SELF-ORGANIZING MAP APPLICATION FOR RETRIEVAL OF MAN-MADE STRUCTURES IN REMOTE SENSING DATA

Matthieu Molinier$^{(1)}$**, Jorma Laaksonen**$^{(2)}$**, Jussi Ahola**$^{(1)}$**, and Tuomas Häme**$^{(1)}$

$^{(1)}$*VTT Technical Research Centre of Finland, Information Technology, P.O. Box 1201, FI-02044 VTT, Finland, Email: firstname.lastname@vtt.fi*

$^{(2)}$*Helsinki University of Technology, Lab. of Computer and Information Science, P.O. Box 5400, FI-02015 HUT, Finland, Email: firstname.lastname@hut.fi*

ABSTRACT

Self-Organizing Maps (SOMs) have been successfully applied to content-based image retrieval (CBIR). In this study, we investigate the potential of PicSOM, an image database browsing system, applied to remote sensing images. Databases of small images were artificially created, either from a single satellite image for object detection, or two satellite images when considering change detection. By visually querying those databases, it was possible to detect targets like houses, roads or man-made structures, as well as changes between two QuickBird images. Preliminary results were encouraging, and open a full range of applications, from structure detection to change detection, to be embedded in a same operative system.

Key words: content-based information retrieval, self-organizing maps, high resolution satellite images, man-made structure detection, change detection.

1. INTRODUCTION

Remote sensing data include large images, often exceeding $10,000 \times 10,000$ pixels (e.g. very high resolution panchromatic QuickBird images). Processing those large images can be computationally unpractical, especially for tasks like objects detection. At the same time, there is an increasing number of Earth Observation data collected and to be processed each day. These needs have led to the emergence of content-based image retrieval systems, for remote sensing image archive management [1–7], or satellite image annotation and interpretation [2, 8, 9].

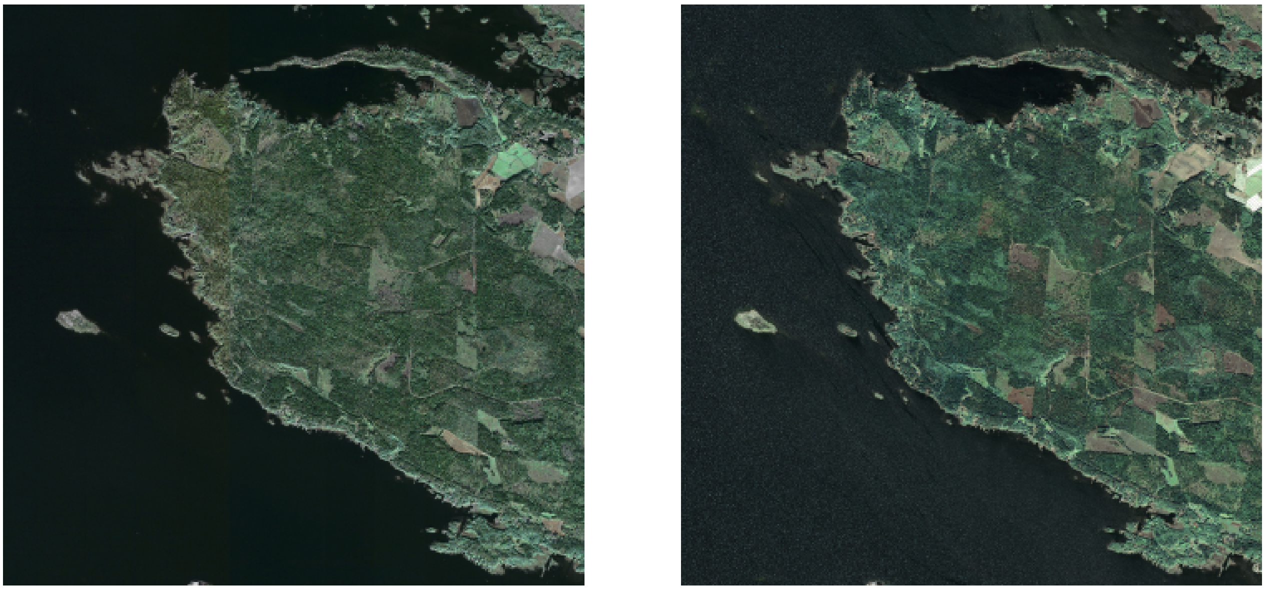
Previous work has been made on databases of relatively small images acquired from medium-resolution sensors. Seidel et al. [5] have experimented a visual-oriented query method on a small test image archive, containing 484 windows extracted from Landsat TM images. Schröder et al. [3] described an intuitive method for semantic labelling of image content suited for query by image content, tested on the same image archive. Schröder [9] and Schröder et al. [4] used a stochastic representation of image content for interactive learning, within a database of about a thousand 1024×1024 Landsat TM scenes – but queries were made by marking training areas. Other work [1, 6, 7] seemed to focus more on managing large databases of full remote sensing scenes. Little was found in the literature about utilizing content-based image retrieval (CBIR) techniques for the purpose of a single scene interpretation, let alone change detection.

We present an original utilization of an existing CBIR system, PicSOM, for the analysis of remote sensing images. In the PicSOM image database browsing system [10], several thousands of images are organized on a Self-Organizing Map (SOM), through the extraction of image descriptors including texture and color features. After the SOM is 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 [11, 12]. The key idea of our study is to artificially create an "image database" from a single satellite image, by dividing it into several thousands of small images, or *imagelets*. PicSOM can then be trained on that virtual "image database", and visually queried for finding objects of interest like man-made structures, or even changes. First results of a PicSOM-based retrieval system applied to very high resolution satellite images are presented in this paper.

2. DATA AND PRE-PROCESSING

2.1. Satellite imagery

Two QuickBird scenes were acquired in the beginning of September 2002 and in mid June 2005, covering a same coastal area in Finland. QuickBird images have four



(a) 2002

(b) 2005

Fig. 1. True-color pan-sharpened QuickBird study scenes

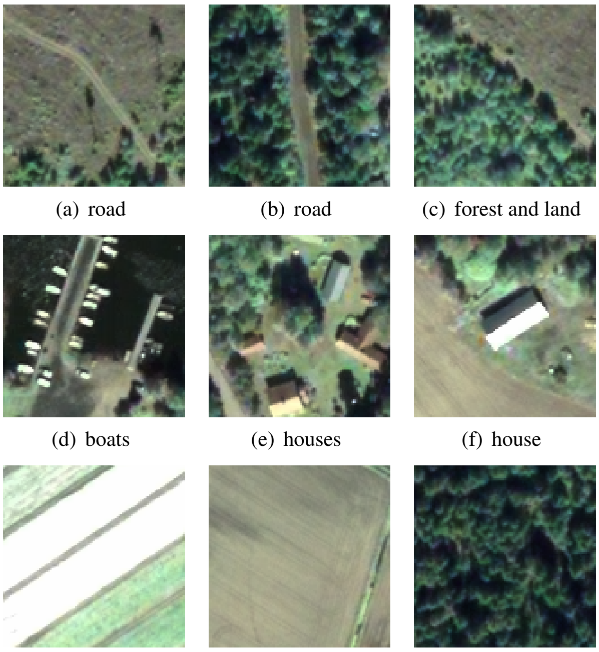
spectral channels with a 2.4 m ground resolution – blue ($450-520$ nm), green ($520-600$ nm), red ($630-690$ nm) and near-infrared NIR ($760-900$ nm) – and a panchromatic channel ($450-900$ nm) with ground resolution of 0.6 m. Both images were remarkably cloud-free, while the sea was quite wavy in the 2005 scene.

2.2. Pre-processing

A study area of size 4×4 km was extracted from both images. The viewing angles were different for the two acquisitions, therefore systematic registration to a common coordinate system would be insufficient for pixel-wise change detection. Scene 2 (June 2005) registration to scene 1 (September 2002) was thus refined by ground control points selection and bicubic transform. The lack of an accurate Digital Elevation Model on the monitored area made orthorectification impossible, thus slight misregistration effects remained.

Because the PicSOM system was originally developed for conventional images (those found e.g. in common web image databases), the dynamic range of QuickBird images had to be reduced from 11 bit to 8 bit. In an attempt to use both spectral and spatial resolution capabilities of the sensor, panchromatic and RGB channels were merged, using ERMapper SFIM Pan Sharpening Wizard. This produced two true-color images with 0.6 m resolution – Fig. 1. Note that the NIR channel was not included in the images used in PicSOM.

PicSOM image retrieval system typically requires several thousands of images in a database, in order to produce relevant indexing. Each pan-sharpened RGB image was then cut into 4900 non-overlapping small im-



(a) road

(b) road

(c) forest and land

(d) boats

(e) houses

(f) house

(g) agricultural field

(h) arable land

(i) forest

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

ages (or *imagelets*), of about 100 × 100 pixels – Fig. 2. Imagelets were named in such a way that it tells their location within the study area and year of acquisition. Data from 2002 and 2005 were kept separated, in two distinct sub-databases of a same database loaded into PicSOM.

2.3. Pixel-based labelling

The 2002 study area was labelled into 7 classes – {*agricultural field, arable land, buildings, clearcuts, forest, roads, water*}. Because automatic classification of very high spatial resolution images is usually challenging, image classification was mostly supervised using Maximum Likelihood algorithm. For small buildings and narrow roads, it was refined by manually selecting the corresponding areas in the image.

Water and forest classes were automatically labelled with the AutoChange software [13], 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 [13]. AutoChange was originally developed for forestry applications, and required to include the NIR band at this stage to perform automatic classification.

Multiple labels were then assigned to each imagelet. The lists of imagelets containing pixels of each class were built and saved as 7 text files, handled by PicSOM. The classification of the 2002 image was not used to help PicSOM recognize objects, only to ease querying or selecting imagelets of interest during the system development and testing.

3. METHODS

The PicSOM system used in this study has originally been developed for content-based image retrieval (CBIR) research [11, 12]. It is based on using the Self-Organizing Map (SOM) [14] as an efficient indexing structure for the images. In PicSOM, multiple SOMs are used in parallel, each created with different low-level visual features. In this paper, we show how this same technique might also be applied in the semi-automated, interactive analysis of satellite images.

3.1. Self-Organizing Maps

The Self-Organizing Map is a neurally-motivated unsupervised learning technique which has been used in many data-analysis tasks. A genuine feature of the Self-Organizing Map is its ability to form a nonlinear mapping of a high-dimensional input space to a typically two-dimensional 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, vectors that reside near each other in the input space are mapped into nearby map units in the output layer. Patterns that are mutually similar in respect to the given feature extraction scheme are thus located near each other on the SOM.

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

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

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

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

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 the small values on later iterations are needed in fine-tuning.

When the SOM has been trained, all the input samples $\mathbf{x}$ are once more mapped to it, each in its best matching unit. Every unit is then assigned as a *visual label* the imagelet whose feature vector was the nearest to the unit's model vector. Fig. 3 and 4 illustrate the most representative imagelets (*visual labels*) on SOMs calculated for two of the feature types introduced in 3.2.

Fig. 3 displays a SOM created from the feature of average RGB values calculated from the imagelets of Fig. 1. We can see how the imagelets whose overall color is dark, i.e. water regions, are mapped to the top-right corner of the map. In the opposite lower-left corner there are imagelets that are the lightest ones, i.e. fields and areas around buildings. The bottom-right corner displays the green forested areas.

In Fig. 4 one can see a different organization of the same imagelet set, this time produced by a texture feature calculated from the imagelets. We can now see that the water regions are mapped in two separate areas due to the different textures of the calm and wavy water surfaces.

3.2. Features

In this study, we used four low-level features automatically extracted from the imagelets. The features were :

xy-**coordinates** This feature used the spatial location of the imagelet in the original scene as a two-dimensional feature. As the images from the two

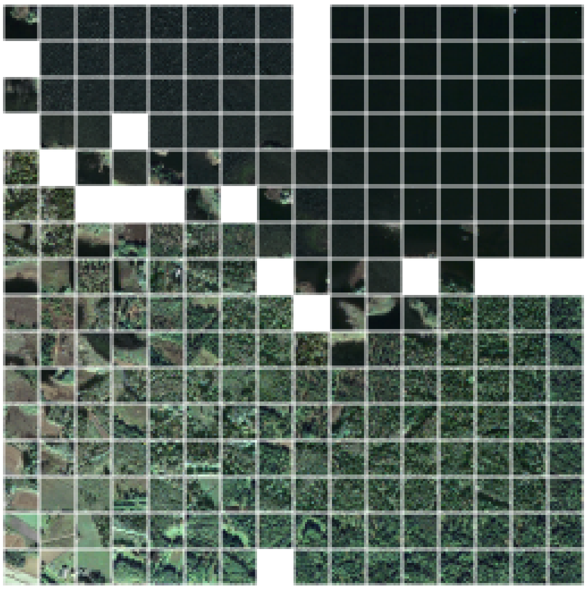


Fig. 3. Organization of the imagelets by their average RGB color on a 16×16 SOM surface.

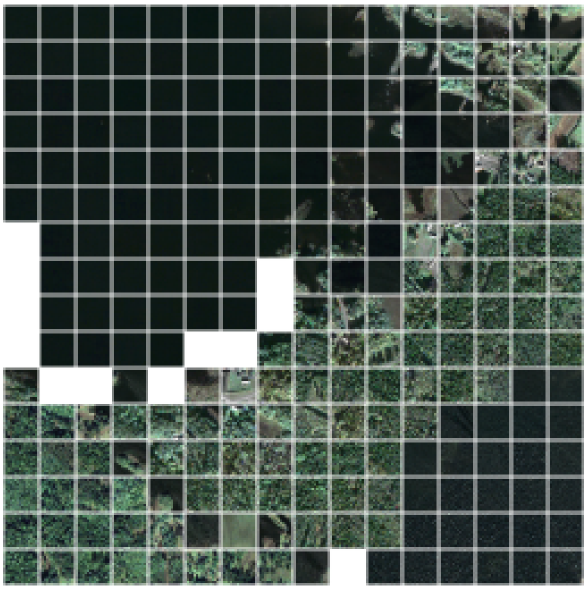


Fig. 4. Organization of the imagelets by their texture content on a 16×16 SOM surface.

years were aligned in preprocessing, the value of this feature was always the same for any imagelet position, regardless of the year and the image content.

average rgb This is a 3-dimensional feature calculated as the average values of the red, green and blue (RGB) channels of the pixels in an imagelet.

color moments For calculating the color moments feature, the RGB color coordinates of the pixels were first transformed to Hue Saturation Value coordinates (HSV). Then the three first moments (mean, variance and skewness) of the HSV values were calculated and used as a 9-dimensional feature.

texture This feature is formed by studying the 8-neighbors of each imagelet pixel. For every 8-neighbor position, a counter is incremented when the illumination on that neighbor pixel is larger than on the center pixel. The final counts are divided by the total number of pixels in the imagelet. The resulting 8-dimensional feature vector then describes local illumination differences, and thus the small-scale texture of the imagelet.

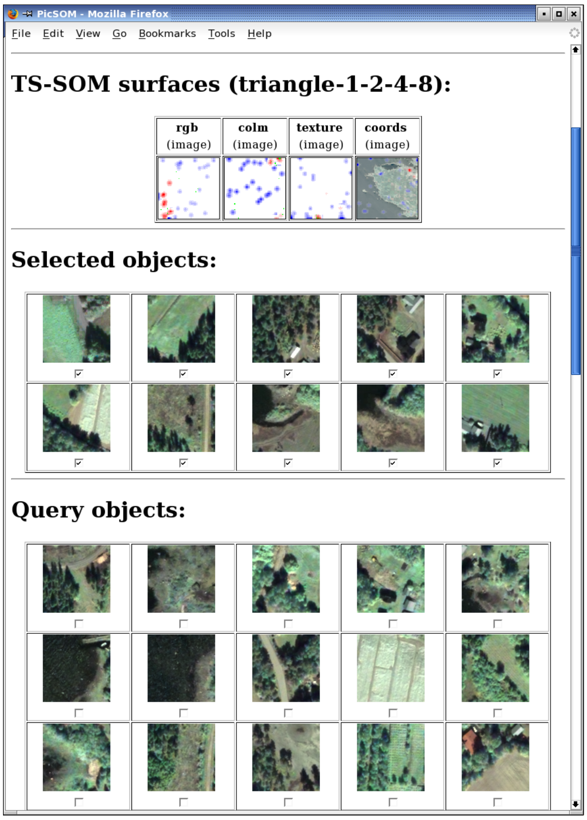
The map sizes were set to 64×64 units for the three visual features SOMs, and 70×70 for the coordinate SOM. Therefore, there were on the average $4900/4096 \approx 1.19$ imagelets mapped in each map unit of the visual SOMs, and exactly one image location on the coordinate map.

3.3. Detection of man-made objects with CBIR

The PicSOM system implements two essential CBIR techniques, query by pictorial examples (QBPE) and relevance feedback. These methods can be used for iterative retrieval of any type of visual content. In iterative QBPE, the system presents some images 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 images and returns them in the next query round.

In our current study, we have used the PicSOM CBIR system to 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 more precisely to more specific semantic targets, such as buildings, roads or clearcuts.

Fig. 5 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. In the top part, the distribution of those imagelets are shown with red colors on the four different SOMs. In the bottom of the interface, some of the new imagelets returned by the system are shown to the user.



TS-SOM surfaces (triangle-1-2-4-8):

Selected objects:

Query objects:

Fig. 5. The web user interface of the PicSOM system in an interactive query for man-made objects.

3.4. Detection of changes

For this study, we devised a method for finding pairs of imagelets, one from the year 2002 and the other from 2005, which differed the most in the sense of some of the extracted features. This means that we did not calculate absolute pixel-wise differences between the imagelets, but defined the change relative to a particular feature extraction scheme. This makes change detection less dependent on small variations in the absolute image coordinates due to inaccurate registration.

Furthermore, we used the SOMs also in the change detection scheme. Some variations in the imagelets are due to various forms of noise and do not correspond to true changes in the land cover. We may assume that the differences caused by noise lead to situations where the best-matching unit for the calculated feature vector remains the same, or is moved to some of the neighboring SOM units. Only the true changes in the imagelet's content would then give rise to such a striking change in the feature vector's value that its projection on the SOM surface is moved to a substantially different location. The substantiality of the change can therefore be measured as the distance between the best matching units (BMUs) of the different years' feature vectors on a same SOM.

Our change detection technique was then as follows. For each of the 4900 imagelet pairs from the years 2002 and 2005, excluding the water areas, we solved the two BMUs on one particular feature's SOM. The Euclidean distance between the BMUs was then calculated and the imagelet pairs were ordered by descending pair-wise BMU distance. A fixed number of imagelet pairs, set to 70 in this study, were then regarded as the locations where the most substantial changes had taken place. This same procedure was repeated for all the three visual features.

4. RESULTS AND DISCUSSION

This research being still in very early stage, quantitative results are not available yet.

4.1. PicSOM for segmentation of satellite images

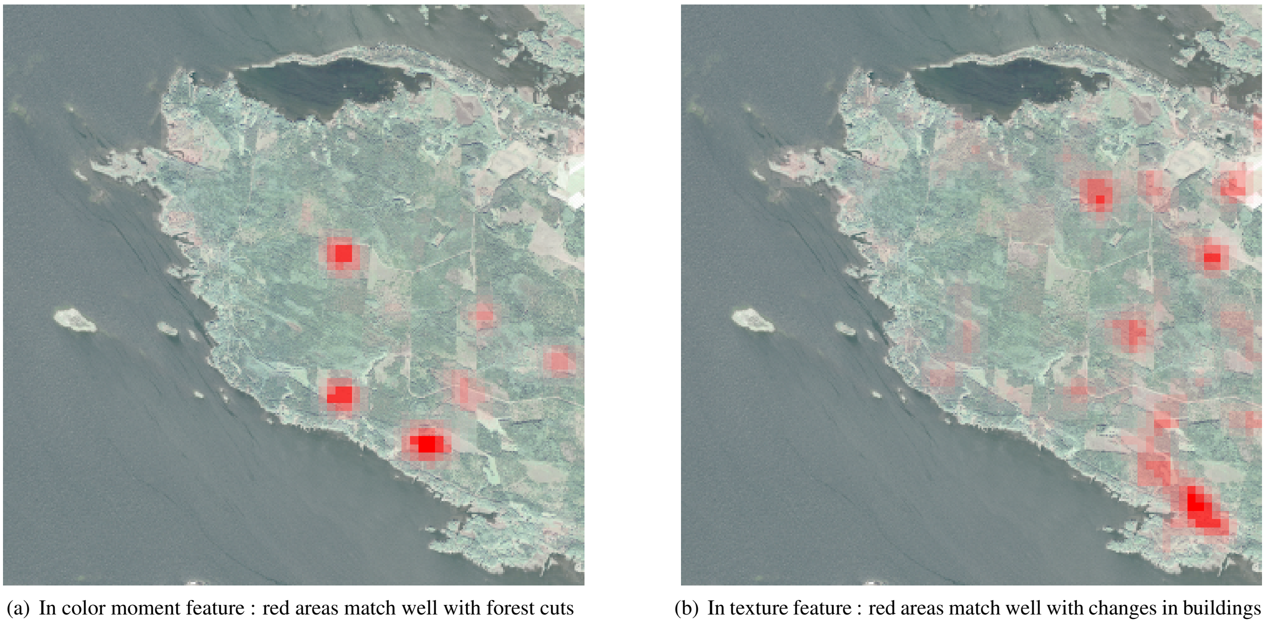
When using PicSOM CBIR system with QuickBird images, the initial random selection of imagelets presented in the user interface didn't include any building. This is probably because there were only small houses in the studied scene, and only a few of them – 137 imagelets out of 4900 – contained buildings in the 2002 study scene.

Already after the first query – made on imagelets containing little or no forest/water – imagelets depicting buildings were retrieved, that could then be selected, refining the detection. Similar results were achieved when visually selecting clearcuts or arable land as target. This already shows a use-case of PicSOM system with remote sensing data, as a supervised, general-purpose and interactive tool for detecting targets within a satellite image, by visual and intuitive querying. A proposed 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. 7.

4.2. PicSOM for change detection

Even with its general feature descriptors, PicSOM system was able to detect zones where changes occurred between 2002 and 2005. Figure 6(a) shows how the system has found the areas where the color moment features had substantial differences. These areas match quite well with those where forest had been cut between 2002–2005. Darker red colors indicate that many adjacent imagelets have changed. On the other hand, Fig. 6(b) displays similar analysis, with the texture feature. These areas correspond this time to changes in built-up areas.

Pixel-based change detection in very high resolution imagery is a challenging task, limited by the requirement of pixel or sub-pixel accuracy registration. A clear advantage of the decomposition in imagelets in the context of change detection, is that it relaxes this constraint – preliminary results suggest that the slight misregistration between the 2002 and 2005 scenes did not affect much the performance of PicSOM for change detection.



(a) In color moment feature : red areas match well with forest cuts

(b) In texture feature : red areas match well with changes in buildings

Fig. 6. Image areas where changes have been detected between 2002–2005

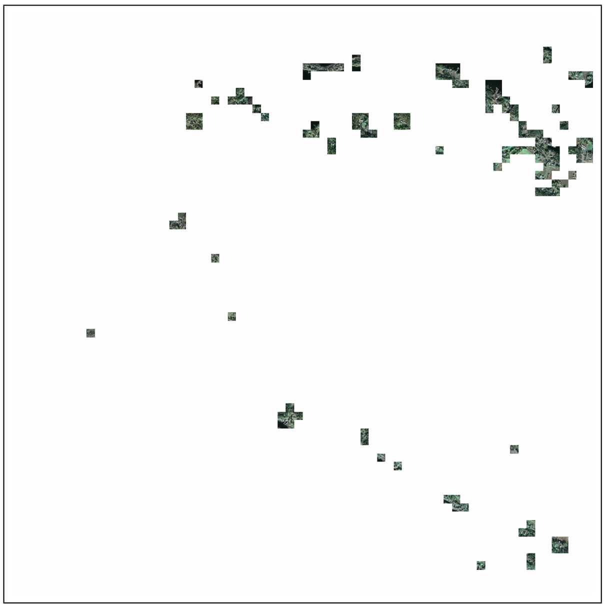


Fig. 7. Imagelets containing buildings (2002 scene), mapped back to original image coordinates.

Seasonal changes were an issue, since many changes were attributed to locations of the study scene containing in fact the same land cover in 2002 and 2005 (often forest) – those were mainly vegetation changes between beginning and end of summer. While this could be interesting for season monitoring applications, it dragged 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 seems to be needed. How this should be done remains open. This could be partly circumvented by calibrating the radiometry of images before loading them into PicSOM.

A way to refine the change detection would be to provide two content targets to 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 by the user of changes of interest – e.g. by selecting imagelets containing forest as earlier target, and buildings as later target, the system would detect newly constructed buildings in forest areas.

An extension of the approach developed here would handle more than 2 satellite images for change detection applications. One could train PicSOM with the imagelets extracted from all images available up to a given time, then query the database for imagelets representative of objects of interest. The system would then return imagelets where significant changes have been detected in the later scene, according to the distance on the SOM between the BMUs of earlier and later imagelets.

4.3. Choosing the size of imagelets

The influence of the imagelet size will have to be studied more deeply. Smaller than the object of interest, it is expected that an imagelet would not restitute all spatial or spectral properties of the target. On the other hand, larger imagelets would contain a too great proportion of perturbating, non-interesting content compared to the spatial extensions of the object 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 contained in the imagelets is representative of the object of interest. Similarly, depending on the scale of expected or interesting changes, the size of imagelets should be smaller or bigger. In an operative system, the expected dimensions of interesting targets or scale of changes could be specified by the user or pre-set, depending on the application, then used to determine an appropriate imagelet size.

The 100×100 pixels imagelets, extracted from QuickBird images used in this study (0.6 m resolution), seemed to provide a trade-off 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 non-overlapping imagelets (namely, generating ”artificial” objects on the borders of imagelets), overlapping imagelets could be used.

4.4. Roadmap for further adapting PicSOM system to remote sensing imagery

The first step will be to handle multi-spectral imagery (typically a number of bands greater than 3) in PicSOM. Using the full dynamic range of the sensors will also be considered. It is expected that more accurate results will be obtained by using all spectral information available when computing features, rather than just 8 bit true-color imagery. As a short-term solution, false-color imagery might also be considered – replacing blue band by NIR, since blue and green bands are usually highly correlated.

Developing sensor-specific feature extractors may lead to improvements in retrieval abilities. Existing features in PicSOM are quite generic image measurements. Widely used measures for remote sensing images, like NDVI, should be included somehow as features. For buildings and man-made structures detection, additional features such as lines or corners may be needed – Rehrauer [15] forecasted that in addition to spectral features, structural features would be suited to content based image retrieval in high resolution satellite images. Selected features should also adapt to various spatial resolutions, depending on the sensor used, and accounting for different target sizes – the importance of scale for satellite image description has also been emphasized in [15].

The relevance of using pan-sharpened images in an operative system has to be investigated. It was used in this preliminary study in order to embed both spatial and spectral information in a same image, easily loaded into PicSOM in its current implementation. If used, the question of preserving spectral information through the pan-sharpening process has to be addressed. The various pan-sharpening techniques do not preserve spectral information equally : some generate more artifacts, some provide suitable results for visual inspection while others are more suited to quantitative evaluation [16]. An appropriate algorithm should be chosen, depending on the level of user interactivity by visual communication needed in an operative system built around PicSOM.

An alternative would be not to fuse panchromatic and multi-spectral channels before processing, but to use them separately. With slight modifications to the PicSOM system, it could be possible to couple spatial features extracted from the panchromatic channel and spectral information from multi-spectral imagery. Preliminary testing on an imagelet database, built from a single panchromatic scene only, showed promising results for building detection, even without any additional spectral information.

The same kind of approach could be applied to radar images, in which case the importance of adapted feature extractors would be even more critical. Still, quick testing of PicSOM on a radar imagelet databases (extracted from an ASAR scene) has been made with satisfying results, that will be published after further research.

5. CONCLUSIONS

We have presented how a content-based image retrieval system, PicSOM, can be used with remote sensing images for tasks like segmentation of man-made structures or clearcuts, as well as change detection. The approach relies on the decomposition of a satellite image into several thousands small images or imagelets, to generate an image database from which the user can query, visually and intuitively. Preliminary results were very encouraging, considering that image features used in the training phase were designed for databases of conventional images. Several improvements of PicSOM are under investigation, that will make it more adequate to the specificities of remote sensing data. The versatility of PicSOM will allow several applications to be embedded in a same system, only to be differentiated by the type of query. Further research will aim at a fully operative and interactive system built around PicSOM, one of the possible applications being long term monitoring of strategic sites.

ACKNOWLEDGMENTS

We are grateful to Ms. Manuela Hirschmugl for the advices she provided on pan-sharpening algorithms. This work was supported by the Academy of Finland in the projects *Neural methods in information retrieval based on automatic content analysis and relevance feedback* and *New information processing principles*, the latter being part of the Finnish Centre of Excellence Programme 2000–2005. Application and system utilization were included in a larger scale project, *ENVIMON*, partly funded by the National Technology Agency of Finland (Tekes).

REFERENCES

[1] Datcu, M., Daschiel, H., Pelizzari, A., Quartulli, M., Galoppo, A., Colapicchioni, A., Pastori, M., Seidel, K., Marchetti, P. G., and D'Elia, S. Information mining in remote sensing image archives - part A: System concepts. *IEEE Trans. on Geoscience and Remote Sensing*, 41:2923–2936, December 2003.

[2] Healey, G. and Jain, A. Retrieving multispectral satellite images using physics-based invariant representation. *IEEE Transactions. on Pattern Analysis and Machine Intelligence*, 18:842–846, August 1996.

[3] Schröder, M., Seidel, K., and Datcu, M. User-oriented content labelling in remote sensing image archives. In *Proceedings of the IEEE International Conference on Geoscience and Remote Sensing*, volume 2, pages 1019–1021, Seattle, USA, July 1998.

[4] Schröder, M., Rehrauer, H., Seidel, K., and Datcu, M. Interactive learning and probabilistic retrieval in remote sensing image archives. *IEEE Trans. on Geoscience and Remote Sensing*, 38:2288–2298, September 2000.

[5] Seidel, K., Schröder, M., Rehrauer, H., and Datcu, M. Query by image content from remote sensing archives. In *IEEE Intern. Geoscience and Remote Sensing Symposium, IGARSS'98*, volume 1, pages 393–396, Seattle, USA, July 1998.

[6] Seidel, K., Mastropietro, R., and Datcu, M. New architectures for remote sensing image archives. In *A Scientific Vision for Sustainable Development, IGARSS'97*, volume 1, pages 616–618, 1997.

[7] Seidel, K. and Datcu, M. Architecture of a new generation of remote sensing ground segments. In *Proceedings of the 19th EARSeL Symposium on Remote Sensing in the 21st Century*, pages 223–228, Valladolid, Spain, June 1999.

[8] Schröder, M. and Dimai, A. Texture information in remote sensing images: A case study. In *Workshop on Texture Analysis, WTA'98*, Freiburg, Germany, 1998.

[9] Schröder, M. Interactive learning in remote sensing image databases. In *IEEE Intern. Geoscience and Remote Sensing Symposium IGARSS'99*, 1999.

[10] PicSOM image browsing system, *http://www.cis.hut.fi/picsom/*. Helsinki University of Technology - Laboratory of Computer and Information Science.

[11] Laaksonen, J., Koskela, M., Laakso, S., and Oja, E. Self-organizing maps as a relevance feedback technique in content-based image retrieval. *Pattern Analysis & Applications*, 4(2+3):140–152, June 2001.

[12] Laaksonen, J., Koskela, M., and Oja, E. PicSOM—Self-organizing image retrieval with MPEG-7 content descriptions. *IEEE Transactions on Neural Networks, Special Issue on Intelligent Multimedia Processing*, 13(4):841–853, July 2002.

[13] Häme, T., Heiler, I., and San Miguel-Ayanz, J. An unsupervised change detection and recognition system for forestry. *International Journal of Remote Sensing*, 19(6):1079–1099, April 1998.

[14] Kohonen, T. *Self-Organizing Maps*, volume 30 of *Springer Series in Information Sciences*. Springer-Verlag, third edition, 2001.

[15] Rehrauer, H. *The Role of Scale for Image Description and Image Query: An Application to Remote Sensing Image, Diss. ETH No. 13622*. PhD thesis, Swiss Federal Institute of Technology, March 2000.

[16] Hirschmugl, M., Gallaun, H., Perko, R., and Schardt, M. "Pansharpening" - methoden für digitale, sehr hoch auflsende fernerkundungsdaten. In *Beiträge zum 17. AGIT Symposium*, Salzburg, Austria, July 06-08 2005.