Author manuscript, published in IEEE Trans. on Geoscience and Remote Sensing, vol. 47, no. 10, pp. 3416-3430, 2009

Change Detection in Optical Aerial Images by a Multi-Layer Conditional Mixed Markov Model

Csaba Benedek and Tamás Szirányi, *Senior Member, IEEE*

***Abstract*— In this paper we propose a probabilistic model for detecting relevant changes in registered aerial image pairs taken with the time differences of several years and in different seasonal conditions. The introduced approach, called the Conditional Mixed Markov model (CXM), is a combination of a mixed Markov model and a conditionally independent random field of signals. The model integrates global intensity statistics with local correlation and contrast features. A global energy optimization process ensures simultaneously optimal local feature selection and smooth, observation-consistent segmentation. Validation is given on real aerial image sets provided by the Hungarian Institute of Geodesy, Cartography and Remote Sensing and Google Earth.**

***Index Terms*— Change detection, aerial images, mixed Markov models**

I. Introduction

AERIAL PHOTO repositories nowadays have a rich and continuously augmenting content. Automatic evaluation of these databases is an important field of research since manual administration is time-consuming, cumbersome and needs uneconomically many human resources. Change detection on photos taken of the same area can be crucial for quick and up-to-date content retrieval. Through the extraction of changes the regions of interest in the images can be decreased drastically in several cases, facilitating applications of e.g. urban development analysis, disaster protection, agricultural monitoring, detection of illegal garbage heaps or wood cuttings. Beside being used as a general preliminary filter, the obtained change map can also provide useful information about size, shape or quantity of the changed areas, which could be applied directly by higher level event detector and object analyser modules [1], [2]. While numerous state-of-the art approaches in remote sensing deal with multispectral [3]–[8] or SAR [9]–[12] imagery, the significance of handling optical photos is also increasing [13], [14]. Here the processing methods should consider that several optical image collections include partially archive data, where the photos are grayscale or contain only poor color information.

This paper focuses on change detection in optical aerial images which were taken with several years time differences partially in different seasons and in different lighting conditions. In this case, simple techniques like thresholding the difference image [15], [16, Sec. IV.] or background modeling [17] cannot be adopted efficiently since the observed pixel levels even in the 'unchanged' image regions may be significantly different. Moreover optical image sensors provide limited information in contrast to [4] where healthy vegetation has been identified by finding peaks in the near infrared wavelength band, or to [10] which uses multi temporal SAR imagery exploiting its insensitivity to atmospheric conditions. We can only assume to have image repositories which contain geometrically corrected and registered [18]–[20] grayscale orthophotos.

The change detection algorithms in the literature follow either the Post-Classification Comparison (PCC) or the direct approach. *PCC* models [7], [13], [14], [21]–[23] segment the input images with different land-cover classes, like arboreous lands, barren lands and artificial structures [21]. Thus changes are obtained indirectly here as regions with different classes in the two image layers. On the other hand *direct* methods [3], [5], [6], [9] derive a similarity-feature map from the input photos [e.g. a difference image (DI)] and then they segment the feature map to separate changed and unchanged areas. As for PCC approaches, besides change detection they classify the observed differences at the same time (e.g. a barren land turns into a built-up area); and the quality of their results can be enhanced by interactive segmentation of the images [22] or exploiting estimated class transition probabilities [21]. However using PCC models we have to fix the clusters a priori in the scenes, and we need to find reliable feature models for each land-cover class with probably various subclasses. In several applications, 'intra-class' transitions - which are ignored by PCC methods - may also be worth for the attention: e.g. inside an urban region, it could be necessary to detect destroyed or re-built buildings, relocated roads. Based on the above remarks, we introduce a *direct* method in this paper which does not use any land-cover class models, and attempts to detect changes which are statistically unusual based on low-level features.

Another important point of view is distinguishing *supervised* [7], [12], [21], [23], [24] and *unsupervised* techniques [5], [6], [8], [9], [25], [26]. Using *unsupervised* approaches is necessary in situations where manually labeled ground truth data cannot be prepared. However many of these methods must use a priori knowledge [27], outlier detection [28] or clustering [29] for separation, which may be difficult if the feature statistics for the different classes is multi-modal and strongly overlapping. Although we can also find unsupervised techniques which do not require any a priori assumptions [6], [8], using them the differences due to atmospheric and light variations may result in artifacts on optical photos [21]. On the other hand if training data is available, it can provide significant additional information for the classification process. Since the photo repositories focused on in this work contain large batches of images from the same year taken with the same quality, camera settings and similar seasonal and illumination conditions, it can be admissible to prepare ground truth for a minor part of the data. Considering

Manuscript received September 24, 2008; revised April 21, 2009. This research was supported by the MUSCLE Shape Modeling E-Team and the Hungarian Scientific Research Fund "Structural information in the space of sensor networks" (OTKA #76159).

The authors are with the Distributed Events Analysis Research Group, Computer and Automation Research Institute, Kende utca 13-17, Budapest, Hungary, H-1111. E-mail: bcsaba@sztaki.hu, sziranyi@sztaki.hu. Cs. Benedek is also with the Ariana Project Team, INRIA Sophia Antipolis, 2004 route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex, France

Digital Object Identifier: 10.1109/TGRS.2009.2022633

Document version of the MTA SZTAKI Publication Repository, http://eprints.sztaki.hu/

Author manuscript, published in IEEE Trans. on Geoscience and Remote Sensing, vol. 47, no. 10, pp. 3416-3430, 2009

these circumstances a *supervised* approach will be introduced here.

The next issue deals with probabilistic feature modeling. In the selected feature space, the different classes can be represented in a parametric [5], [9], [12], [30], [31], semiparametric [6] or nonparametric [3], [7], [21], [30], [32] manner. On one hand nonparametric (e.g. neural networks, kernel density estimation) or semiparametric models may approximate a general density function. However problems of over-learning and incomplete training data provide difficult challenges, which can only partially be handled by model dimensionality reduction, like fixing the number of hidden neurons in a multi-layer perceptron. In these 'black box' techniques, the responsibility of choosing appropriate training and test data is highly increased, both in the supervised and unsupervised cases.

On the other hand parametric models approximate the feature statistics by particular probability density functions (*pdf*), such as Gaussian [5], generalized Gaussian [9] or bivariate gamma distribution [12]. Here the chosen *pdf* may represent additional a priori or experimental information about the feature-models, number of classes, or type of the expected noise. We propose a parametric approach here for the classes *change* and *background* (i.e. unchanged region), where *pdf* selection is validated by experiments.

Comparing the goals, several previous methods deal with purely natural [4] or built-up [14] territories, or they are dedicated to a specific task like detecting urban areas [13], [22], [33] or destructions due to earthquakes [34]. The method in [14] considers the task of change detection as a problem of training-based object recognition, assuming that objects can be detected efficiently by simple features like region entropy, edge points, intensity mean etc. However the latter approach can only be used if the changes can be interpreted at object levels, which needs a precisely restricted environment. As Fig. 1, 10 and 12 show, several photos of our interest contain both built-in and unpopulated regions, including forests, fields and agricultural lands as well, presenting various types of differences.

Similar environmental conditions are expected by the PCA-based model [5]. This method assumes that the 'unimportant' differences are caused by alteration of illumination (like direct and diffuse light, transmittance of the atmosphere) and camera settings (dark current and gain settings of sensors), which influence the observed sensor values in a multiplicative or additive fashion. Accordingly the pixel levels corresponding to the same surface point at two different times are related by a global linear transform which is estimated by an iterative procedure over the image. Other examples for linear intensity transformation can be found in [15, Sec. III.]. However these models do not take into consideration that the scene may 'regularly' alter as well, primarily due to the seasonal vegetation changes. Moreover in agricultural areas which follow crop rotation, the shape and arrangement of the neighboring tracks of a plough-land may be changed significantly. We will show that the regularity of these changes can also be measured in a statistical way, although they may cause significant deviations from the estimated linear approach.

We continue with the formal approach of the problem. Change detection can be considered as an image segmentation task with *change* and *background* classes. Since the seminal work of Geman and Geman [35] Markov Random Fields (MRFs) have been used extensively for various segmentation problems: they can simultaneously ensure the consistency of the class labels with local pixel-level cluster descriptors, and spatial smoothness of the label map through interaction between neighboring pixels. MRFs have also been adopted for extracting changes in remote sensing [5], [6]. As separating complex classes may be inaccurate by a single feature, integration of multiple observations is an increasing challenge. For this purpose, multi-layer Markovian approaches have recently been established [16], [36], [37] showing often significant advantages [37] versus multinomial feature density modeling [38] or decision fusion [39] techniques. In multi-layer models, the different observations are assigned to individual layers. The segmentation of each layer is directly influenced by its corresponding measurement and indirectly by features of the other layers, which interact through inter-layer connections. In this way, data-driven and label-based inferences can be encapsulated in a consistent probabilistic framework providing a robust segmentation approach.

Since context dependent class models can be hardly encoded in conventional MRFs, different schemas have been proposed recently to overcome this limitation. Triplet Markov fields [40] contain an auxiliary latent process which can be used to describe various subclasses of each class in different manners. On the other hand, Mixed Markov models [41] extend MRFs by admitting data-dependent links between the processing nodes, which enables configurable structures in feature integration. Since the later property proved to be crucial in our multiple feature based approach we follow the Mixed Markovian schema. As a key contribution of this paper, we exploit the dynamic connections in a multi-layer framework. Here the different layer-regions are considered or ignored upon local statistical estimation of the reliability of their corresponding features.

In the paper, as an extension of our previous work [42], we propose a robust multi-layer Conditional MiXed Markov model (CXM) model to tackle the change detection problem in remote sensing optical images. Changes are identified through complementary features: global intensity statistics and local correlation. A contrast based selection process is responsible for choosing locally the more reliable feature in the different image regions, while a smooth change map is ensured using local connectivity constraints.

The paper is organized into five parts. In Section II we address feature selection, probabilistic description of the change/background classes and principles of feature integration. In Section III a novel mixed Markovian image segmentation model is introduced for the change detection problem. Section IV deals with experimental evaluation: we give a detailed description of the test datasets and ground generation. Thereafter we explain the manner of parameter estimation and we compare the proposed model to four state-of-the-art approaches qualitatively and quantitatively. Finally concluding remarks are given in Section V.

II. Image model and feature extraction

In this section we focus on the issues of feature selection and modeling. Let G_1 and G_2 be the two registered grayscale images which we wish to compare. G_1 and G_2 have an identical pixel lattice S. The gray values are denoted by $g_1(s)$ and $g_2(s)$ for a pixel $s \in S$ of G_1 and G_2, respectively. Our first task is to extract local features at each $s \in S$ which give us information for classifying s as a *changed* (ch) or *background* (bg) i.e. unchanged surface point. Taking a probabilistic approach, we consider the $\mathrm{ch/bg}$

Document version of the MTA SZTAKI Publication Repository, http://eprints.sztaki.hu/

Author manuscript, published in IEEE Trans. on Geoscience and Remote Sensing, vol. 47, no. 10, pp. 3416-3430, 2009

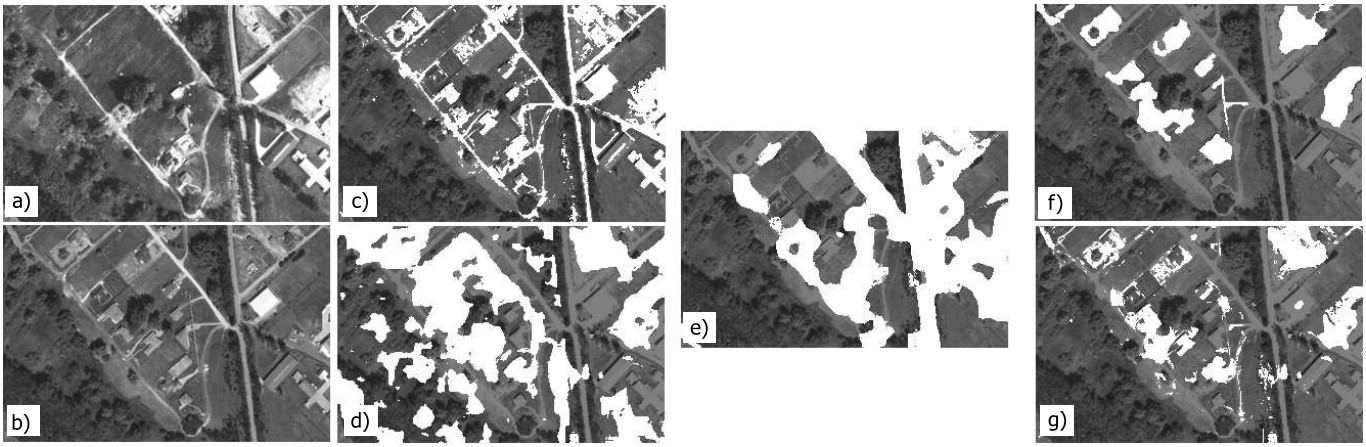


Fig. 1. Feature selection: a) image 1 (G_1), b) image 2 (G_2), c) intensity based change detection ($\phi_g(.)$, changes are marked with white), d) correlation based change detection ($\phi_c(.)$) , e) local variance based segmentation, white if $\phi_\nu(s) = c$, f) ground truth, g) change detection results obtained by per pixel integration of $\phi_g(.)$, $\phi_c(.)$ and $\phi_\nu(.)$ maps

classes as random processes generating the features according to different distributions. Validation of the upcoming class models is experimental: we demonstrate the consecutive modeling steps using a selected image pair shown in Fig. 1(a) and (b). Since the proposed method is supervised, we assume that each *test set* contains a few training images with ground truth. The photo pairs included in the same *test set* are taken from nearby locations, and within a time-layer the illumination conditions and the camera properties/settings are similar. Note that detailed evaluation will be given in Section IV.

A. Segmentation based on global intensity statistics

We start our investigations in the joint intensity domain of the two images. Here, instead of prescribing a global linear transform between $g_1(s)$ and $g_2(s)$ for the background areas [5], we give a multi-modal description of the observed data. Let us consider the 2-D histogram of the $\overline{g}(s) = [g_1(s), g_2(s)]^T$ vectors extracted over the *background* regions of the training images [see Fig 2(a) regarding the image pair of Fig. 1]. Thereafter we approximate this histogram by a mixture of K Gaussian distributions, where K is a parameter of the model. In this way, we measure which intensity values occur often together in the corresponding images. Thus the probability of the $\overline{g}(s)$ observation in the background is calculated as:

$$P\big(\overline{g}(s)\big|\mathrm{bg}\big) = \sum_{i=1}^{K} \kappa_i \cdot \eta\Big(\overline{g}(s), \overline{\mu}_i, \overline{\overline{\Sigma}}_i\Big), \tag{1}$$

where $\eta(.)$ denotes a two dimensional Gaussian density function with $\overline{\mu}_i$ mean vector and $\overline{\overline{\Sigma}}_i$ covariance matrix, while the κ_i terms are positive weighting factors ($\sum_{i=1}^{K} \kappa_i = 1$). Fig 2(b) shows the EM estimate [43] of the density using $K = 5$ mixture components, however only two of them have significant weights.

While the background's intensity model exploits the presence of a few frequently co-occuring gray level pairs in the two images (e.g. the mean color of plough lands or forests), the $\overline{g}(s)$ histogram of the changed regions [see Fig 2(c)] has usually several peaks covering a significant part of the 2-D intensity domain. Moreover we have observed that inside the *test set*s the background histograms of the different image pairs are fairly similar, meanwhile the change statistics shows large variety from one pair to another one. This phenomenon is demonstrated in Table I: first we measured the Bhattacharyya distances [4] between the empirical change distributions of different image pairs from the same set, thereafter a similar experiment was achieved regarding the background as well. The results show that for each comparison the distance of the change histograms is three-four times higher than the distance measured in the background. Therefore we use here a common simplification [15], [17]: expressing that any $\overline{g}(s)$ value may occur in the changed areas with similar probabilities, the 'ch' class is modeled by a uniform density [44] [Fig. 2(d)]:

TABLE I

BHATTACHARYYA DISTANCES BETWEEN THE EMPIRICAL 'CHANGE' (RESP. 'BACKGROUND') $\overline{g}$ STATISTICS OF THREE DIFFERENT IMAGE PAIRS (P1, P2 AND P3) FROM THE SAME SET.

Pairs compared:	P1-P2	P1-P3	P2-P3
Dist. of ch histograms:	0.1309	0.1751	0.0963
Dist. of bg histograms:	0.0404	0.0455	0.0321

$$P\big(\overline{g}(s)\big|\mathrm{ch}\big) = \begin{cases} \frac{1}{(b_1-a_1)\cdot(b_2-a_2)}, & \text{if } \overline{g}(s) \in \Gamma \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where $\overline{g}(s) \in \Gamma$ iff $a_1 \le g_1(s) \le b_1$ and $a_2 \le g_2(s) \le b_2$.

In summary, the $\overline{g}(s)$ feature-based model part for ch/bg separation is described by the following parameters:

$$\Theta_g = \{\kappa_i, \overline{\mu}_i, \overline{\overline{\Sigma}}_i | i = 1 \dots K\} \cup \{a_1, b_1, a_2, b_2\} \tag{3}$$

Next we demonstrate the limitations of using the above intensity based approach. After supervised estimation of the Θ_g distribution parameters, we derive the ϕ_g change map as the pixel by pixel maximum likelihood (ML) estimate, where the label of s is

$$\phi_g(s) = \mathrm{argmax}_{\psi \in \{\mathrm{ch},\mathrm{bg}\}} P\big(\overline{g}(s)\big|\psi\big). \tag{4}$$

Fig. 1(c) shows ϕ_g projected to the original second image. One can observe that the procedure erroneously marks several unaltered regions as changes compared to the proposed ground truth segmentation [Fig. 1(f)]. However the mistakes are mainly limited to highly textured regions (e.g. areas of buildings and roads) since the $\overline{g}(s)$ gray values occurring there are less frequent in the global image statistics. Since these artifacts cannot be

Document version of the MTA SZTAKI Publication Repository, http://eprints.sztaki.hu/

Author manuscript, published in IEEE Trans. on Geoscience and Remote Sensing, vol. 47, no. 10, pp. 3416-3430, 2009

handled in the above approach, we introduce a second feature in the following section.

B. Segmentation based on local correlation

Normalized cross correlation is an efficient similarity measure between image parts, assuming the two regions are identical if and only if the corresponding pixel values are related via an arbitrary (but for the whole region constant) linear transform. Although we have previously refused using a globally constant linear intensity transform for the whole image, the linear dependency often holds locally for small unchanged image parts.

Denote by $N_{s,z} \subset S$ the rectangular neighborhood of s, with a fixed window size $z \times z$ (used $z = 17$). Let $\lambda_i(s)$ and $\nu_i(s)$ be the empirical mean and variance values of the gray levels over the $N_{s,z}$ subimage of G_i, $i \in \{1, 2\}$. Derive $c(s)$ as the normalized cross correlation coefficient between the neighborhoods of s in the two images:[1]

$$c(s) = \frac{\sum_{r \in N_{s,z}} \big(g_1(r) - \lambda_1(s)\big) \cdot \big(g_2(r) - \lambda_2(s)\big)}{z^2 \sqrt{\nu_1(s) \cdot \nu_2(s)}} \tag{5}$$

In Fig 3(a), we plot the histogram of the obtained $c(s)$ values over the changed respectively background regions of the training images. Considering the asymmetry of the empirical distributions, we have found that Beta density approximations [46] are appropriate for the classes [Fig 3(b)]:

$$P(c(s)|\mathrm{ch}) = B\big([c(s)+1]/2, \alpha_{\mathrm{ch}}, \beta_{\mathrm{ch}}\big), \tag{6}$$

where

$$B(x, \alpha, \beta) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, & \text{if} \quad x \in (0,1) \\ 0 & \text{otherwise} \end{cases}$$

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt.$$

Note that scaling $x = [c(s)+1]/2$ is necessary to transform the correlation values into the $[0, 1]$ interval where the Beta density is defined.

Similarly regarding the background class:

$$P\big(c(s)\big|\mathrm{bg}\big) = B\big([c(s)+1]/2, \alpha_{\mathrm{bg}}, \beta_{\mathrm{bg}}\big), \tag{7}$$

As expected, the $c(s)$ features in the changed regions follow approximately a zero-mean distribution, while the background values lie within a higher domain of the $[-1, 1]$ interval. Thus the corresponding parameter-set is:

$$\Theta_c = \{\alpha_{\mathrm{ch}}, \beta_{\mathrm{ch}}, \alpha_{\mathrm{bg}}, \beta_{\mathrm{bg}}\} \tag{8}$$

Next we calculate the ML estimation of the segmentation based on the $c(.)$ descriptor:

$$\phi_c(s) = \mathrm{argmax}_{\psi \in \{\mathrm{ch},\mathrm{bg}\}} P\big(c(s)\big|\psi\big). \tag{9}$$

As the segmentation result in Fig. 1(d) shows, this approach is also weak in itself. However we should observe that $\overline{g}(s)$ and $c(s)$ are efficient complementary features. In low contrasted, homogenous image regions, where the noisy $c(s)$ may be irrelevant, the decision based on $\overline{g}(s)$ seems to be fairly reliable. On the other hand in textured areas one should choose $c(s)$ instead of $\overline{g}(s)$. In the following section, we formulate the *contrast based feature selection* in a probabilistic manner.

[1]Using the integral image trick [45] the calculation of the whole correlation map can be performed efficiently.

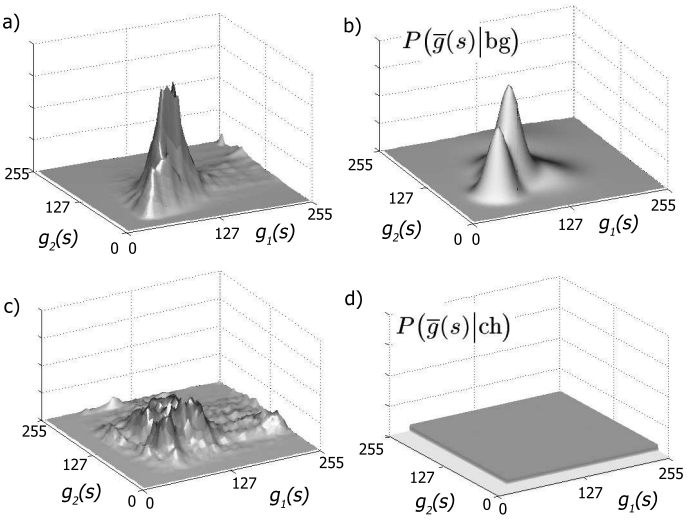


Fig. 2. a) $\overline{g}$-histogram of background pixels b) Mixture of Gaussians approximation of $P(\overline{g}(s)|\mathrm{bg})$ obtained by the EM algorithm [43] c) $\overline{g}$-histogram of the changed pixels d) Uniform density estimation [44] for $P(\overline{g}(s)|\mathrm{ch})$

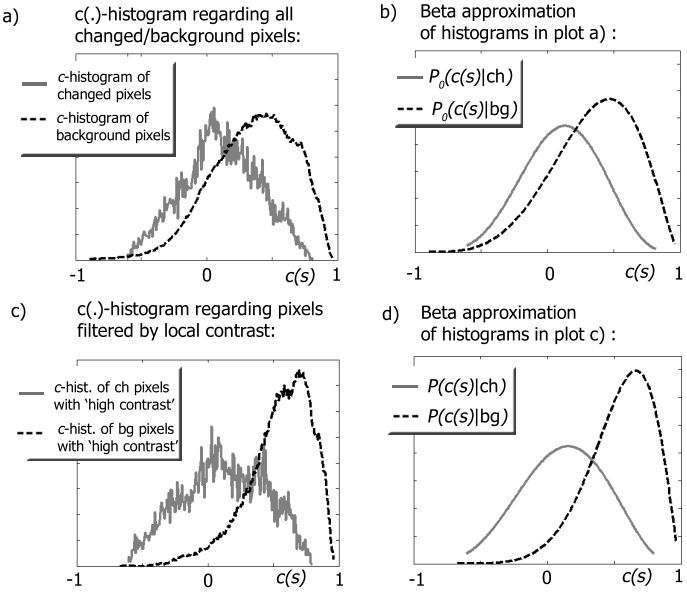


Fig. 3. c histogram and Beta density approximation [46] of the $P(c(s)|\mathrm{ch})$ and $P(c(s)|\mathrm{bg})$ probabilities. (a) and (b): initial estimation; (c) and (d): optimized estimation

C. Contrast based feature selection

Previous observations suggest that considering local contrast, we may estimate the reliability of the segmentation based on the $\overline{g}(s)$ intensity respectively $c(s)$ correlation features at each pixel s. In this section, we give a statistical model for the feature selection.

We will measure the local contrast over image G_i by $\nu_i(s)$ ($i \in \{1, 2\}$), i.e. the variance of the gray levels in the $N_{s,z}$ neighborhood of s (as defined in Section II-B). Let be $\overline{\nu}(s) = [\nu_1(s), \nu_2(s)]^T$. We denote by T the manually generated ground truth mask with $t(s) \in \{\mathrm{ch}, \mathrm{bg}\}$ labels $\forall s \in S$, and δ is the Kronecker-delta.

Next we examine quantitatively the correspondence between the observed $\overline{\nu}(s)$ value and the ML classification performance

Document version of the MTA SZTAKI Publication Repository, http://eprints.sztaki.hu/

Author manuscript, published in IEEE Trans. on Geoscience and Remote Sensing, vol. 47, no. 10, pp. 3416-3430, 2009

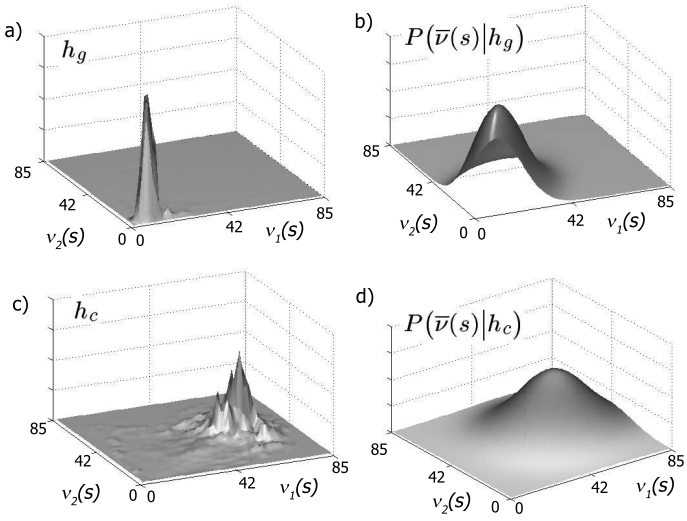


Fig. 4. Illustration of the 2 dimensional h_g and h_c histograms as function of the corresponding $\nu_1(s)$ and $\nu_2(s)$ values

using the $\overline{g}(s)$ and $c(s)$ features, respectively. We particionate the domain of the occurring $\nu_1(s)$ [similarly $\nu_2(s)$] values with L equal bins: $b_1, \ldots, b_L$ (each b_n is a line segment in $\mathbb{R}$.) We say that $\overline{\nu}(s) \in \overline{b}_{m,n}$ if $\nu_1(s) \in b_m$ and $\nu_2(s) \in b_n$ ($\overline{b}_{m,n}$ is a rectangle in $\mathbb{R}^2$). Next we build the following *ratio histogram* h_g, which measures for each $\overline{b}_{m,n}$ bin the ratio of the number of correctly and erroneously classified pixels through $\phi_g(.)$, where the corresponding $\overline{\nu}(s)$ values lie in $\overline{b}_{m,n}$. With $S_{m,n} = \{s | s \in S, \overline{\nu}(s) \in \overline{b}_{m,n}\}$:

$$h_g[m,n] = \frac{\sum_{s \in S_{m,n}} \delta\big(t(s), \phi_g(s)\big)}{\sum_{s \in S_{m,n}} \Big(1 - \delta\big(t(s), \phi_g(s)\big)\Big)} \tag{10}$$

h_c can be defined similarly for the $c(.)$ feature.

We illustrate the h_g and h_c 2-D *ratio histograms* in Fig. 4(a) and 4(c). High peaks of h_g (resp. h_c) indicate domains of $\overline{\nu}(s)$ where the decision based on the $\overline{g}(.)$ [resp. $c(.)$] feature is reliable. After normalization, the histograms can be considered as probability distributions which we approximate again with parametric density functions. In this case, the two classes being modeled are g and c, indicating the $\overline{\nu}(s)$ domains where the $\overline{g}(s)$ respectively $c(s)$ features are more reliable regarding the ch/bg classification of pixel s. In the experiments the two domains proved to be fairly separable with 2-D Gaussian density approximations of the h_g and h_c histograms as it is shown in Fig. 4(b) and 4(d) (see Fig. 4: the histograms are unimodal and only slightly overlapping). Thus we use the following distributions:

$$P\big(\overline{\nu}(s)\big|h_g\big) = \eta\Big(\overline{\nu}(s), \overline{\mu}_g, \overline{\overline{\Sigma}}_g\Big) \tag{11}$$

$$P\big(\overline{\nu}(s)\big|h_c\big) = \eta\Big(\overline{\nu}(s), \overline{\mu}_c, \overline{\overline{\Sigma}}_c\Big) \tag{12}$$

The parameter set assigned to the contrast feature is:

$$\Theta_\nu = \{\overline{\mu}_g, \overline{\overline{\Sigma}}_g, \overline{\mu}_c, \overline{\overline{\Sigma}}_c\} \tag{13}$$

Thereafter we can obtain the ML contrast map [Fig. 1(e)] as:

$$\phi_\nu(s) = \mathrm{argmax}_{\chi \in \{g,c\}} P\big(\overline{\nu}(s)\big|h_\chi\big). \tag{14}$$

The classes' feature models: iterative algorithm for parameter estimation and refinement

Notations: $\Theta_g^{[k]}$, $\Theta_c^{[k]}$ and $\Theta_\nu^{[k]}$ – parameter sets describing the intensity $\overline{g}(s)$, correlation $c(s)$ respectively contrast $\overline{\nu}(s)$ features at the kth iteration. $\phi_g^{[k]}$, $\phi_c^{[k]}$ and $\phi_\nu^{[k]}$ – label maps at the kth iteration.

Steps of the algorithm:

1) Initialization: using the labeled training data, determine $\Theta_g^{[0]}$ and $\Theta_c^{[0]}$ as calculating the ML estimates of the $P\big(\overline{g}(s)\big|\psi\big)$ resp. $P\big(c(s)\big|\psi\big)$ distributions for $\psi \in \{\mathrm{ch}, \mathrm{bg}\}$.
2) Let be $k = 0$.
3) Update the label maps corresponding to the intensity resp. correlation features: calculate $\phi_g^{[k]}$ using (4) with $\Theta_g^{[k]}$ parameter set; similarly $\phi_c^{[k]}$ using (9) and $\Theta_c^{[k]}$.
4) Update the contrast based g/c reliability-histogram: calculate $h_g^{[k]}$ according to (10) using $\phi_g^{[k]}$; similarly $h_c^{[k]}$ based on $\phi_c^{[k]}$.
5) Update the contrast parameters: estimate $\Theta_\nu^{[k]}$ using histograms $h_g^{[k]}$ and $h_c^{[k]}$.
6) Update the contrast based label map: calculate $\phi_\nu^{[k]}$ using (14) with $\Theta_\nu^{[k]}$ parameter set.
7) Update the intensity and correlation parameters: determine $\Theta_g^{[k+1]}$ through ML estimation, considering only the pixels of the training images where $\phi_\nu^{[k]}(s) = g$. Estimate $\Theta_c^{[k+1]}$ through training pixels with $\phi_\nu^{[k]}(s) = c$
8) If the Θ_g, Θ_c and Θ_ν parameters converged, or $k = k_{\max}$ $\rightarrow$ STOP; otherwise $\rightarrow$ $k := k + 1$ and GOTO step 3.

Fig. 5. Iterative algorithm for estimating the Θ_g, Θ_c and Θ_ν parameter sets. Θ_g resp. Θ_c model the $\overline{g}(s)$ intensity resp. $c(s)$ correlation features generated by the *change* and *background* classes, meanwhile Θ_ν describes the $\nu(s)$ contrast observation, on condition that the *intensity* (h_g) resp. *correlation* (h_c) factors are reliable (used $k_{\max} = 5$)

For estimating the final change mask, ϕ_*, the following pixel-by-pixel segmentation process can be taken:

$$\phi_*(s) = \begin{cases} \phi_g(s) & \text{if } \phi_\nu(s) = g \\ \phi_c(s) & \text{if } \phi_\nu(s) = c \end{cases} \tag{15}$$

Moreover the above classification approach enables us to refine the distribution parameters by using purely the ch/bg labeled training data. Observe that in Sections II-A and II-B, the parameters in the Θ_g and Θ_c sets were estimated considering all the pixels of the training images. For example, the background c-histogram in Fig. 3(a) also encapsulates $c(s)$ features extracted from 'low contrasted' areas, where the correlation coefficient proved to be unreliable and irrelevant regarding the final change map. Thus according to (15), we only need to model the $c(s)$ statistics over the 'high contrasted', while $\overline{g}(s)$ distributions over the 'low contrasted' image regions. Therefore we can re-estimate the Θ_g parameters considering only the pixels of the training images with $\phi_\nu(s) = g$; and Θ_c for the training pixels with $\phi_\nu(s) = c$. Note that in this linear parameter estimation schema, there is a mutual dependency between the parameter sets Θ_ν and $\Theta_g \cup \Theta_c$. Thus the parameters can be refined by an iterative algorithm detailed in Fig. 5. In our experiments, the algorithm converged in 3-5 iterations and caused a notable evolution especially regarding the Θ_c parameter set. Fig 3(b) and (d) demonstrate the initial and final Gaussian density functions from Θ_c.

We close this section with an experimental evaluation of the

Document version of the MTA SZTAKI Publication Repository, http://eprints.sztaki.hu/

Author manuscript, published in IEEE Trans. on Geoscience and Remote Sensing, vol. 47, no. 10, pp. 3416-3430, 2009

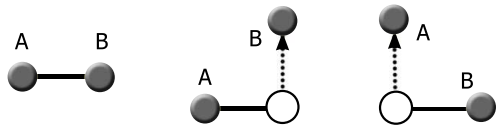


Fig. 6. Three different configurations, where A and B regular nodes may directly interact in mixed Markov models. Empty circles mark address nodes, continuous lines are edges, dotted arrows denote address pointers.

pixel-by-pixel segmentation schema of (15). In Fig. 1(f) and 1(g) images we can compare the proposed ground truth to the ϕ_*-segmentation output. Although the feature integration significantly enhances the results of the individual descriptors [see also Fig. 1(c) and (d)], the combined segmentation in Fig. 1(g) is still quite noisy. Particularly we expect in the segmented image smooth and connected blobs representing the changed regions unlike in Fig. 1(g). To overcome this artifact, neighborhood interaction should be involved in the process besides the pixel level features.

As mentioned before, using MRFs for observation-consistent and smooth image segmentation is well established. However the previously introduced segmentation manner [see (15)] needs a different approach from conventional single-layer MRF models (like the Potts model [47]) for two reasons. First we should integrate efficiently the results of different segmentations, which will be solved by a multi-layer technique similarly to [16], [36]. Secondly, the $\overline{\nu}(s)$ feature plays a particular role: it can locally switch ON and OFF the $\overline{g}(s)$ respectively $c(s)$ features into the integration process. Since in MRFs the interactions between the processing nodes must be static, we should use an extended structure called the mixed Markov model [41], which will be investigated in the next section.

III. A Conditional Mixed Markov image segmentation model

In this section, we introduce a robust segmentation model for the addressed change detection problem. The proposed approach, called the Conditional MiXed Markov Model (CXM), is a combination of a mixed Markov model [41] and a conditionally independent random field of signals. We start with a brief introduction into mixed Markov models and afterwards we present the proposed segmentation approach in detail.

A. Introduction into mixed Markov models

Mixed Markov models [41] extend the modeling capabilities of Markov random fields: beside a priori *static* connections, they enable using observation-dependent *dynamic* links between the processing nodes. A mixed Markov model – similarly to a conventional MRF – is defined over an undirected graph $\mathcal{G} = (Q, E)$, where Q and E denote the sets of nodes and edges, respectively. A label, i.e. a random variable $\omega(q)$, is assigned to each node $q \in Q$ as well, and the node labels over the graph determine a *global labeling*:

$$\underline{\omega} = \{\omega(q) | q \in Q\}$$

However in mixed Markov models two types of nodes are discriminated: J contains *regular nodes* and A is the set of *address nodes* ($Q = J \cup A,\ J \cap A = \emptyset$). Regular nodes $j \in J$ have the same roles as nodes in MRFs: in our application the corresponding variable $\omega(j)$ will encode a segmentation label getting values from the binary $\{\mathrm{ch}, \mathrm{bg}\}$ label set. On the other hand address nodes provide configurable links in the graph by creating pointers to other (regular) nodes. Thus for a given address node $a \in A$, the domain of its 'label' $\omega(a)$ is the set $J \cup \{\mathrm{nil}\}$. In the case of $\omega(a) \neq \mathrm{nil}$, let us denote by $\tilde{\omega}(a)$ the label of the regular node addressed by a:

$$\tilde{\omega}(a) := \omega(\omega(a)). \tag{16}$$

There is no restriction on the graph topology: edges can link any two nodes [41]. The edges define the set of cliques of $\mathcal{G}$, which is denoted here by $\mathcal{C}$.

In a given configuration, two regular nodes may interact directly if they are connected by a static edge or by a chain of a static edge and a dynamic address pointer (see Fig. 6). Particularly with notation for each clique $C \in \mathcal{C}$: $\omega_C = \{\omega(q) | q \in C\}$ and $\omega_C^A = \{\tilde{\omega}(a) | a \in A \cap C, \omega(a) \neq \mathrm{nil}\}$ the a priori probability of a given global labeling $\underline{\omega} = \{\omega(q) | q \in Q\}$ is given by:

$$P(\underline{\omega}) = \alpha \prod_{C \in \mathcal{C}} \exp\Big(- V_C\left(\omega_C, \omega_C^A\right)\Big) \tag{17}$$

where V_C is a $\mathcal{C} \to \mathbb{R}$ *clique potential function*, which has a 'low' value if the labels within the set $\omega_C \cup \omega_C^A$ are semantically consistent, while V_C is 'high' otherwise. Scalar $\alpha = 1/\sum_{\underline{\omega}} P(\underline{\omega})$ is a normalizing constant, which could be calculated over all the possible global labelings. Note that a detailed analysis of analytical and computational properties of mixed Markov models can be found in [41], which confirms the efficiency of the approach in probabilistic inference.

B. A mixed Markovian approach of the change detection problem

In the proposed method, we construct a mixed Markovian probabilistic model on a graph $\mathcal{G}$ whose structure is shown in Fig. 7(c). In Section II, we segmented the images in three different ways, and derived the final result through pixel by pixel label operations using the three segmentations. Therefore we arrange the nodes of $\mathcal{G}$ into four layers: S^g, S^c, S^ν and S^*, where each layer has the same size as the S image lattice. S^g, S^c and S^ν are called the *feature layers*, and S^* is the combined segmentation layer. We assign to each pixel $s \in S$ a unique node in each layer: e.g. s^g is the node corresponding to pixel s on the layer S^g. We denote $s^c \in S^c$, $s^\nu \in S^\nu$ and $s^* \in S^*$ similarly. However, instead of segmenting the layers independently (as in Section II), we obtain the result here by stochastic optimization of a single energy function which encapsulates all constraints of the model: spatial smoothness, optimal local feature selection and observation-consistent classification.

First we introduce a labeling random process, which assigns a label $\omega(q)$ to each q node of $\mathcal{G}$. As usual in mixed models [41], graph edges and address pointers express direct dependencies between the corresponding node labels.

The S^g, S^c, and S^* layers of the model contain *regular nodes*, where the label denotes a possible ch/bg segmentation class:

$$\forall s \in S, i \in \{g, c, *\}:\ \omega(s^i) \in \{\mathrm{ch}, \mathrm{bg}\}$$

For each s, $\omega(s^g)$ resp. $\omega(s^c)$ corresponds to the segmentation directly influenced by the $\overline{g}(s)$ resp. $c(s)$ feature; while the labels at the S^* layer present the final change mask.

On the other hand the S^ν layer is responsible for matching the regions of the final change map S^* to appropriately segmented

Document version of the MTA SZTAKI Publication Repository, http://eprints.sztaki.hu/

Author manuscript, published in IEEE Trans. on Geoscience and Remote Sensing, vol. 47, no. 10, pp. 3416-3430, 2009

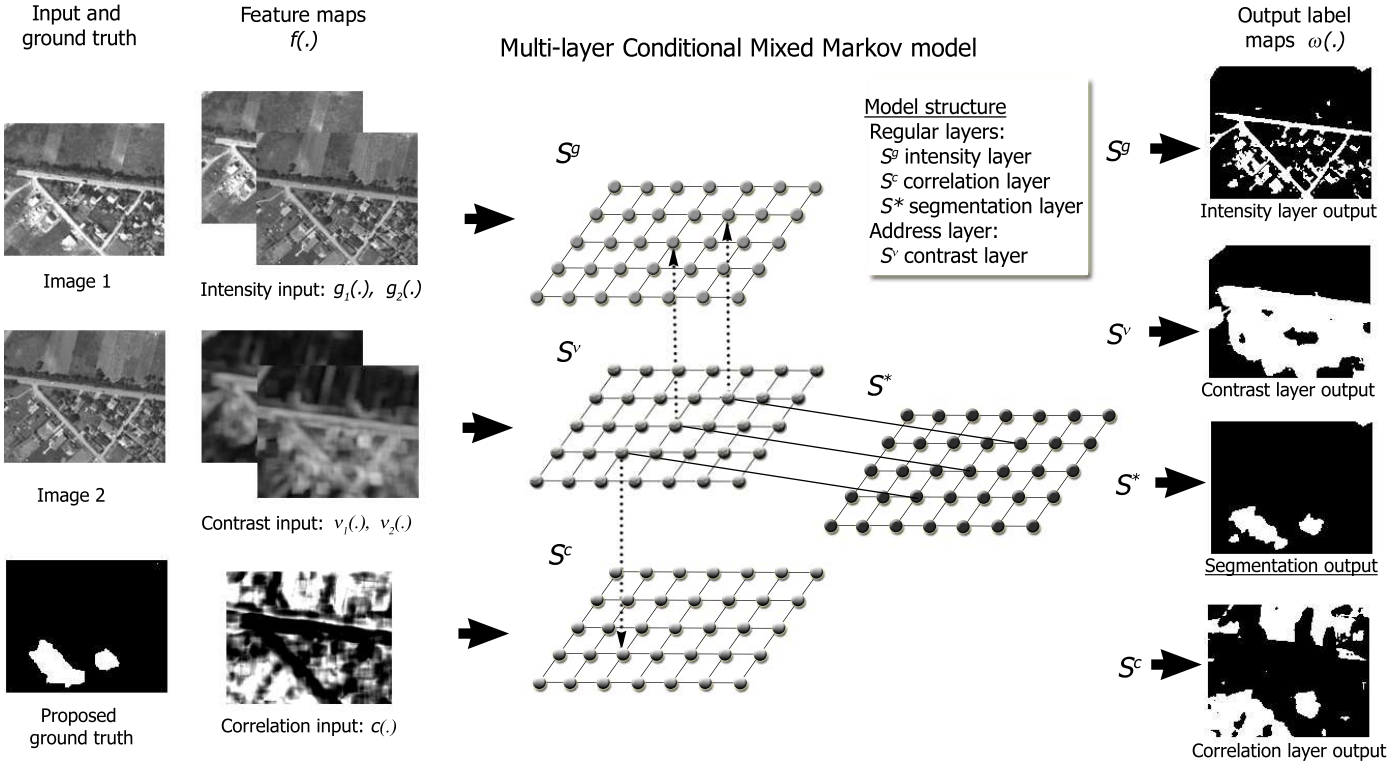


Fig. 7. Structure of the proposed model and overview of the segmentation process. (a) registered input photos and ground truth change mask for validation (b) $\overline{g}(.)$, $\overline{\nu}(.)$ and $c(.)$ feature maps extracted from the input image pair. (c) Structure diagram of the CXM model. (note: the inter-layer connections are only shown regarding three selected pixels) (d) Output label maps of the four layers after MMD optimization. The segmentation result is obtained as the labeling of the S^* layer.

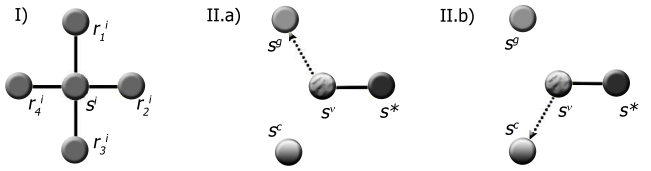


Fig. 8. Demonstration of (I) intra- and (II.a,II.b) inter-layer connections regarding nodes associated to pixel s. Continuous line is an edge of $\mathcal{G}$, dotted arrows denote the two possible a destinations of the address node s^ν. (in I: $i \in \{g, c, \nu, *\}$)

regions either in the S^g or in the S^c layers. Hence S^ν will be an *address layer*, with node-pointer labels $\{\omega(s^\nu)|\forall s \in S\}$.

One can find here semantic analogy between, for example, the $\omega(s^g)$ CXM label and the $\phi_g(s)$ label from Section II: both labels mark the estimated class (*change* or *background*) of pixel s based on the gray level feature. However we emphasize with the different notations that using the Markovian concept, $\omega(s^g)$ depends not only on $\overline{g}(s)$, but also on other model constraints defined over the entire $\mathcal{G}$ graph.

The next issue describes how the model encapsulates the information extracted from the input images. We introduce a $f(.)$ operator which assigns to the nodes of the *feature layers* S^g, S^c and S^ν the corresponding local observations, so that $f(s^g) = \overline{g}(s)$, $f(s^c) = c(s)$ and $f(s^\nu) = \overline{\nu}(s)$, $\forall s \in S$. Denote the global observation process by $\mathcal{F} = \{f(q)|q \in \mathcal{O}\}$, where $\mathcal{O} = S^g \cup S^c \cup S^\nu$.

Our proposed CXM segmentation model follows the Maximum a Posteriori (MAP) approach [16], [35], [36]. The goal is here to find the global labeling $\widehat{\underline{\omega}}$ which maximizes the following conditional probability:

$$\widehat{\underline{\omega}} = \arg\max_{\underline{\omega}\in\Omega} P(\underline{\omega}|\mathcal{F}) = \arg\max_{\underline{\omega}\in\Omega} \Big\{P(\mathcal{F}|\underline{\omega}) \cdot P(\underline{\omega})\Big\} \qquad (18)$$

As for the $P(\mathcal{F}|\underline{\omega})$ probability component of (18), we use the common assumption [17] that the observed local features in the different nodes are conditionally independent given a global labeling. Thus $P(\mathcal{F}|\underline{\omega})$ can be obtained by as a product of *singleton* probability terms assigned to the nodes of the feature layers:

$$P(\mathcal{F}|\underline{\omega}) = \prod_{q\in\mathcal{O}} P\big(f(q)|\omega(q)\big) \qquad (19)$$

In the S^g and S^c layers, we calculate the node by node $P\big(f(q)|\omega(q)\big)$ singletons using the same probability density functions which have already been defined by (1), (2), (6) and (7) in Section II. Thus $\forall s \in S$ and $\psi \in \{\mathrm{ch}, \mathrm{bg}\}$:

$$P\big(f(s^g)|\omega(s^g) = \psi\big) = P\big(\overline{g}(s)|\psi\big)$$
$$P\big(f(s^c)|\omega(s^c) = \psi\big) = P\big(c(s)|\psi\big)$$

Singletons of S^ν will be defined later.

On the other hand using CXM the $P(\underline{\omega})$ prior probability derives from a mixed Markov model, thus it follows (17). Accordingly to calculate $P(\underline{\omega})$, we have to define appropriately the edges (or cliques) of $\mathcal{G}$ and the corresponding V_C clique potential functions. To fulfill the desired constraints, we use in the model

Document version of the MTA SZTAKI Publication Repository, http://eprints.sztaki.hu/

Author manuscript, published in IEEE Trans. on Geoscience and Remote Sensing, vol. 47, no. 10, pp. 3416-3430, 2009

Iterative label optimization in CXM using the Modified Metropolis algorithm (MMD)

1. Pick up randomly an initial global labeling (i.e. configuration) $\underline{\omega} := \underline{\omega}^{[0]}$.
   Set the iteration counter $k := 0$ and the temperature $T := T_0$.
2. Create a list from the nodes in the four-layer model: assign to each node a unique ordinal number between 1 and $|Q|$, applying a sequential scanning strategy for the consecutive layers. Let denote the index of the *actual node* by j, and initialize it as $j := 1$.
3. Choose the j^{th} node from the node-list, and denote it by q. Let be $i \in \{g, c, \nu, *\}$ the index of the layer which contains q, and $s \in S$ the corresponding image pixel for which $q = s^i$.
4. Denote the *actual* label of q in $\underline{\omega}$ by $\omega(q)$. Considering that binary label sets are used in each layer, flip the label of q and denote it by $\pi(q)$:
   - Examples: if $q \in S^g$ and $\omega(q) = \text{ch}$ then $\pi(q) := \text{bg}$; if $q = s^\nu \in S^\nu$ and $\omega(q) = s^g$ then $\pi(q) := s^c$

   Let $\underline{\pi}^q$ be the global configuration which differs from $\underline{\omega}$ only in the label of q.
5. Let us calculate the difference of the field energies corresponding to $\underline{\pi}^q$ and $\underline{\omega}$ configurations: with notation $U(\underline{\omega}) = -\log P(\underline{\omega}|\mathcal{F})$ [see (22)], compute $\Delta U = U(\underline{\pi}^q) - U(\underline{\omega})$ as follows

$$\Delta U := \Delta U_1 + \Delta U_2 + \Delta U_3, \quad \text{where}$$

(a) ΔU_1 is the observation-dependent term, which can be calculated using (1), (2), (6), (7), (11) and (12) from Section II:

$$\Delta U_1 := \begin{cases} -\log P\left(\overline{g}(s)|\pi(s^g)\right) + \log P\left(\overline{g}(s)|\omega(s^g)\right) & \text{if } \ i = d, \\ -\log P\left(c(s)|\pi(s^c)\right) + \log P\left(c(s)|\omega(s^c)\right) & \text{if } \ i = c, \\ -\log P\left(\overline{\nu}(s)|h_{\pi(\chi)}\right) + \log P\left(\overline{\nu}(s)|h_\chi\right) & \text{if } \ i = \nu, \ \text{where } \chi \in \{g, c\} : \omega(s^\nu) = s^\chi, \pi(s^\nu) = s^{\pi(\chi)} \\ 0 & \text{if } \ i = * \end{cases}$$

(b) ΔU_2 is the aggregated difference of the local intra-layer smoothness components based on (20):

$$\Delta U_2 := \sum_{r \in \Phi_s} V_{C_2}\left(\pi(s^i), \omega(r^i)\right) - V_{C_2}\left(\omega(s^i), \omega(r^i)\right).$$

where Φ_s is the first ordered neighborhood of pixel s.
(c) ΔU_3 is related to the differences in the inter-layer potentials [see (21)]:

$$\Delta U_3 := \begin{cases} V_{C_3}\left(\omega(s^*), \pi(s^g)\right) - V_{C_3}\left(\omega(s^*), \omega(s^g)\right) & \text{if } \ i = g \text{ AND } \omega(s^\nu) = s^g \\ V_{C_3}\left(\omega(s^*), \pi(s^c)\right) - V_{C_3}\left(\omega(s^*), \omega(s^c)\right) & \text{if } \ i = c \text{ AND } \omega(s^\nu) = s^c \\ V_{C_3}\left(\omega(s^*), \omega(s^{\pi(\chi)})\right) - V_{C_3}\left(\omega(s^*), \omega(s^\chi)\right) & \text{if } \ i = \nu \text{ (where } \omega(s^\nu) = s^\chi) \\ V_{C_3}\left(\pi(s^*), \tilde{\omega}(s^\nu)\right) - V_{C_3}\left(\omega(s^*), \tilde{\omega}(s^\nu)\right) & \text{if } \ i = * \\ 0 & \text{otherwise} \end{cases}$$

9. Replace the actual configuration by $\underline{\pi}^q$, if ΔU is lower than a positive threshold value which is proportional to the (decreasing) temperature parameter. Otherwise keep the original configuration $\underline{\omega}$. Considering that $\underline{\omega}$ and $\underline{\pi}^q$ differ only in the label of q:

$$\omega(q) := \begin{cases} \pi(q) & \text{if } \ \log \tau \leq -\frac{\Delta U}{T}, \\ \omega(q) & \text{otherwise.} \end{cases} \qquad \text{(MMD label update rule)}$$

   where τ is a constant threshold ($\tau \in (0, 1)$).
10. If $j < |Q|$: $\{j := j + 1$ and GOTO step 3.$\}$
11. STOP if convergence is reached, i.e. the number of the changes between the k^{th} and $k + 1^{\text{th}}$ iteration is lower than a threshold.
12. Increase the iteration counter $k := k + 1$, decrease the temperature $T := T_k$ and jump to the first node $j := 1$, thereafter, GOTO step 3.

Fig. 9. Steps of the Modified Metropolis algorithm (MMD [48]) used for the proposed CXM model. Corresponding notations are given in Sections II and III-A. Following the suggestion of [48] we used in the tests $\tau = 0.3$, $T_0 = 4$, and an exponential cooling strategy: $T_{k+1} = 0.96 \cdot T_k$

two types of cliques representing intra- and inter-layer interactions (see Fig. 8).

For the sake of obtaining smooth segmentations, we put connections within each layer among node pairs corresponding to neighboring[2] pixels on the S image lattice. Denote the set of the resulting *intra-layer* cliques by $\mathcal{C}_2$. The prescribed potential function of a clique in $\mathcal{C}_2$ penalizes neighboring nodes having different labels. Assuming r and s to be neighboring pixels on S, the potential of the doubleton clique $C_2 = \{r^i, s^i\} \in \mathcal{C}_2$ for each $i \in \{g, c, \nu, *\}$ is calculated as:

$$V_{C_2}\left(\omega(s^i), \omega(r^i)\right) = \begin{cases} -\varphi^i & \text{if } \ \omega(s^i) = \omega(r^i) \\ +\varphi^i & \text{if } \ \omega(s^i) \neq \omega(r^i) \end{cases} \qquad (20)$$

with a constant $\varphi^i > 0$.

[2]We use first order neighborhoods in S, where each pixel has 4 neighbors.

Now let us continue with the description of the inter-layer interactions. Based on previous investigations [see (15)], $\omega(s^*)$ should mostly be equal either to $\omega(s^g)$ or to $\omega(s^c)$, depending on the 'vote' of the $\nu(s)$ feature. Hence we put an edge among s^* and s^ν as well as we prescribe that address node s^ν should point either to s^g or to s^c:

$$\forall s \in S: \ \omega(s^\nu) \in \{s^g, s^c\}$$

The directions of the address pointers are influenced by the singletons of S^ν [from (19)] where we use the distributions defined by (11) and (12):

$$P\big(f(s^\nu)|\omega(s^\nu) = s^\chi\big) = P\big(\overline{\nu}(s)|h_\chi\big), \quad \chi \in \{g, c\}$$

Finally we get the potential function of the inter-layer clique $C_3 =$

Document version of the MTA SZTAKI Publication Repository, http://eprints.sztaki.hu/

Author manuscript, published in IEEE Trans. on Geoscience and Remote Sensing, vol. 47, no. 10, pp. 3416-3430, 2009

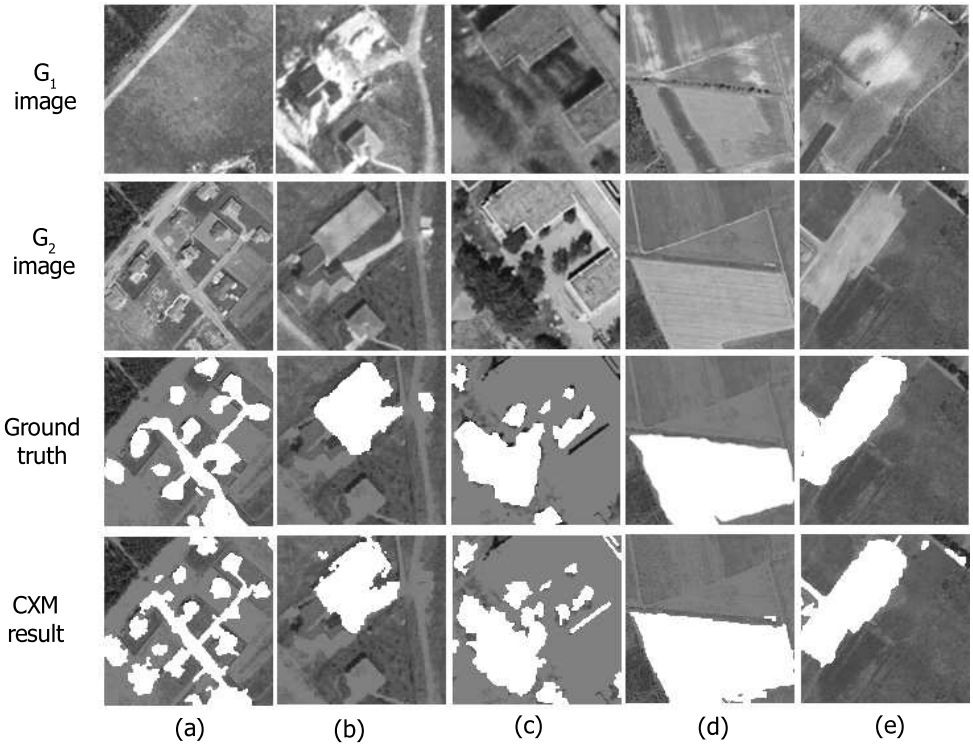


Fig. 10. Change prototypes considered for ground truth generation (a) new built-up regions (b) building operations (c) planting of trees (d) fresh plough-land (e) groundwork before building over

$\{s^*, s^\nu\}$ as

$$V_{C_3}\Big(\omega(s^*), \tilde{\omega}(s^\nu)\Big) = \begin{cases} -\rho & \text{if } \omega(s^*) = \tilde{\omega}(s^\nu) \\ +\rho & \text{otherwise} \end{cases} \tag{21}$$

where $\rho > 0$, and using (16): $\tilde{\omega}(s^\nu) = \omega\big(\omega(s^\nu)\big)$.

According to (18), the observation-dependent term (19) and the prior potential functions (20) and (21) substituted into (17) determine $[P(\mathcal{F}|\underline{\omega}) \cdot P(\underline{\omega})] \propto P(\underline{\omega}|\mathcal{F})$ for an arbitrary global labeling. Thus taking the negative logarithm of the probabilities the optimal $\widehat{\underline{\omega}}$ can be calculated as:

$$\widehat{\underline{\omega}} = \arg\min_{\underline{\omega}\in\Omega} \Bigg\{ \sum_{s\in S} -\log P\big(\overline{g}(s)|\omega(s^g)\big) + \\ + \sum_{s\in S} -\log P\big(c(s)|\omega(s^c)\big) + \sum_{s\in S} -\log P\big(\overline{\nu}(s)|\omega(s^\nu)\big) + \\ + \sum_{i;\{s,r\}\in\mathcal{C}_2} V_{C_2}\big(\omega(s^i), \omega(r^i)\big) + \sum_{s\in S} V_{C_3}\big(\omega(s^*), \tilde{\omega}(s^\nu)\big) \Bigg\} \tag{22}$$

where $i \in \{g, c, \nu, *\}$ and Ω denotes the set of all the possible global labelings.

The energy term of (22) can be minimized by conventional iterative techniques, like ICM [49] or simulated annealing [35]. For choosing a good compromise between the quality factor and processing speed, we adapted to our CXM model the deterministic Modified Metropolis relaxation algorithm [48], which is detailed in Fig. 9. Accordingly the four layers of the model are optimized simultaneously, and their interactions develop the final segmentation, which is taken at the end as the labeling of the S^* layer. Note that due to the fully modular structure the introduced model could be completed straightforwardly with additional sensor information (e.g. color or infrared sensors) or task-specific features depending on availability.

IV. Experiments

A. Test databases and ground truth generation

For evaluation we used *three* sets of optical aerial image pairs provided by the Hungarian Institute of Geodesy Cartography & Remote Sensing (FÖMI) and Google Earth. Data set SZADA contains images taken by FÖMI in 2000 and in 2005, respectively. This test set consists of *seven* - also manually evaluated - photo pairs, covering in aggregate 9.5km^2 area at *1.5m/pixel* resolution (the size of each image in the test set is 952×640 pixels). One image pair has been used here for training and the remaining six ones for validation. The second test set called TISZADOB includes *five* photo pairs from 2000 resp. 2007 (6.8km^2) with similar size and quality parameters to SZADA. Finally, in the test ARCHIVE, we compared an aerial image taken by FÖMI in 1984 to a corresponding Google Earth photo from around 2007. The latter case is highly challenging, since the photo from 1984 has a poor quality, and several major differences appear due to the 23 years time difference between the two shots.

In addition we have performed a few experiments with high resolution (*0.5m/pixel*) images as well [Fig. 10(b) and (c)]. In those cases, already very fine differences caused change alarms, such as planting single trees [Fig. 10(c)], unlike in photos with 1.5m resolution (Fig. 12).

Ground truth masks have been generated manually for each image pair of the training and test sets. The following changes have been considered (see Fig. 10): new built-up regions, building operations, planting forests or individual trees (the latter only

Document version of the MTA SZTAKI Publication Repository, http://eprints.sztaki.hu/

Author manuscript, published in IEEE Trans. on Geoscience and Remote Sensing, vol. 47, no. 10, pp. 3416-3430, 2009

at high resolution), fresh plough-lands and groundworks before building over.

B. Parameter settings

The introduced CXM segmentation model has the following parameters:

- Preliminary parameters of feature calculation: $\Lambda = \{K, z\}$
- Parameters of the probability density functions introduced by (3), (8) and (13): $\Theta = \Theta_g \cup \Theta_c \cup \Theta_\nu$
- Parameters of the intra- and inter-layer potential functions: $\Phi = \left\{\rho, \varphi^i : i \in \{g, c, *, \nu\}\right\}$

As indicated in the introduction, this paper deals with a *supervised* approach of change detection. The first step is determining the Λ preliminary parameters. K, which is the number of the Gaussian mixture components in the background's intensity model, was set by trial and error. We considered the distribution sequence $P^K(\overline{g}(s)|\mathrm{bg})$ for $K = 1, 2, \ldots$ where P^K is obtained by EM estimation [43] from the training data using K mixture components. Thereafter we measured the Bhattacharyya distances [4] between the empirical histogram and the approximated distributions. The results are shown in Fig. 11 for the SZADA training set: the minimal error has been observed at $K = 5$ which is used in the following.

On the other hand the size of the correlation window z particularly depends on the image resolution and textureness. Since in the considered photos the buildings were the principal sources of texture, we have chosen a correlation window which narrowly covers an average house (for the three test sets we used $z = 17$). During the tests we found this choice optimal: with significantly larger windows ($z > 30$) some individual building changes have been erroneously ignored, while small rectangles ($z < 5$) reported many false changes.

After fixing Λ, the Θ parameters can be obtained automatically from the training image pairs using conventional estimators [43], [44], [46] embedded in the iterative framework of Fig. 5.

While Θ parameters strongly depend on the input image data, factors in Φ are largely independent of it. Experimental evidence suggests that the model is not sensitive to a particular setting of Φ within a wide range, which can be estimated a priori. Although one can also find a few automatic estimation methods for similar smoothing terms [50], Φ could rather be considered as a hyper-parameter-set. We used $\rho = \varphi^i = 1$, $i \in \{g, c, *, \nu\}$ in the tests.

C. Evaluation

The aim of this section is to compare quantitatively and qualitatively the proposed approach to results in the literature. An overview on related state-of-the-art methods has been given in Section I. As concluded there, the different approaches use significantly different assumptions, for example, they are either supervised or unsupervised. From another aspect, the methods may have different goals as well, such as change detection (direct methods) or joint image segmentation (PCC). These methodological differences must be considered also in comparative evaluation. Since our proposed method is supervised, for the sake of fair validation we utilize the same training data for our CXM model and regarding the selected reference methods as well. Therefore we implemented supervised modifications of three previous *unsupervised* methods [3], [5], [6] for evaluation.

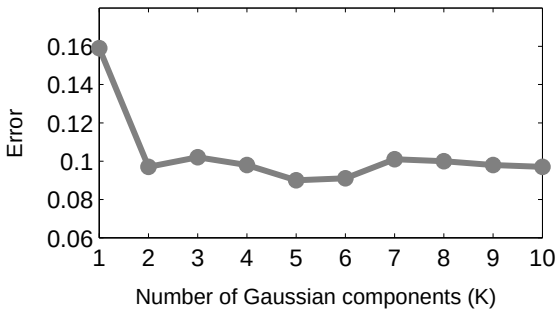


Fig. 11. Error of the mixture of Gaussians approximation of the joint intensity statistics as a function of the number of mixture components (K). The error is characterized by the Bhattacharyya distance between the empirical $\overline{g}$ background histogram and the estimated density.

On the other hand in PCC models (e.g. [7], [13], [21]) the ground truth used for training and evaluation marks the land-cover classes in each frame, while changes in our proposed approach could not be interpreted as such class transition occurrences. For the above reason, we only investigate direct methods in this section.

Considering the above remarks, we compared our method to four previous solutions, which will be introduced briefly in the following. For easier notation, we mark the pixels of the difference image (DI) by $d(s) = g_1(s) - g_2(s)$.

1) PCA: Implementation of the model [5] with supervised training. The observed $\overline{g}(s) = [g_1(s), g_2(s)]$ joint gray level vectors are projected into the space of the principal components estimated over the background training regions. Thereafter the feature value at pixel s is quantified by the magnitude of the second principal component normalized by local contrast. Finally the change-background segmentation of the feature map is obtained by a Potts-like [47] MRF model.

2) Hopfield: Following [3], a Hopfield-type neural network is constructed, which is initialized by a DI-based pixel-by-pixel classification process. The final change mask is obtained by an iterative procedure, which (sub-)minimizes the global energy of the network, while inter-node interactions are responsible for getting smooth change and background regions.

3) Parzen: A nonparametric supervised approach to segment the DI. The $P(d(s)|\omega(s) = \mathrm{bg})$ and $P(d(s)|\omega(s) = \mathrm{ch})$ conditional *pdf*-s are approximated by Parzen kernel density estimators [51] and a MRF model [6] generates the smooth change map. Apart from the hereby adopted supervised training, this solution follows the method [6].

4) MLP: Unlike the previous *Parzen* approach, this method approximates the $P(\omega|d)$ probabilities instead of $P(d|\omega)$. A MRF segmentation of the DI is investigated again, but here the $P(\mathrm{ch}|d(s))$ and $P(\mathrm{bg}|d(s))$ distributions are estimated by a multilayer perceptron trained with backpropagation from ground truth data. Note that this approach has been applied in a similar manner by *PCC* methods [7], [21] to segment the different image layers, which makes the above *'direct'* MLP technique also worth for the comparison.

During the numerical tests, we use the same metric as [3], [6]: we compare the segmentation results provided by the different techniques to the ground truth (GT), and measure the numbers of false alarms (unchanged pixels which were detected as changes),

Document version of the MTA SZTAKI Publication Repository, http://eprints.sztaki.hu/

Author manuscript, published in IEEE Trans. on Geoscience and Remote Sensing, vol. 47, no. 10, pp. 3416-3430, 2009

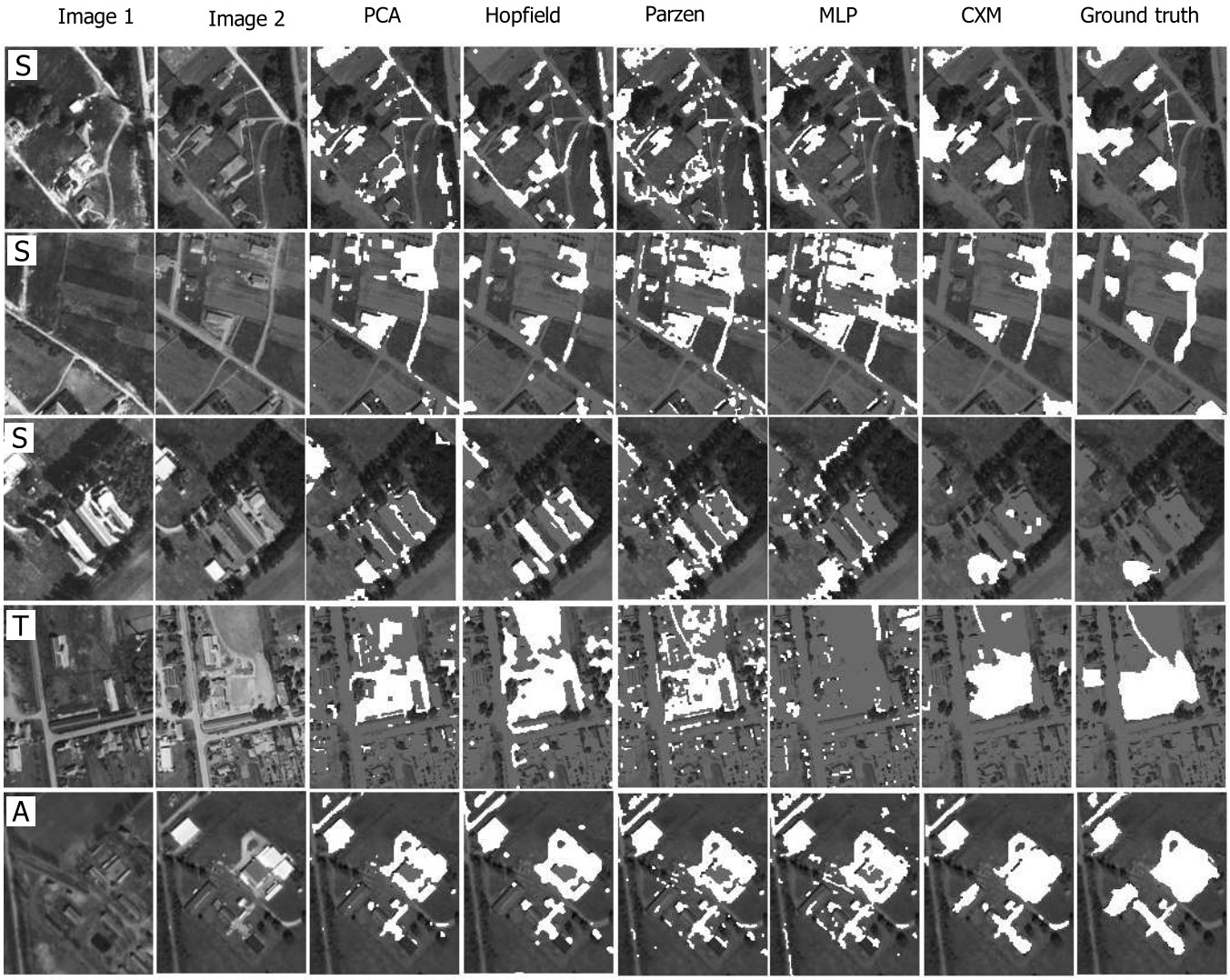


Fig. 12. Qualitative comparison of the change detection results with the different test methods defined in Sec. IV-C and the proposed CXM model. White regions mark the detected/ground truth changes. Each image part covers a $45\mathrm{m}^2$ sized area (128×156 pixels at 1.5 resolution). Capital letters at the beginning of the rows refer to the corresponding datasets: SZADA (S), TISZADOB (T) and ARCHIVE (A).

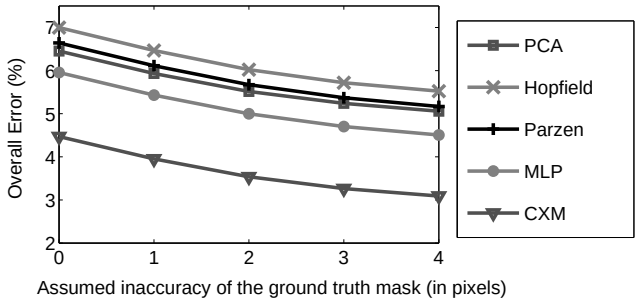


Fig. 13. Overall errors on the SZADA test set according to the i-evaluation metrics as a function of i.

missed alarms (erroneously ignored changed pixels) and overall error (sum of the previous two quantities).

A practical problem of GT-based quantitative evaluation is that the accuracy of the manually generated reference masks may impact the results. To achieve a robust comparison between the different methods we investigated the error rates at different GT-quality assumptions. Let us assume that in the GT masks the borders of the change regions may be inaccurate up to $\pm i$ pixels. Then we exclude from the false and missed alarm candidates the pixels which are at most at i distance from the change blobs' borders. As i increases the metrics become more permissive, thus the number of false/missed alarms obviously decrease monotonically. In Fig. 13 we show the overall errors on the SZADA test set according to this 'i-evaluation metrics' for $i = 0, 1, \ldots, 4$. We can observe here that the order of methods is the same for all i-s, and the differences are similar, therefore we use only the case $i = 0$ (the most strict assumption) in the upcoming comparisons.

The further evaluation rates measured on thehree test sets are given in Fig. 14 in percent of the number of processed image pixels. As this figure shows, the overall error of the proposed CXM model is below the error of the reference methods by about $2-5$ percent. Note that the generally weaker results in the ARCHIVE tests are primarily caused by the lower image quality.

For the sake of visual demonstration, we show the comparative

Document version of the MTA SZTAKI Publication Repository, http://eprints.sztaki.hu/

Author manuscript, published in IEEE Trans. on Geoscience and Remote Sensing, vol. 47, no. 10, pp. 3416-3430, 2009

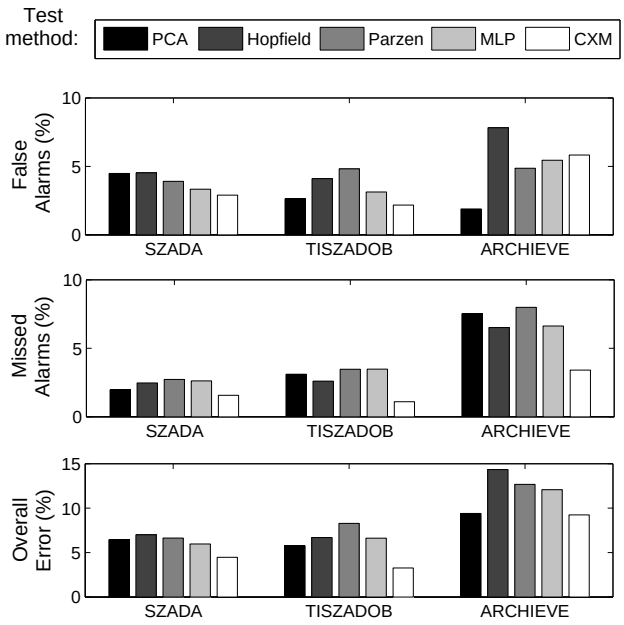


Fig. 14. Quantitative comparison of the different test methods regarding the three test sets: SZADA, TISZADOB and ARCHIVE. False alarm, missed alarm and overall error rates are given in percent of the checked pixels.

change detection results of some relevant image parts in Fig. 12. Each image covers here a $45\mathrm{m}^2$ area. We can observe that the CXM model produces smooth and more accurate change regions in the selected areas, compared to the reference techniques.

D. Discussion of the results

The experimental validation on real image sets (Section IV-C) demonstrates that the introduced CXM model offers measurable advantages versus earlier techniques for the change detection problem. The main sources of the improvements are the joint usage of *multiple features* and the proposed probabilistic *feature integration* schema.

As for *feature selection*, artifacts of the Hopfield, Parzen respectively MLP methods show that in the considered optical photos the changed and background regions cannot be efficiently separated via pure pixel level intensity differences due to the large overlap between the classes in the $d(s)$ descriptor's domain. The PCA approach uses also a scalar feature: it constructs a linear physical model for the irrelevant global illumination changes, and the classification is performed based on the second principal component of the joint intensity vector. However the experiments confirm that in optical images, several regular changes (like vegetation changes) do not fit the linear model, meanwhile noise and sensor saturation causes additional false alarms. (See Fig. 12, 3^{rd} row's PCA image: a false building change appears in the left.) Moreover both of the above pixel level descriptors are less efficient in areas containing dense 3-D structures where various illumination artifacts may appear including mirroring effects and shadows. As a main difference, our method uses different features in different scene parts exploiting that while the intensity value is a relevant attribute in homogenous regions; it is rather a source of noise inside the high contrasted areas where it is preferred to apply texture comparison with block correlation. Choosing locally the more reliable feature leads us to obtain more accurate change regions both in homogenous and highly textured territories (Fig. 10 and 12).

The next key issue has been the construction of the *segmentation model*. Since pixel by pixel classification proved to be quite noisy [see Fig. 1(g) and Fig. 15(d)] we considered Markovian neighborhood interaction for obtaining a smoothed change map [35]. However the introduced label-based feature integration approach required a special model structure which oversteps the single-layer techniques used in the reference methods. We also emphasize that in the proposed multi-layer model the segmentations of the different layers are performed in a synchronised way through the inter-layer interactions. To illustrate the advantages of this synchronisation we compared in Fig. 15 the CXM result to an ensemble of independent MRFs [image (e)]. The later model applies first three (conventional) MRF-segmentations for the $\overline{g}(s)$, $c(s)$ and $\overline{\nu}(s)$ feature maps which step is followed by pixel level label fusion based on (15). Although most erroneous discontinuities can be removed from the change mask by the ensemble [compare Fig. 15(d) and (e)], the asynchronous smoothing of the different label maps causes several artifacts [22], [37] compared to the CXM result [image (f)].

We tested the proposed method using real world aerial images with various content, quality and resolution (Fig. 10). Considering the results our contributions proved to be the most efficient in processing 1.5m/pixel resolution photos from sparsely populated areas (Fig. 12). Those images contain equally low contrasted natural regions and highly textured built-up territories, thus the improvements of choosing from multiple descriptors become significant versus the single feature based techniques. Note that at a resolution of several meters [6] fine land texture is usually not observable reducing the role of the correlation descriptor. On the other hand at very fine resolution ($\leq$0.5m) misregistration and parallax decrease the efficiency of the approach which can be partially controlled by choosing large regions for $c(s)$ and $\overline{\nu}(s)$ calculation (z parameter in Section IV-C) or applying a moving correlation window [16]. However dealing here with dense urban areas large occlusions and moving objects should be handled at a higher processing level. False alarms may also appear due to fine rarefaction of vegetation which changes the texture and color of the area at the same time in unusual way (Fig. 12, 1^{st} row).

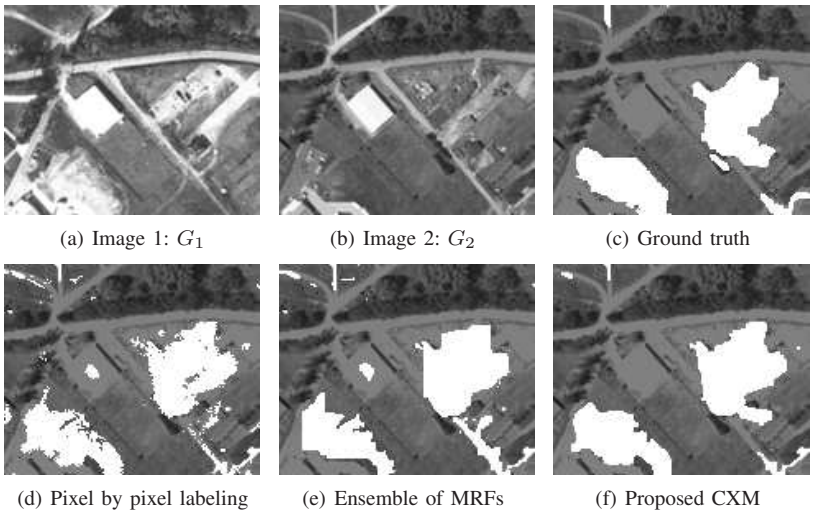
As the image quality becomes lower, the method shows graceful degradation like in Fig. 14 regarding the ARCHIVE test set. Here additional sources of artifacts appear like blurred textures and blotches due to film errors which could be filtered as in [52].

V. CONCLUSIONS

This paper has addressed the detection of statistically unusual changes in optical aerial image pairs taken with significant time differences. A novel Conditional Mixed Markov model has been proposed, which integrates the robustness of MRF-based segmentation techniques [35], the modularity of multi-layer approaches [37] and semantic flexibility of mixed Markov models [41]. The introduced method utilizes information from three different observations: global intensity statistics, local correlation and contrast. The performance of the method has been validated using real-world aerial images, and its superiority versus four earlier reference methods has been shown quantitatively and qualitatively.

Document version of the MTA SZTAKI Publication Repository, http://eprints.sztaki.hu/

Author manuscript, published in IEEE Trans. on Geoscience and Remote Sensing, vol. 47, no. 10, pp. 3416-3430, 2009



(a) Image 1: G_1 (b) Image 2: G_2 (c) Ground truth

(d) Pixel by pixel labeling (e) Ensemble of MRFs (f) Proposed CXM

Fig. 15. Impacts of the multi-layer CXM structure for the quality of the change mask. We compare the results of (d) the pixel by pixel classification without spatial smoothing, (e) the ensemble of three independent, single-layer MRFs and (f) the proposed multi-layer model

The proposed model presents an efficient and scalable change detection filter for several remote sensing applications. The method is based purely on low-level features, working without object extraction or identification of land cover classes. Therefore it can be used for a large variety of scenes and purposes, even in situation where the concept of 'interesting changes' is not well defined. The method can help in manual evaluation of large data sets by focusing the operator's attention, and also in automated systems with decreasing the field of interest and presenting shape or region based descriptors for higher level image interpretation modules.

ACKNOWLEDGEMENT

The authors would like to thank Josiane Zerubia from INRIA for her kind pieces of advice, and the anonymous reviewers for their valuable comments and suggestions.

REFERENCES

[1] T. Peng, I. H. Jermyn, V. Prinet, and J. Zerubia, "Incorporating generic and specific prior knowledge in a multi-scale phase field model for road extraction from VHR images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 1, no. 2, pp. 139–146, June 2008.

[2] F. Lafarge, X. Descombes, J. Zerubia, and M. Pierrot-Deseilligny, "Automatic building extraction from DEMs using an object approach and application to the 3D-city modeling," *Journal of Photogrammetry and Remote Sensing*, vol. 63, no. 3, pp. 365–381, May 2008.

[3] S. Ghosh, L. Bruzzone, S. Patra, F. Bovolo, and A. Ghosh, "A context-sensitive technique for unsupervised change detection based on Hopfield-type neural networks," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 778–789, March 2007.

[4] G. Perrin, X. Descombes, and J. Zerubia, "2D and 3D vegetation resource parameters assessment using marked point processes," in *Proc. International Conference on Pattern Recognition*, Hong-Kong, 2006.

[5] R. Wiemker, "An iterative spectral-spatial bayesian labeling approach for unsupervised robust change detection on remotely sensed multispectral imagery," in *Proc. Int. Conf. on Computer Analysis of Images and Patterns*, vol. LNCS 1296, Kiel, Germany, 1997, pp. 263–270.

[6] L. Bruzzone and D. F. Prieto, "An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images," *IEEE Trans. on Image Processing*, vol. 11, no. 4, pp. 452–466, 2002.

[7] L. Bruzzone, D. Fernandez Prieto, and S. Serpico, "A neural-statistical approach to multitemporal and multisource remote-sensing image classification," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 37, no. 3, pp. 1350–1359, May 1999.

[8] F. Bovolo, L. Bruzzone, and M. Marconcini, "A novel approach to unsupervised change detection based on a semisupervised SVM and a similarity measure," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 46, no. 7, pp. 2070–2082, July 2008.

[9] Y. Bazi, L. Bruzzone, and F. Melgani, "An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 874–887, April 2005.

[10] P. Gamba, F. Dell'Acqua, and G. Lisini, "Change detection of multitemporal SAR data in urban areas combining feature-based and pixel-based techniques," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 44, no. 10, pp. 2820–2827, 2006.

[11] G. Moser, S. Serpico, and G. Vernazza, "Unsupervised change detection from multichannel SAR images," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 2, pp. 278–282, April 2007.

[12] F. Chatelain, J.-Y. Tourneret, and J. Inglada, "Change detection in multisensor SAR images using bivariate gamma distributions," *IEEE Trans. on Image Processing*, vol. 17, no. 3, pp. 249–258, March 2008.

[13] P. Zhong and R. Wang, "A multiple conditional random fields ensemble model for urban area detection in remote sensing optical images," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 3978–3988, December 2007.

[14] W. Liu and V. Prinet, "Probabilistic modeling for structural change inference," in *Asian Conference on Computer Vision*, vol. LNCS 3851, Hyderabad, India, 2006, pp. 836–846.

[15] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: A systematic survey," *IEEE Trans. on Image Processing*, vol. 14, no. 3, pp. 294–307, 2005.

[16] Cs. Benedek, T. Szirányi, Z. Kato, and J. Zerubia, "Detection of object motion regions in aerial image pairs with a multi-layer Markovian model," *IEEE Trans. on Image Processing*, 2009, in press.

[17] Cs. Benedek and T. Szirányi, "Bayesian foreground and shadow detection in uncertain frame rate surveillance videos," *IEEE Trans. on Image Processing*, vol. 4, no. 17, pp. 608–621, 2008.

[18] G. Wen, J. Lv, and W. Yu, "A high-performance feature-matching method for image registration by combining spatial and similarity information," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 46, no. 4, pp. 1266–1277, April 2008.

[19] A. Wong and D. Clausi, "Arrsi: Automatic registration of remote-sensing images," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 45, no. 5, pp. 1483–1493, May 2007.

[20] C. Shah, Y. Sheng, and L. Smith, "Automated image registration based on pseudoinvariant metrics of dynamic land-surface features," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 46, no. 11, pp. 3908–3916, November 2008.

[21] L. Castellana, A. D'Addabbo, and G. Pasquariello, "A composed supervised/unsupervised approach to improve change detection from remote sensing," *Pattern Recogn. Lett.*, vol. 28, no. 4, pp. 405–413, 2007.

[22] Cs. Benedek and T. Szirányi, "Markovian framework for structural change detection with application on detecting built-in changes in

Document version of the MTA SZTAKI Publication Repository, http://eprints.sztaki.hu/

Author manuscript, published in IEEE Trans. on Geoscience and Remote Sensing, vol. 47, no. 10, pp. 3416-3430, 2009

airborne images," in *Proc. Int. Conf. on Signal Processing, Pattern Recognition and Applications*, Innsbruck, Austria, 2007, pp. 68–73.

[23] L. Bruzzone and S. Serpico, "An iterative technique for the detection of land-cover transitions in multitemporal remote sensing images," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 858–867, July 1997.

[24] S. Serpico and G. Moser, "Weight parameter optimization by the Ho-Kashyap Algorithm in MRF models for supervised image classification," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 44, no. 12, pp. 3695–3705, December 2006.

[25] Y. Qi and Z. Rongchun, "A CMRF-based approach to unsupervised change detection in multitemporal remote-sensing images," in *Proc. Int. Conf. on Electronic Measurement and Instruments*, vol. 2, Xi'an, China, 2007, pp. 898–904.

[26] S. Patra, S. Ghosh, and A. Ghosh, "Unsupervised change detection in remote-sensing images using modified self-organizing feature map neural network," in *Proc.International Conference on Computing: Theory and Applications*, Washington, DC, USA, 2007, pp. 716–720.

[27] T. Fung and E. LeDrew, "The determination of optimal threshold levels for change detection using various accuracy indices," *Photogrammetric Engineering and Remote Sensing*, vol. 54, pp. 1449–1454, 1988.

[28] V. Hodge and J. Austin, "A survey of outlier detection methodologies," *Artif. Intell. Rev.*, vol. 22, no. 2, pp. 85–126, 2004.

[29] R. Xu and I. Wunsch, "Survey of clustering algorithms," *IEEE Trans. on Neural Networks*, vol. 16, no. 3, pp. 645–678, 2005.

[30] M. Chini, F. Pacifici, W. Emery, N. Pierdicca, and F. Del Frate, "Comparing statistical and neural network methods applied to very high resolution satellite images showing changes in man-made structures at rocky flats," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 46, no. 6, pp. 1812–1821, June 2008.

[31] G. Mercier, G. Moser, and S. Serpico, "Conditional copulas for change detection in heterogeneous remote sensing images," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1428–1441, May 2008.

[32] G. Camps Valls, L. Gomez Chova, J. Munoz Mari, J. Rojo Alvarez, and M. Martinez Ramon, "Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 46, no. 6, pp. 1822–1835, June 2008.

[33] A. Lorette, X. Descombes, and J. Zerubia, "Texture analysis through a Markovian modelling and fuzzy classification: Application to urban area extraction from satellite images," *International Journal of Computer Vision*, vol. 36, no. 3, pp. 221–236, 2000.

[34] Y. Kosugi, M. Sakamoto, M. Fukunishi, T. Wei Lu Doihara, and S. Kakumoto, "Urban change detection related to earthquakes using an adaptive nonlinear mapping of high-resolution images," *IEEE Geoscience and Remote Sensing Letters*, vol. 1, no. 3, pp. 152–156, 2004.

[35] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.

[36] Z. Kato and T. C. Pong, "A Markov random field image segmentation model for color textured images," *Image and Vision Computing*, vol. 24, no. 10, pp. 1103–1114, 2006.

[37] P. Jodoin, M. Mignotte, and C. Rosenberger, "Segmentation framework based on label field fusion," *IEEE Trans. on Image Processing*, vol. 16, no. 10, pp. 2535–2550, October 2007.

[38] D. Clausi and H. Deng, "Design-based texture feature fusion using Gabor filters and co-occurrence probabilities," *IEEE Trans. on Image Processing*, vol. 14, no. 7, pp. 925–936, July 2005.

[39] A. Kushki, P. Androutsos, K. Plataniotis, and A. Venetsanopoulos, "Retrieval of images from artistic repositories using a decision fusion framework," *IEEE Trans. on Image Processing*, vol. 13, no. 3, pp. 277–292, 2004.

[40] D. Benboudjema and W. Pieczynski, "Unsupervised statistical segmentation of nonstationary images using triplet Markov fields," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 8, pp. 1367–1378, 2007.

[41] A. Fridman, "Mixed Markov models." *Proc. National Academy of Sciences of USA*, vol. 100, no. 14, pp. 8092–8096, July 2003.

[42] Cs. Benedek and T. Szirányi, "A Mixed Markov model for change detection in aerial photos with large time differences," in *Proc. International Conference on Pattern Recognition*, Tampa, Florida, USA, 2008.

[43] J. A. Bilmes, "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models," International Computer Science Institute and Computer Science Division, University of California at Berkley, Berkley, CA, Tech. Rep. TR-97-021, April 1998.

[44] M. Benšić and K. Sabo, "Estimating the width of a uniform distribution when data are measured with additive normal errors with known variance," *Computational Statistics & Data Analysis*, vol. 51, no. 9, pp. 4731–4741, 2007.

[45] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 1, Hawaii, USA, 2001, pp. 511–518.

[46] G. J. Hahn and S. S. Shapiro, *Statistical models in engineering*. New York: John Wiley & Sons, 1994, p. 95.

[47] R. Potts, "Some generalized order-disorder transformation," in *Proceedings of the Cambridge Philosophical Society*, no. 48, 1952, p. 106.

[48] Z. Kato, J. Zerubia, and M. Berthod, "Satellite image classification using a modified Metropolis dynamics," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, March 1992, pp. 573–576.

[49] J. Besag, "On the statistical analysis of dirty images," *Journal of Royal Statistics Society*, vol. 48, pp. 259–302, 1986.

[50] Z. Kato, J. Zerubia, and M. Berthod, "Unsupervised parallel image classification using Markovian models," *Pattern Recognition*, vol. 32, no. 4, pp. 591–604, 1999.

[51] K. Fukunaga and R. Hayes, "The reduced Parzen classifier," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, no. 4, pp. 423–425, April 1989.

[52] A. Licsár, T. Szirányi, and L. Czúni, "Trainable blotch detection on high resolution archive films minimizing the human interaction," *Machine Vision and Applications*, November 2007, online.

Csaba Benedek received the M.Sc. degree in computer sciences in 2004 from the Budapest University of Technology and Economics, and the Ph.D. degree in image processing in 2008 from the Pázmány Péter Catholic University, Budapest. Since 2008, he holds a research fellow position at the Distributed Events Analysis Research Group at the Computer and Automation Research Institute, Hungarian Academy of Sciences. He is currently working as a postdoctoral researcher with the Ariana Project Team at INRIA Sophia-Antipolis, France. His research interests include Bayesian image segmentation, change detection, video surveillance and aerial image processing.

Tamás Szirányi received the Ph.D. degree in electronics and computer engineering in 1991 and the D.Sci. degree in 2001 from the Hungarian Academy of Sciences, Budapest. He was appointed to a Full Professor position an 2001 at Veszprém University, Hungary, and in 2004, at the Pázmány Péter Catholic University, Budapest. He is currently a scientific advisor at the Computer and Automation Research Institute, Hungarian Academy of Sciences, where he is the head of the Distributed Events Analysis Research Group. His research activities include texture and motion segmentation, surveillance systems for panoramic and multiple camera systems, measuring and testing the image quality, digital film restoration, Markov Random Fields and stochastic optimization, image rendering and coding.

Dr. Szirányi was the founder and first president (1997 to 2002) of the Hungarian Image Processing and Pattern Recognition Society. Between 2002 and 2008 he was an Associate Editor of IEEE TRANSACTIONS ON IMAGE PROCESSING. He was honored with the Master Professor award in 2001. He is a fellow of IAPR.

Document version of the MTA SZTAKI Publication Repository, http://eprints.sztaki.hu/