28].

B. Change Detection

Pixel-based change detection in very high resolution imagery is a challenging task, which is limited by the requirement of pixel or subpixel accuracy registration. In the context of change detection, a clear advantage of the decomposition in imagelets is that it relaxes this constraint. Slight misregistrations remained between the 2002 and 2005 images, but it did not influence the efficiency of our methods.

A building appearing between 2002 and 2005 was a rare event—only 40 imagelet pairs in our database. In addition, if the earlier imagelet already contains a building, the appearance of a new building in the later image could remain undetected. In such a case, the increase of the discriminative value associated to buildings might not be significant enough that it would suggest a new building has appeared. Furthermore, our current ground-truth labeling does not take into account the number of objects of a given class in an imagelet.

The proposed change detection method performed slightly better than the baseline method in the supervised case, while needing a less user interaction than a postclassification change detection. However, our unsupervised change detection method did not perform as well as simple image differencing on panchromatic or NIR channel. This may be because the PicSOM system was not tuned for unsupervised detection, but rather, it takes advantage of the relevance feedback in the supervised case—either from class models or user interaction.

Some changes were attributed to the locations containing in fact the same land cover (often forest) in 2002 and 2005. Those cases were mainly vegetation changes between the beginning and end of summer. While this could be interesting for season monitoring applications, it dragged the PicSOM away from the goal of detecting appearing or disappearing man-made structures. A clear definition within the system of what changes are of interest is needed. How this should be done remains

open. This could be partly circumvented by radiometrically calibrating the images before loading them into the PicSOM.

A way to refine our change detection method would be to provide two content targets to the PicSOM: a content from which the change occurs (earlier target) and a content to which the change occurs (later target). This would allow an intuitive and interactive definition of interesting changes by the user—e.g., by selecting imagelets containing forest as the earlier target and buildings as the later target, the system would detect newly constructed buildings in forested areas.

The approach developed here could handle more than two satellite images for change detection applications. One could train the PicSOM on the imagelets extracted from all images available up to a given time, then query the database for imagelets representative of the objects of interest. The system would then return the imagelets, where significant changes have been detected in the later scene, according to the distance on a feature SOM between the BMUs of earlier and later imagelets.

C. Size of Imagelets and Objects of Interest

The influence of the imagelet size has to be studied more carefully. If smaller than the objects of interest, it is expected that an imagelet would not reconstitute all spatial or spectral properties of the target. On the other hand, larger imagelets would contain a too great proportion of perturbing noninteresting content compared to the spatial extensions of the objects of interest—typically in our study scene, a small isolated building surrounded by forest. In that case, the imagelets would be clustered in the SOM according to their dominant content, which would not necessarily be the content of interest.

Therefore, the size of imagelets has to be adjusted so that the information they contain is representative of the target. Similarly, the imagelet size should adapt to the scale of interesting changes. In an operative system, the expected dimensions of the interesting targets or scale of changes could be specified by the user or preset, depending on the application, then could be used to determine an appropriate imagelet size.

The 100×100 pixel imagelets, extracted from QuickBird images, seemed to provide a tradeoff between the two undesirable situations. Luckily (and surprisingly), not too many buildings in the study scene were split into two or more imagelets. In order to reduce the consequences of “cutting” an object of interest into several nonoverlapping imagelets (namely, generating “artificial” objects on the borders of imagelets), overlapping imagelets could be used.

The methods should also make use of the spatial information with respect to imagelets or context: the neighboring imagelets of a given imagelet may bring a useful information for the analysis of satellite images. For example, built-up areas are often concentrated, and the proposed methods currently do not specifically handle objects bigger than the imagelets. The use of the coordinate feature could help detecting those bigger structures by considering the neighboring imagelets. The coordinate feature introduced in this paper (but not really used in the described experiments) is a step in that direction, as the relevance spreading in this case is directly related to favoring neighbors of relevant imagelets in retrieval.

VIII. CONCLUSION AND FUTURE WORK

We have presented how a CBIR system, PicSOM, can be used with the remote sensing images for tasks like detection of man-made structures, as well as for supervised or unsupervised change detection. The approaches rely on the decomposition of a satellite image into several thousands of small images or imagelets to generate a database. After extracting features from the imagelets and training SOMs, the imagelet database can be queried. Relevant images can be retrieved either automatically or interactively by a user in a visual and intuitive manner. The same framework allows the detection of man-made structures as well as changes between two images or both tasks simultaneously.

Quantitative and qualitative evaluations of the proposed methods were performed. The results were encouraging, considering we have presented a new approach to the challenging problem of change detection in very high resolution images. Future work might include better shape descriptors for the purpose of building detection. The same kind of approach could be applied to radar images, in which case the importance of the adapted feature extractors would be critical. Preliminary results on a database of imagelets constructed from an ASAR scene were encouraging.

The versatility of PicSOM will allow several applications to be embedded in the same system, only to be differentiated by the type of query. Further research will aim at a fully operative and interactive system built around the PicSOM. Possible applications of the proposed methods include fast interactive analysis of satellite images. After the SOMs have been trained, a new image can be presented to the system in order to be automatically analyzed or visually queried for structures of interest. The approach could also be used in long-term monitoring of strategic sites. In such a setting, each new image would first be analyzed, then integrated in the training set for improving the knowledge of the system about the area under surveillance.

ACKNOWLEDGMENT

The authors would like to thank V. Viitaniemi and M. Sjöberg for taking part in the interactive retrieval experiment, and also the anonymous reviewers for their fruitful comments.

REFERENCES

- [1] G. Healey and A. Jain, “Retrieving multispectral satellite images using physics-based invariant representation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 842–846, Aug. 1996.
- [2] K. Seidel, R. Mastropietro, and M. Datcu, “New architectures for remote sensing image archives,” in *Proc. IGARSS—Scientific Vision Sustainable Development*, Singapore, Aug. 1997, vol. 1, pp. 616–618.
- [3] K. Seidel, M. Schröder, H. Rehrauer, and M. Datcu, “Query by image content from remote sensing archives,” in *Proc. IGARSS*, Seattle, WA, Jul. 1998, vol. 1, pp. 393–396.
- [4] M. Schröder, K. Seidel, and M. Datcu, “User-oriented content labelling in remote sensing image archives,” in *Proc. IGARSS*, Seattle, WA, Jul. 1998, vol. 2, pp. 1019–1021.
- [5] K. Seidel and M. Datcu, “Architecture of a new generation of remote sensing ground segments,” in *Proc. 19th EARSel Symp. Remote Sens. 21st Century*, Valladolid, Spain, Jun. 1999, pp. 223–228.
- [6] M. Schröder, H. Rehrauer, K. Seidel, and M. Datcu, “Interactive learning and probabilistic retrieval in remote sensing image archives,” *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 5, pp. 2288–2298, Sep. 2000.

- [7] M. Datcu, H. Daschiel, A. Pelizzari, M. Quartulli, A. Galoppo, A. Colapicchioni, M. Pastori, K. Seidel, P. G. Marchetti, and S. D'Elia, "Information mining in remote sensing image archives—Part A: System concepts," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 12, pp. 2923–2936, Dec. 2003.
- [8] J. Li and R. N. Narayanan, "Integrated spectral and spatial information mining in remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 673–685, Mar. 2004.
- [9] H. Daschiel and M. Datcu, "Information mining in remote sensing image archives—System evaluation," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 1, pp. 188–199, Jan. 2005.
- [10] M. Schröder and A. Dimai, "Texture information in remote sensing images: A case study," in *Proc. WTA*, Freiburg, Germany, 1998, pp. 181–192.
- [11] M. Schröder, "Interactive learning in remote sensing image databases," in *Proc. IGARSS*, Hamburg, Germany, Jun./Jul. 1999, vol. 5, pp. 2725–2728.
- [12] M. Ferecatu, N. Boujemaa, and M. Crucianu, "Active relevance feedback for interactive satellite images retrieval," in *Proc. ESA-EUSC Workshop Image Inf. Mining—Theory and Application Earth Observation*, Frascati, Italy, Oct. ESA WPP-257, 2005.
- [13] M. Campedel, E. Moulines, H. Maître, and M. Datcu, "Feature selection for satellite image indexing," in *Proc. ESA-EUSC Workshop Image Inf. Mining—Theory and Application Earth Observation*, Frascati, Italy, Oct. ESA WPP-257, 2005.
- [14] L.-K. Soh and C. Tsatsoulis, "Segmentation of satellite imagery of natural scenes using data mining," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 2, pp. 1086–1099, Mar. 1999.
- [15] T. Kohonen, *Self-Organizing Maps*, 3rd ed., vol. 30. Berlin, Germany: Springer-Verlag, 2001.
- [16] W. Wan and D. Fraser, "Multisource data fusion with multiple self-organizing maps," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1344–1349, May 1999.
- [17] J.-B. Henry, K. Fellah, S. Clandillon, B. Allenbach, and P. De Fraipont, "Earth observation and case-based systems for flood risk management," in *Proc. IGARSS*, Toronto, ON, Canada, Jun. 2002, vol. 3, pp. 1496–1498.
- [18] J.-F. Zhang, L.-L. Xie, and X.-X. Tao, "Change detection of earthquake-damaged buildings on remote sensing image and its application in seismic disaster assessment," in *Proc. IGARSS*, Toulouse, France, Jul. 2003, vol. 4, pp. 2436–2438.
- [19] S. Gopal and C. Woodcock, "Remote sensing of forest change using artificial neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 2, pp. 398–404, Mar. 1996.
- [20] I. Niemeyer, S. Nussbaum, and M. Canty, "Automation of change detection procedures for nuclear safeguards-related monitoring purposes," in *Proc. IGARSS*, Seoul, Korea, Jul. 2005, vol. 3, pp. 2133–2136.
- [21] T. M. Lillesand, R. W. Kiefer, and J. W. Chipman, *Remote Sensing and Image Interpretation*, 5th ed. Hoboken, NJ: Wiley, 2004.
- [22] PicSOM Development Group, *PicSOM Online Demonstration*, 1998–2006, Helsinki Univ. Technol. [Online]. Available: <http://www.cis.hut.fi/picsom>
- [23] J. Laaksonen, M. Koskela, S. Laakso, and E. Oja, "Self-organizing maps as a relevance feedback technique in content-based image retrieval," *Pattern Anal. Appl.*, vol. 4, no. 2/3, pp. 140–152, Jun. 2001.
- [24] J. Laaksonen, M. Koskela, and E. Oja, "PicSOM—Self-organizing image retrieval with MPEG-7 content descriptions," *IEEE Trans. Neural Netw.*, vol. 13, no. 4, pp. 841–853, Jul. 2002.
- [25] M. Molinier, J. Laaksonen, J. Ahola, and T. Häme, "Self-organizing map application for retrieval of man-made structures in remote sensing data," in *Proc. ESA-EUSC Workshop Image Inf. Mining—Theory and Application Earth Observation*, Frascati, Italy, Oct. ESA WPP-257, 2005.
- [26] T. Häme, I. Heiler, and J. San Miguel-Ayanz, "An unsupervised change detection and recognition system for forestry," *Int. J. Remote Sens.*, vol. 19, no. 6, pp. 1079–1099, Apr. 1998.
- [27] J. Laaksonen, M. Koskela, and E. Oja, "Class distributions on SOM surfaces for feature extraction and object retrieval," *Neural Netw.*, vol. 17, no. 8/9, pp. 1121–1133, Oct./Nov. 2004.
- [28] H. Rehrauer, "The role of scale for image description and image query: An application to remote sensing image," Ph.D. dissertation, Swiss Federal Inst. Technol., Zurich, Switzerland, Mar. 2000.
- [29] H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-8, no. 6, pp. 460–473, Jun. 1978.
- [30] V. D. Lecce, A. Guerriero, G. Dimauro, S. Impedovo, G. Pirlo, and A. Salzo, "Image basic features indexing techniques for video skimming," in *Proc. ICIAP*, 1999, pp. 715–720.
- [31] G. Milillo, L. Dini, Serra, Guerriero, and D. Lecce, "Grid enabled feature extraction and signature generation system for EO images database," in *Proc. ESA-EUSC Workshop Image Inf. Mining—Theory and Application Earth Observation*, Frascati, Italy, Oct. ESA WPP-257, 2005.
- [32] D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*. New York: Wiley, 1966.
- [33] J. Kittler, *Feature Set Search Algorithms*, C. Chen, Ed. Gronigen, The Netherlands: Sijthoff and Noordhoff, 1978.



Matthieu Molinier was born in France in 1980. He received the Engineering degree from the École Nationale Supérieure de Physique de Strasbourg and the M.S. degree in image processing from the Université Louis Pasteur, Strasbourg, France, both in 2004.

Since October 2004, he has been a Research Scientist with the VTT Technical Research Centre of Finland, Espoo, and a postgraduate student with the Laboratory of Computer and Information Science, Helsinki University of Technology, Espoo. His current research interests include change detection in satellite imagery, machine learning, video processing, and motion tracking.



Jorma Laaksonen (S'96–A'97–SM'02) received the Dr.Sci.Tech. degree from the Helsinki University of Technology (TKK), Espoo, Finland, in 1997.

He is currently a Senior Research Scientist with the Laboratory of Computer and Information Science, TKK. He has authored several journals and conference papers on pattern recognition, statistical classification, and neural networks. His research interests include content-based image retrieval and recognition of handwriting. He is a Founding Member of the SOM and LVQ Programming Teams and the PicSOM Development Group.

Dr. Laaksonen is a member of the International Association of Pattern Recognition (IAPR) Technical Committee 3: Neural Networks and Machine Learning.



Tuomas Häme received the M.S. and Ph.D. degrees from the University of Helsinki, Helsinki, Finland, in 1979 and 1992, respectively.

Since 1979, he has been with the VTT Technical Research Centre of Finland, Espoo, where he is a Manager with the Remote Sensing Team. Since 2005, he has held a Research Professorship on Earth Observation (EO) with VTT. He was a Visiting Researcher with North Carolina State University in 1989, and with the Joint Research Centre, Italy, from 1995 to 1996. He is an expert in remote sensing in forestry, and has developed particularly methods for forest change detection, forest area mapping, and biomass estimation from local to continental extents. He has managed several international projects on EO and is working in close cooperation with value-adding industry.