# Multilayer Markov Random Field models for change detection in optical remote sensing images

CrossMark

Csaba Benedek [a,*], Maha Shadaydeh [a], Zoltan Kato [b], Tamás Szirányi [a], Josiane Zerubia [c]

[a] Distributed Events Analysis Research Laboratory, Institute for Computer Science and Control, Hungarian Academy of Sciences, H-1111 Budapest, Kende u. 13-17, Hungary
[b] Institute of Informatics, University of Szeged, 6701 Szeged, P.O. Box 652, Hungary
[c] INRIA, AYIN team, 2004 route des Lucioles, 06902 Sophia Antipolis, France

## ARTICLE INFO

## ABSTRACT

In this paper, we give a comparative study on three Multilayer Markov Random Field (MRF) based solutions proposed for change detection in optical remote sensing images, called *Multicue MRF*, *Conditional Mixed Markov* model, and *Fusion MRF*. Our purposes are twofold. *On one hand*, we highlight the significance of the focused model family and we set them against various state-of-the-art approaches through a thematic analysis and quantitative tests. We discuss the advantages and drawbacks of class comparison vs. direct approaches, usage of training data, various targeted application fields and different ways of Ground Truth generation, meantime informing the Reader in which roles the Multilayer MRFs can be efficiently applied. *On the other* hand we also emphasize the differences between the three focused models at various levels, considering the model structures, feature extraction, layer interpretation, change concept definition, parameter tuning and performance. We provide qualitative and quantitative comparison results using principally a publicly available change detection database which contains aerial image pairs and Ground Truth change masks. We conclude that the discussed models are competitive against alternative state-of-the-art solutions, if one uses them as pre-processing filters in multitemporal optical image analysis. In addition, they cover together a large range of applications, considering the different usage options of the three approaches.

© 2015 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

## 1. Introduction

Automatic evaluation of aerial image repositories is an important field of research, since periodically repeated manual processing is time-consuming and cumbersome in cases of high number of images and dynamically changing content. Change detection is an important part of many remote-sensing applications. Some country areas are scanned frequently (e.g. year-by-year) to spot relevant changes, and several repositories contain multitemporal image samples for the same area. Through the extraction of changes, the regions of interest in the images can be significantly decreased in several cases, helping applications of 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 analyzer modules.

However, the definition of "relevant change" is highly task-specific, leading to a large number of change detection methods with significantly different goals, assumptions and applied tools. Even for a given specific problem the data comparison may be notably challenging, considering that due to the large time lag between two consecutive image samples, one must expect seasonal changes, differences in the obtained data quality and resolution, 3D geometric distortion effects, various viewpoints, different illumination, or results of irrelevant human intervention (such as crop rotation in the arboreous lands).

### 1.1. Related work

The change detection algorithms in the literature can be grouped based on various aspects. First, they may follow either the *Post-Classification Comparison (PCC)* or the *direct* approach. Second, depending on the availability of training data, they can be *supervised* or *unsupervised*. Third, they may realize *region based*

* Corresponding author.
E-mail addresses: benedek.csaba@sztaki.mta.hu (C. Benedek), shadaydeh.maha@sztaki.mta.hu (M. Shadaydeh), kato@inf.u-szeged.hu (Z. Kato), sziranyi.tamas@sztaki.mta.hu (T. Szirányi), josiane.zerubia@inria.fr (J. Zerubia).

(e.g. detecting new forest regions), or *object based* (e.g. searching for changed buildings) scene interpretation.

### 1.1.1. PCC vs. direct approaches

*PCC* methods (Liu and Prinet, 2006; Castellana et al., 2007; Zhong and Wang, 2007; Szirányi and Shadaydeh, 2014) segment first the input images into various land-cover classes, like urban areas, forests, plough lands, etc. In this case, changes are obtained indirectly as regions with different class labels in the different time layers. On the other hand *direct* methods (Wiemker, 1997; Bruzzone and Fernandez-Prieto, 2002; Bazi et al., 2005a; Ghosh et al., 2007; Benedek and Szirányi, 2009; Singh et al., 2014) derive a similarity-feature map between the multitemporal input images (e.g. a difference image (DI) or a block correlation map) and then they cluster the feature map to separate changed and unchanged areas. A straightforward advantage of PCC approaches is that besides change detection, they classify the observed differences at the same time (e.g. a forest region turns into a built-in area). A difficulty with region comparison is that, in several cases we must rely on noisy cluster descriptors during the classification step, and the exact borders of the clusters in the images may be ambiguous. For this reason, if we apply two independent segmentation algorithms for the two images, the segmented regions may have slightly different shapes and sizes, even if the image parts have not changed in fact. As possible solutions, the segmentation quality can be enhanced by interactive segmentation of the images (Benedek and Szirányi, 2007) or exploiting estimated class transition probabilities (Castellana et al., 2007; Liu et al., 2008).

Since *direct* methods do not use explicit land cover class models, the change detection process does not require defining various segmentation classes with reliable feature models. In addition, in several applications 'intra-class' transitions – which are ignored by *PCC* methods – may also be worth of attention: e.g. inside an urban region, it could be necessary to detect destroyed or re-built buildings, relocated roads, etc. On the other hand, due to lack of exact change definition, it is much more difficult to describe the validity range of these models.

### 1.1.2. Supervised and unsupervised models

Another important point of view is distinguishing *supervised* (Serpico and Moser, 2006; Castellana et al., 2007, 2008; Fernandez-Prieto and Marconcini, 2011) and *unsupervised* techniques (Wiemker, 1997, 2006,, 2007,,, 2008, 2011, 2014). Since *unsupervised* methods do not use manually labeled Ground Truth data, they usually rely on prior assumptions (Fung and LeDrew, 1988), such as the area of unchanged regions is significantly larger. In that case, changes can be obtained through outlier detection (Hodge and Austin, 2004) or clustering (Xu and Wunsch, 2005) in the feature space. However, as shown by Benedek and Szirányi (2009), in optical images the feature statistics for the different classes may be multi-modal and strongly overlapping, therefore unsupervised separation is usually more challenging than in models using multispectral measurements (Bruzzone and Fernandez-Prieto, 2002; Bovolo et al., 2008). Especially, atmospheric and light variations may result in artifacts for change detection on optical images (Castellana et al., 2007). On the other hand if training data is available, we can gain a significant amount of additional information for the classification process. In many real applications, the image repositories contain large batches of images from the same year taken with the same quality, camera settings and similar seasonal and illumination conditions, where it can be admissible to prepare Ground Truth from a minor part of the available data.

### 1.1.3. Targeted scenarios

Differences between approaches can also be taken regarding the exact application goals and circumstances. Several methods deal only with either agricultural (Kumar et al., 2012) or urban (Liu and Prinet, 2006) territories, moreover, they often focus on a specific task like built-up area extraction (Lorette et al., 2000; Zhong and Wang, 2007; Benedek and Szirányi, 2007), disaster assessment after earthquakes (Kosugi et al., 2004), floods (Martinis et al., 2011; Wang et al., 2013), and mine countermeasure in oceans (Shuang and Leung, 2012). Besides region level change monitoring, a number of approaches consider the change detection task as a problem of training-based object recognition, for example Liu and Prinet (2006) proposed applications for building development monitoring. However the latter approach can only be used if the changes can be interpreted at object levels, where we need a precisely restricted environment.

### 1.2. Markovian change detection models

As the above discussion already foreshows, visual change detection is in itself a largely diversified topic, and giving a complete overview would extend the scope of this article. Therefore we introduce various specifications for our investigations in this comparative paper: we limit our survey to *region level* non-object-based approaches working on *optical* remote sensing images. Among the different modeling tools we focus on the comparison of Multilayer Markov Random Field based techniques.

At the region level of change detection, Markov Random Fields (MRFs) (Kato and Zerubia, 2012) are widely used tools since the early eighties (Kalayeh and Landgrebe, 1986; Solberg et al., 1996; Bruzzone and Fernandez-Prieto, 2000). MRFs are able to simultaneously embed a data model, reflecting the knowledge on the measurements; and prior constraints, such as spatial smoothness of the solution through a graph based image representation, where nodes belong to different pixels and edges express direct interactions between the nodes. Although a number of the corresponding MRF based state-of-the art models deal with multispectral (Bruzzone and Fernandez-Prieto, 2002; Ghosh et al., 2007; Xu et al., 2012; Chen and Cao, 2013; Ghosh et al., 2013; Subudhi et al., 2014) or SAR (Melgani and Serpico, 2003; Bazi et al., 2005b; Carincotte et al., 2006; Gamba et al., 2006; Martinis et al., 2011; Wang et al., 2013; Baselice et al., 2014) imagery, the significance of handling optical images is also increasing (Zhong and Wang, 2007; Benedek and Szirányi, 2009; Moser et al., 2011; Szirányi and Shadaydeh, 2014; Hoberg et al., 2015).

Since conventional MRFs show some limitations regarding context dependent class modeling, different modified schemes have been recently proposed to increase their flexibility. Triplet Markov fields (Wang et al., 2013) contain an auxiliary latent process which can be used to describe various subclasses of each class in different manners. Mixed Markov models (Fridman, 2003) extend MRFs by admitting data-dependent links between the processing nodes, which enables configurable structures in feature integration. Conditional Random Fields (CRFs) directly model the data-driven posterior distributions of the segmentation classes (Chen et al., 2007; Hoberg et al., 2012; Hoberg et al., 2015).

### 1.3. Multilayer segmentation models

We continue with the discussion of feature selection. For many problems, scalar valued features alone may be weak to model complex segmentation clusters appropriately, thus the integration of multiple observations is a crucial issue. In a straightforward solution called *observation fusion*, the different feature components are integrated into an *n* dimensional feature vector, and for each class, the distribution of the features is approximated by an *n* dimensional multinomial density function (Clausi and Deng, 2005; Kato and Pong, 2006). For example, one can fit a Gaussian mixture to the multivariate feature histogram of the training

images (Kato and Pong, 2006), where the different mixture components correspond to the different classes or subclasses. However, in the above case, each relevant prototype of a given class should be represented by a significant peak in the joint feature histogram, otherwise the observation fusion approach becomes generally less efficient.

The *multilayer segmentation models* can overcome the before mentioned limitation (Kato et al., 2002; Reed et al., 2006; Jodoin et al., 2007; Benedek et al., 2009). Here the layers correspond usually to different segmentations which interact through prescribed inter-layer constraints. The model is called *decision fusion* if the layers are first segmented independently by e.g. MRFs, thereafter, a pixel by pixel fusion process inferences purely on the obtained labels (Reed et al., 2006), followed by a smoothing step. We can also mention here Factorial MRF models (Kim and Zabih, 2002) with multiple interacting MRF layers, or the *fusion-reaction* framework proposed by Jodoin et al. (2007), which implements a sequential process of over-segmentation and label fusion.

A *multilayer MRF* framework has been introduced in Kato et al. (2002), where a single energy function encapsulates all the constraints of the model, and the result is obtained by a global optimization process in one step. Here, in contrast to decision (Reed et al., 2006) or label (Jodoin et al., 2007) fusion, the observed features are in interaction with the final label map during the whole segmentation process.

### 1.4. Outline of the paper

Different multilayer MRF techniques have recently been proposed for change detection, which differ in both feature selection and in model structure. Since the literature of multilayer MRFs is not as deeply established as in the single-layer case, it is often not straightforward to decide from the application point of view, which are the advantages and drawbacks of the different approaches in a given situation.

The goal of this paper is to present a comparative study about three multilayer MRF techniques, developed earlier in the research laboratories of the authors. Section 2 provides an overview on the three methods, focusing briefly on the similarities and differences between them. In Sections 3–5, each method is introduced following the same presentation scheme so that the Reader can follow the main similarities and differences in the model structures, used features and the working constraints. Sections 6 and 7 cover the optimization and parameter settings issues, respectively. In the experimental Section 8, quantitative and qualitative comparison will be provided relying principally on the SZTAKI AirChange Benchmark Set (Benedek and Szirányi, 2009).

## 2. Overview on the three compared models

In the paper, we compare three state-of-the-art multilayer MRF techniques for change detection, which have been developed for optical remote sensing image analysis. For a graphical comparison, Figs. 1–3 show the structures and the processing workflows of the three models. The main properties of the different techniques are summarized in Table 1.

The *Multicue MRF model* (L₃MRF, Fig. 1) (Singh et al., 2014) integrates the modified Histogram of Oriented Gradients (HOGs) and graylevel difference (GLD) features into the original Multi-MRF structure framework proposed by Kato et al. (2002), where two layers correspond to the two feature maps and the third one is the final segmentation layer. The class models and the inter-layer interaction terms are both affected by observation dependent and prior constraints, while within a given layer the prior Potts terms (Potts, 1952) ensure a smooth segmentation.

The second discussed method is the *Conditional Multilayer Mixed MRF* (CXM, Fig. 2) proposed by Benedek and Szirányi (2009). CXM has – in a first view – a similar structure to the L₃MRF, however, we can find a number of key differences. First, CXM is a multilayer implementation of Mixed Markov models (Fridman, 2003), which uses besides regular MRF nodes the so called address nodes in the graph. Address nodes can link regular nodes in a flexible way based on various data- and prior conditions, ensuring a configurable graph structure. Second, in CXM the feature maps only affect directly the individual graph nodes, while the interaction terms implement purely prior label fusion soft-constraints. Third, the applied features are also different from L₃MRF. In CXM, the change detection result is based on two weak features: global intensity co-occurrence statistics (ICS) and block correlation; while as a third feature, a contrast descriptor locally estimates the reliability of the weak change descriptors at each pixel position.

Regarding the targeted application fields and the expected image inputs, the scopes of the L₃MRF and CXM model are quite similar. Both work on grayscale inputs and their output is a binary change mask. As practical differences, CXM may handle better scenarios when large radiometric differences may occur between unchanged regions, while L₃MRF is quicker and less sensitive to the presence of registration errors.

While the L₃MRF and CXM models have a similar structure, the *Fusion-MRF Model* (FMRF, Fig. 3) proposed by Szirányi and Shadaydeh (2014) follows a significantly different approach. First, while L₃MRF and CXM are *direct* techniques working without any land cover class models, FMRF is a Post-Classification Comparison (PCC) method which simultaneously implements an adaptive segmentation and change detection model for optical remote sensing images. Even the concept of layer is significantly different here. While in the first two models the different layers correspond to different image features, in FMRF each layer represents given input image; Thus 'multi-layers' refers to multi-temporal images.

As another important difference between the models, FMRF has been designed to compare several images (two or more) from different time instances, while the L₃MRF and CXM methods can compare always two multitemporal images. It is also possible to use the FMRF with an input of an image pair, but the use of three or more images are generally preferred to enhance cluster definitions depending on the quality of the images and the degree of similarity between these images. The FMRF method applies clustering on a fused image series by using the Cluster Reward Algorithm (CRA) (Inglada and Giros, 2004) as cross-layer similarity measure, followed by a multilayer MRF segmentation (see Fig. 3). The resulted label map is applied for the automatic training of the single layers. After the segmentation of each single layer separately, changes are detected between the single label-maps.

Although the selected techniques have been mainly tested on optical images, no specific information of image sources is prescribed. However, since L₃MRF and CXM are based on single channel inputs, they typically expect intensity images provided by airborne or spaceborne sensors. On the other hand, the FMRF method can deal with multi-modal as well as multispectral images. Here during the tests with FMRF, the chrominance (i.e. color) channels of the images are also exploited, while in the original paper (Szirányi and Shadaydeh, 2014), the Reader may also find an example for fusion with an infra-red image.

There are also differences in the applied MRF optimization techniques, which affect the quality and computational speed of the change detection process. Due to the sub-modular structures of the L₃MRF and FMRF models, the energy function can be optimized by the efficient graph-cut based technique. On the other hand, the complex structure components of CXM yield that the energy optimization process is more complicated, and the computationally more expensive simulated annealing algorithm should be adopted.
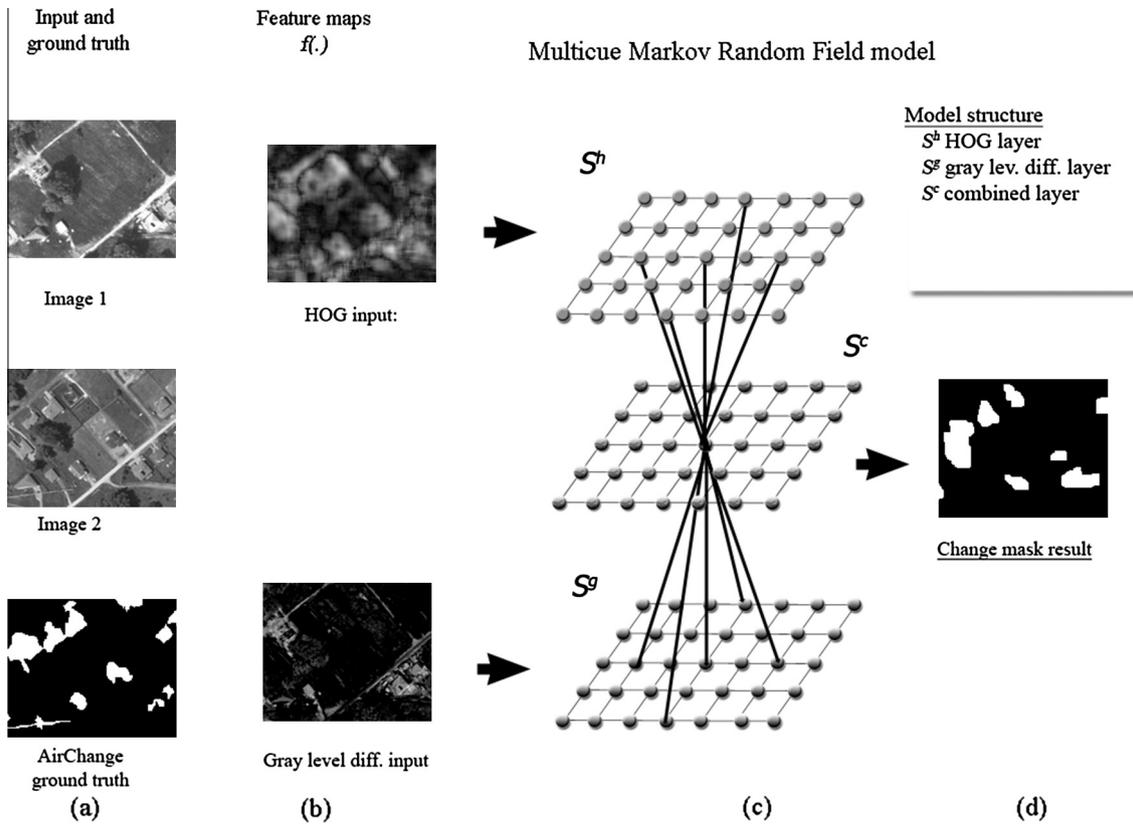
**Fig. 1.** Structure of the L₃MRF model and overview of the Multicue change detection process. Column (a): registered input images and Ground Truth change mask for validation, Column (b): feature maps, Column (c): structure diagram of the L₃MRF model, Column (d): output change mask.
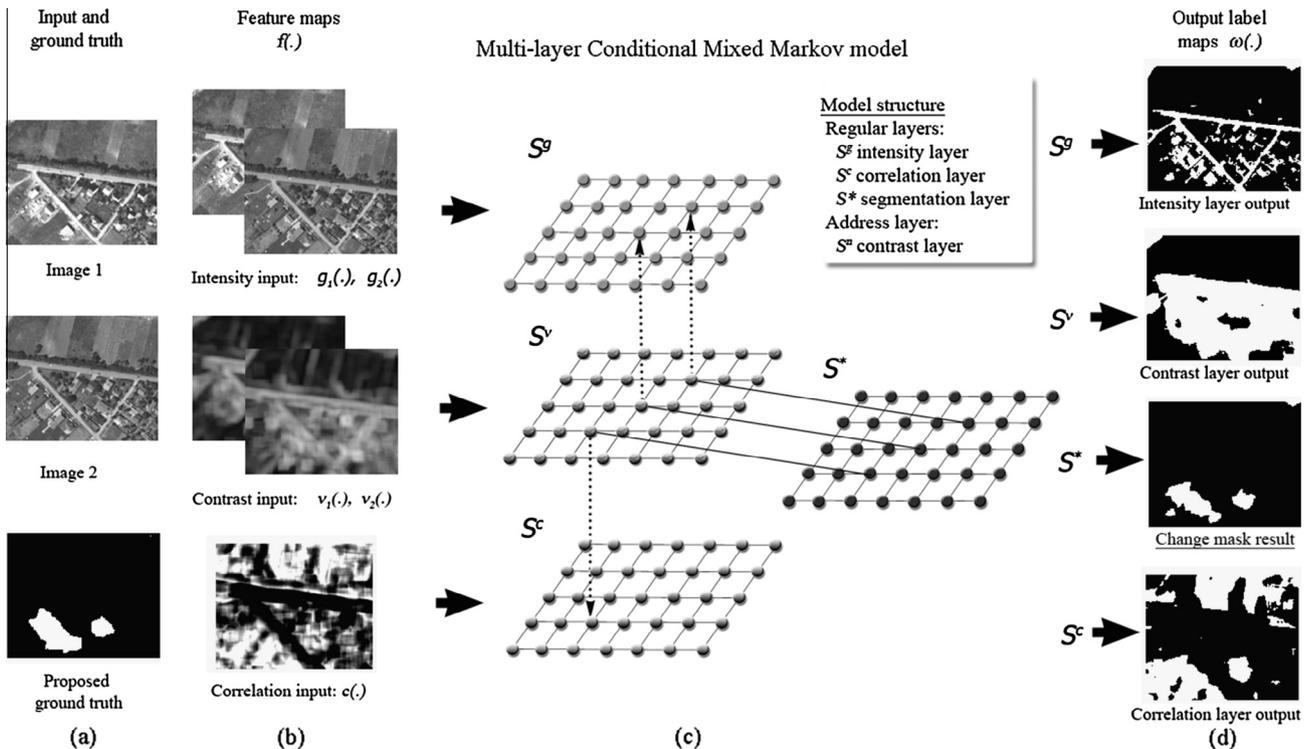


**Fig. 2.** Structure of the CXM model and overview of the segmentation process. Column (a): inputs and Ground Truth, Column (b): $\overline{g}(\cdot)$, $\overline{v}(\cdot)$ and $c(\cdot)$ feature maps extracted from the input image pair. Column (c): Structure diagram of the CXM model. (note: the inter-layer connections are only shown regarding three selected pixels), Column (d): Output label maps of the four layers after MMD optimization. The segmentation result is obtained as the labeling of the $S^*$ layer.
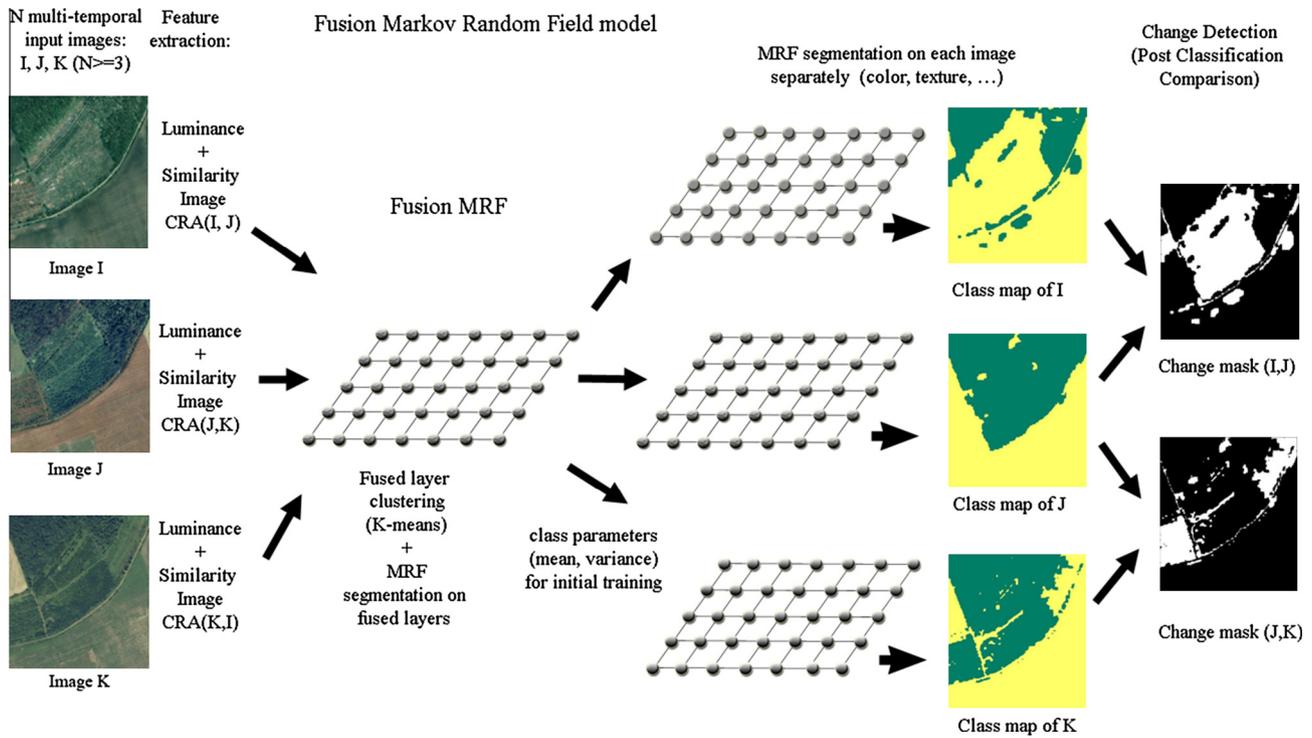
**Fig. 3.** Structure of the FMRF model and workflow of the implemented *Post-Classification Comparison* process.

**Table 1**
Main properties of the discussed three models.

| Method | $L_3$MRF | CXM | FMRF |
|---|---|---|---|
| Model type | Multicue MRF | Mixed Markov model | Fusion-MRF |
| Change category | Direct approach | Direct approach | Post-Classification Comparison (PCC) |
| Number of input images | 2 images | 2 images | $\geqslant 2$ images |
| Number of layers | 3 | 4 | Equal to the num. of input images |
| Used image channels | Gray level intensity | Gray level intensity | Luminance and color channels |
| Key image features | Histogram of Oriented Gradients, gray level difference | Joint intensity histogram, block correlation, variance | Color, texture, multi-structure |
| Inter layer interactions | Data- and label fusion terms | Purely label fusion | Segmentation in the joint multitemporal image domain, CRA similarity calculation |
| Parameter estimation | Supervised training | Supervised training | Unsupervised or supervised |
| Optimization | Graph cut | Modified Metropolis | Graph cut |

As for the use of training data, L₃MRF and CXM are supervised models, i.e. the feature model parameters are set using training regions. On the other hand, the FMRF model may be used both in an unsupervised and in a supervised way upon the availability of manually labeled sample regions.

Regarding the necessary image resolution, our experiments showed that a minimum of 0.5 m/pixel is expected, if we want to highlight e.g. built-in changes in semi-urban areas. There is no explicit upper limit for the image resolution, since the focused techniques use pixel-level and block-based features, and the sizes of blocks can be upscaled for photos with larger resolution (some experiments with 1.5 m resolution images were presented in (Benedek and Szirányi, 2009)). For the FMRF model, image resolution affects the accuracy of detected changes as discussed in Section 7.3.

Although the discussed three models are able to support various applications, they also face some joint limitations. First, since the methods are purely based on pixel level or rectangular block-based image features, only very limited amount of geometric object information can be incorporated in the models. Therefore at object level, geometric approaches such as Marked Point Processes

(Benedek et al., 2012) could be used more efficiently. Secondly, since the outputs are binary change/no change masks, the techniques are *directly* not suitable for highlighting specific kinds of changes or novelty detection (such as distinguishing new and demolished buildings). However, there are some *indirect* options for change classification. Due to the nature of the PCC approach, if we train the supervised FMRF with semantically meaningful training classes (such as built-in vs. natural areas), we can provide a classification of the changes through the different class transitions. We can also cluster the changes in the CXM model in a limited way, by highlighting only the ICS based (homogeneous regions) or the correlation based (textured urban areas) differences.

## 3. Change detection with the Multicue MRF model

In Singh et al. (2014), a *Multicue Markovian model* has been proposed for change detection in registered optical aerial image pairs with large time differences. A *Multicue Markov Random Field* takes into account information at two feature layers from two different sets of features:

- difference of a modified Histogram of Oriented Gradients (HOGs) and
- graylevel (GL) difference.

The third layer provides the final change detection by a probabilistic combination of the two feature layers via MRF interactions. Thus we integrate both the texture level as well as the pixel level information to generate the final result. The proposed model uses pairwise interactions which also ensures that sub-modularity condition is satisfied. Hence a global energy optimum is achieved using a standard max-flow/min-cut algorithm ensuring homogeneity in the connected regions (Singh et al., 2014).

### 3.1. Image model and features

Let us consider a registered gray scale image pair, $G_1$ and $G_2$, over the same pixel lattice $S = \{s_1, s_2, \ldots, s_N\}$. We denote the grayscale values of a given pixel $s \in S$ by $g_1(s)$ and $g_2(s)$ in the first and second image, respectively. The goal is to classify each site $s \in S$ as changed (*foreground*) or unchanged (*background*). Hence the assignment of a label to a particular site is from the set: $\lambda = \{fg, bg\}$ where fg refers to foreground class and bg refers to the background class. The background/foreground classes are modeled as random processes generating the observed image features. These random processes are modeled by fitting a suitable distribution function over the histograms corresponding to each of the foreground and background classes using a set of training image pairs. The training image pairs contain a Ground Truth having all the pixels manually labeled by an expert. The features adopted in Singh et al. (2014) characterize the changes in terms of intensity (GL difference) and in terms of texture/structure (HOG difference).

The graylevel (GL) difference feature $d(s)$ computed over the aligned image pairs for each corresponding pixel is simply

$$d(s) = \|g_1(s) - g_2(s)\|. \tag{1}$$

Analysing the histogram of the background class using the WAFO toolbox (Brodtkorb et al., 2000), the generalized gamma density proved to be a good parametric model to represent these features, thus

$$P(d(s)|bg) = f(d(s)|a, b, c) = \frac{c}{b^{ac}\Gamma(a)} d(s)^{ac-1} e^{-\left(\frac{d(s)}{b}\right)^c}, \tag{2}$$

where $\Gamma(\cdot)$ is the gamma function, and $(a, b, c)$ are the background parameters.

As for the foreground, basically any $d(s)$ value is allowed, hence it is represented by a uniform density function given as

$$P(d(s)|fg) = \begin{cases} \frac{1}{b_d - a_d}, & \text{if } d(s) \in [a_d, b_d]. \\ 0, & \text{otherwise}. \end{cases} \tag{3}$$

using $(a_d, b_d)$ foreground parameters.

The Histogram of Oriented Gradient (HOG) is a feature descriptor that has mainly been used for object detection. It basically involves counting the number of occurrences of different orientations of gradients inside fixed sized windows and then rounding it to the correct bin of the histogram. In Dalal and Triggs (2005) the image was divided into blocks and then further into small cells for which the histogram of gradients is computed. Finally the concatenation of all resulting histograms leads to the descriptor for the entire image. The HOG feature used in Singh et al. (2014) is a somewhat modified version of the original method proposed in Dalal and Triggs (2005): instead of cells, a sliding window of size $11 \times 11$ is used to compute HOG. Given the gradients $(I_x, I_y)$ computed at each pixel via a standard finite difference operator, the magnitude and orientation can be calculated using the following equations:

$$\|H\| = \sqrt{I_x^2 + I_y^2} \quad \theta = \arctan\left(\left\|\frac{I_y}{I_x}\right\|\right)$$

Note that $\theta \in [0, \pi/2]$. Then a HOG with *9 bins* is computed at every position of the sliding window over the entire image, yielding a 9 dimensional vector $\overrightarrow{f_s}$ associated with every pixel $s \in S$. The HOG difference feature $h(s)$ corresponding to a particular pixel $s$ is thus given by:

$$h(s) = \|\overrightarrow{f_{s_1}} - \overrightarrow{f_{s_2}}\| \tag{4}$$

As for the GL difference feature, the background is modeled as a generalized gamma density function $P(h(s)|bg) = f(h(s)|u, v, w)$ with parameters $u$, $v$, $w$ and the foreground is represented as a uniform distribution over the $[a_h, b_h]$ interval.

### 3.2. Multicue model

Using the features $d(s)$ and $h(s)$, a three-layer MRF model is constructed in Singh et al. (2014) to solve the change detection problem as a foreground/background segmentation. As shown in Fig. 1, the proposed MRF segmentation model (Kato et al., 2002; Kato and Zerubia, 2012) is built over a Graph $\mathcal{G}$ composed of three different layers, namely $S^h$, $S^c$ and $S^g$, all being of the same size as the lattice $S$ of the input images. Each pixel $s \in S$ has a corresponding site associated with it in each of these layers denoted as

$$s^h \in S^h, s^c \in S^c, s^g \in S^g \tag{5}$$

where $S^h$, $S^c$ and $S^g$ are the layers representing the Modified HOG difference feature, final desired change map, and GL difference feature respectively.

Every site $s^i$, $i \in \{h, c, g\}$ has also a class label associated to it, which is denoted by $\omega(s^i)$ and modeled as a discrete random variable taking values from the label set $\lambda = \{fg, bg\}$. The hidden label process is thus the set of all the labels over the entire graph $\mathcal{G}$ as follows:

$$\underline{\omega} = \{\omega(s^i)|s \in S, i \in \{h, c, g\}\} \tag{6}$$

The neighborhood system (representing the conditional dependencies of nearby sites) are shown in Fig. 1. The intra-layer interactions consist of *singleton* and *doubleton* cliques denoted by $C_1$ and $C_2$, respectively. These cliques correspond to a standard first order neighborhood system (Kato and Zerubia, 2012), in which *singletons* with single individual sites are linking the model to the two observation features, while *doubletons* ensure homogeneity within each layer. Note that *singletons* are not defined for the combined layer $S^c$ as it has no direct interaction with the observations, while for the other two layers, *singletons* represent the HOG features for $S^h$ and GL features for $S^g$, yielding the observation process

$$\mathcal{F} = \{h(s^h)|s \in S\} \cup \{d(s^g)|s \in S\}, \tag{7}$$

The inter-layer cliques, marked by $C_5$, are doubletons as displayed in Fig. 1, and they are responsible for feature integration. Hence the graph $\mathcal{G}$ has the set of cliques

$$C = C_1 \cup C_2 \cup C_5 \tag{8}$$

The goal is to find the optimal labeling $\widehat{\underline{\omega}}$ which maximizes the a posteriori probability $P(\underline{\omega}|\mathcal{F})$, which is the maximum a posteriori (MAP) estimate (Geman and Geman, 1984) given as

$$\widehat{\underline{\omega}} = \underset{\underline{\omega} \in \Omega}{\operatorname{argmax}} P(\underline{\omega}|\mathcal{F}) \tag{9}$$

where $\Omega$ denotes the set of all the possible labelings. Based on the Hammersley–Clifford theorem, the posterior probability for a particular labeling follows a Gibbs distribution:

$$P(\underline{\omega}|\mathcal{F}) = \frac{exp(-U(\underline{\omega}))}{Z} = \frac{1}{Z}exp\left(-\sum_{C\in\mathcal{C}}V_C(\underline{\omega}_C)\right) \tag{10}$$

where $U(\underline{\omega})$ is the energy function, $V_C$ denotes the *clique potential* of a clique $C \in \mathcal{C}$ having the label configuration $\underline{\omega}_C$, and $Z$ is a normalization constant independent of $\underline{\omega}$ given by $Z = \sum_{\underline{\omega}\in\Omega}\exp(-U(\underline{\omega}))$. Therefore by defining the potentials $V_C$ for the cliques completes the MRF model definition and the above MAP problem becomes a standard energy minimization problem.

Since the labelings for $S^h$ and $S^g$ layers are directly influenced by the values of $h(\cdot)$ and $d(\cdot)$ respectively, $\forall s \in S$ the singleton will link these layers to the respective observations as:

$$V_{\{s^h\}}(\omega(s^h)) = -\log P(h(s)|\omega(s^h)) \tag{11}$$

$$V_{\{s^g\}}(\omega(s^g)) = -\log P(d(s)|\omega(s^g)), \tag{12}$$

where the conditional probabilities $P(\cdot|\omega(\cdot))$ given that the label being a background or a foreground class generate the $h(s)$ or $d(s)$ observations, has already been defined in Section 3.1. However, the labels corresponding to the sites of the $S^c$ layer have no direct influence by these observations and hence the singleton cliques on $S^c$ are not used.

For the intra-layer cliques $C_2 = \{s^i, r^i\}$ where $C_2 \in \mathcal{C}_2$ and $i \in \{h, c, g\}$, a simple Ising-type potential (Besag, 1986) can be used to ensure local label homogeneity:

$$V_{C_2}(\omega(s^i), \omega(r^i)) = \begin{cases} 0, & \text{if } \omega(s^i) = \omega(r^i). \\ 2\mathcal{K}^i, & \text{if } \omega(s^i) \neq \omega(r^i). \end{cases} \tag{13}$$

where $i \in \{h, c, g\}$. $\mathcal{K}^i \geqslant 0$ is a parameter controlling the homogeneity of the regions. As $\mathcal{K}^i$ increases, the resulting regions in the corresponding layer (indexed $i$) become more homogeneous.

Finally, the inter-Layer cliques are $C_5 = \{s^i, r^j\}$ where $C_5 \in \mathcal{C}_5$ and $\{i, j\} \in \{\{h, c\}, \{c, g\}\}$, and the potentials are given by:

$$V_{C_5} = f(\omega(s^i), \omega(r^j)) = \begin{cases} \rho^{hc} * W_r * \|V_{\{s^h\}}(\omega(s^h)) - V_{\{s^h\}}(\omega(r^h))\|, & \text{(a)}. \\ \rho^{cg} * W_r * \|V_{\{s^g\}}(\omega(s^g)) - V_{\{s^g\}}(\omega(r^g))\|, & \text{(b)}. \end{cases} \tag{14}$$

where (a) and (b) refer to $\{i, j\} = \{h, c\}$ and $\{i, j\} = \{c, g\}$ respectively. Also $V_{\{s^h\}}(\omega(\cdot))$ and $V_{\{s^g\}}(\omega(\cdot))$ are the singleton potentials for the sites $s^h \in S^h$ and $s^g \in S^g$ dependent on the labeling $\omega(\cdot)$. Both parameters ($\rho^{hc} \geqslant 0$ and $\rho^{cg} \geqslant 0$) control the influence of the feature layers ($S^h$ and $S^g$) to the combined ($S^c$) layer. $W_r$ is the weight of the interaction. Higher weight ($W_r$ = 0.6) is assigned to the corresponding site whereas smaller weights ($W_r$ = 0.1 each) to the other 4 neighboring sites.

Hence the optimal MAP labeling $\widehat{\underline{\omega}}$, which maximizes $P(\underline{\omega}|\mathcal{F})$ can be calculated as the minimum energy configuration (Singh et al., 2014)

$$\widehat{\underline{\omega}} = \underset{\underline{\omega}\in\Omega}{\operatorname{argmin}} - \sum_{s\in S}\log P(h(s)|\omega(s^h)) - \sum_{s\in S}\log P(d(s)|\omega(s^g))$$
$$+ \sum_{C_2\in\mathcal{C}_2}V_{C_2}(\underline{\omega}_{C_2}) + \sum_{C_5\in\mathcal{C}_5}V_{C_5}(\underline{\omega}_{C_5}). \tag{15}$$

## 4. Change detection with the Multilayer Conditional Mixed Markov model

In Benedek and Szirányi (2009), a Multilayer Conditional Mixed Markov model (CXM) has been proposed for change detection in optical image pairs. The CXM is defined as the combination of a mixed Markov model (Fridman, 2003) and a conditionally independent random field of signals. The model has four layers: three of them are feature layers, while the fourth one is the final segmentation layer, having a similar role to the combined layer of the L$_3$MRF model. The feature layers correspond to the following descriptors:

- joint 2D gray level histogram of the multitemporal image inputs,
- normalized block correlation using a fix sized window around each pixel,
- variance of gray levels within a sliding window for local contrast estimation.

Since the gray level histogram and the correlation are complementary features working either in homogeneous or in highly textured image region, the variance descriptor plays an indicator role: Based on the local contrast, one can decide which feature is more reliable at a given pixel location. The fourth layer provides the final change detection by a probabilistic combination of the three feature layers via MRF interactions. Similarly to L$_3$MRF, pairwise interactions are used between the nodes, but as a significant difference, the variance layer is composed of address nodes of the Mixed Markov model, while all the other layers contain regular nodes.

### 4.1. Image model and features

Here the same image model and the corresponding notations are used as introduced in Section 3.1 by the L$_3$MRF model. The starting step of the modeling process is again the extraction of *local features* at each $s \in S$ which give us information for classifying $s$ as a changed *foreground* (fg) or unchanged *background* (bg) surface point. The fg/bg classes as considered henceforward as random processes generating the features according to different distributions.

The *first feature* in the CXM model is based on the investigations in the joint intensity domain of the two images. Here, instead of prescribing an intensity offset or other global linear transform between the corresponding $g_1(s)$ and $g_2(s)$ gray levels (as seen by the L$_3$MRF), we give a multi modal description of the observed data. We approximate the 2-D histogram of the $\overline{g}(s) = [g_1(s), g_2(s)]^T$ vectors by a mixture of Gaussians distribution. In this way, one can measure which intensity values occur often together in the two images. Thereafter, the probability of the $\overline{g}(s)$ observation in the background is calculated as: $P(\overline{g}(s)| \text{ bg}) = \sum_{i=1}^{K}\kappa_i \cdot \eta(\overline{g}(s), \overline{\mu}_i, \Sigma_i)$, where $\eta(\cdot)$ denotes a two dimensional Gaussian density function with $\overline{\mu}_i$ mean vector and $\Sigma_i$ covariance matrix, while the $\kappa_i$ terms are positive weighting factors. Using a fixed $K = 5$, the distribution parameters are estimated automatically by the conventional EM algorithm. On the other hand, any $\overline{g}(s)$ value may occur in the changed regions, hence the 'ch' class is modeled by a uniform density: $P(\overline{g}(s)|\text{fg}) = u$. However, this multi-Gaussian intensity based approach (MGI) may erroneously mark several unaltered regions as changes compared: the miss-classifications would mainly be limited to highly textured regions (e.g. buildings and roads) since the $\overline{g}(s)$ gray values occurring there are less frequent in the global image statistics.

For obtaining the *second* feature, denoted by $c(s)$, we calculate the correlation between the rectangular $z \times z$ neighborhoods of $s$ in $G_1$ and in $G_2$ (used $v = 17$). Pixels with higher $c(s)$ values lie more likely in unchanged image regions. The experiments of the authors showed that the $P(c(s)|\text{bg})$ and $P(c(s)|\text{fg})$ probabilities can be approximated by different Gaussian distributions. Note that in itself, a simple Maximum Likelihood (ML) classification based on $c(\cdot)$ would also results in a fairly poor segmentation, but the $\overline{g}(s)$ and $c(s)$ are efficient complementary features. In low contrasted image regions, where the noisy $c(s)$ may be irrelevant, the decision based on $\overline{g}(s)$ is reliable. In textured areas one should choose $c(s)$ instead of $\overline{g}(s)$.

As the previous paragraphs suggest, we may estimate the reliability of the segmentation based on the $\bar{g}(s)$ intensity respectively $c(s)$ correlation features at each pixel $s$. Let $v_i(s)$ be the variance of the gray levels in the neighborhood of $s$, and let be $\bar{v}(s) = [v_1(s), v_2(s)]^T$. For implementing a probabilistic feature selection process, we approximate the $\bar{v}(s)$ variances of the *low* and *high* contrasted image regions by 2-D Gaussian density functions, where the parameters are estimated with an iterative algorithm presented in Benedek and Szirányi (2009).

### 4.2. Multilayer Mixed Markov model

As mentioned in the introduction, Mixed Markov models (Fridman, 2003) extend the modeling capabilities of Markov random fields: they enable using both static and observation-dependent dynamic links between the processing nodes. We can take here the advantage of this property, since the $\bar{v}(s)$ feature plays a particular role: it may locally switch ON and OFF the $\bar{g}(s)$ and $c(s)$ features into the integration procedure. We consider the task as a composition of four interactive segmentation processes. Thus we project the problem to a graph $\mathcal{G}$ with four layers: $S^g$, $S^c$, $S^v$ and $S^*$. We assign to each pixel $s \in S$ a unique graph node in each layer: e.g. $s^g$ is the node corresponding to pixel $s$ on the layer $S^g$. Denote $s^c \in S^c$, $s^v \in S^v$ and $s^* \in S^*$ similarly.

Following the approach of Section 3.2, we introduce next the labeling random process, which assigns a label $\omega(\cdot)$ to all nodes of $\mathcal{G}$. As usual, graph edges express direct dependencies between the corresponding node labels. The present approach exploits that Mixed Markov models distinguish two types of processing units, called *regular* and *address* nodes (Fridman, 2003). The $S^g$, $S^c$, and $S^*$ layers contain *regular nodes*, where the label denotes a possible fg/bg segmentation class: $\forall s \in S, i \in \{g, c, *\} : \omega(s^i) \in \{fg, bg\}$. For each $s$, $\omega(s^g)$ resp. $\omega(s^c)$ corresponds to the segmentation based on the $\bar{g}(s)$ resp. $c(s)$ feature; while the labels at the $S^*$ layer present the final change mask.

On the other hand, the $S^v$ layer contains *address nodes*, where for $s^v \in S^v$ the label $\omega(s^v)$ is a pointer to a regular node of $\mathcal{G}$. In contrast with static edges, address pointers represent dynamic connections between the nodes.

We use the following notations: $\tilde{\omega}(s^v) := \omega(\omega(s^v))$ is the label of the (*regular*) node addressed by $s^v$, and $\underline{\omega} = \{\omega(s^i)|s \in S, i \in \{g, c, v, *\}\}$ denotes a global labeling. Let $\mathcal{F} = \{\mathcal{F}_s|s \in S\}$ be the global observation, where $\mathcal{F}_s$ is the union of the $\bar{g}(s)$, $\bar{v}(s)$ and $c(s)$ local features extracted at pixel $s$. By definition of Mixed Markov models (Fridman, 2003), (static) edges may link any two nodes, and the a posteriori probability of a given global labeling $\underline{\omega}$ is given by:

$$P(\underline{\omega}|\mathcal{F}) = \alpha \prod_{C \in \mathcal{C}} \exp\left(-V_C(\omega_C, \omega_C^v, \mathcal{F})\right), \tag{16}$$

where $\mathcal{C}$ is the set of cliques in $\mathcal{G}$. For $C \in \mathcal{C}: \omega_C = \{\omega(q)|q \in C\}$ and $\omega_C^v = \{\tilde{\omega}(s^v)|s^v \in S^v \cap C\}$. $V_C$ is a $C \to \mathbb{R}$ *clique potential function*, which has a 'low' value if the labels within the set $\omega_C \cup \omega_C^v$ are semantically consistent, while $V_C$ is 'high' otherwise. Scalar $\alpha$ is a normalizing constant, which is independent of $\underline{\omega}$. Note that we will also use *singleton cliques* which contain single nodes.

Next, we define the cliques of $\mathcal{G}$ and the corresponding $V_C$ clique potential functions. The observations affect the model through the singleton potentials. As we stated previously, the labels in the $S^g$ and $S^c$ layers are directly influenced by the $\bar{g}(\cdot)$ respectively $c(\cdot)$ values, while the labels in $S^*$ have no direct links with these measurements. For this reason, let be $V_{\{s^g\}} = -\log P(\bar{g}(s)|\omega(s^g))$, $V_{\{s^c\}} =$

$-\log P(c(s)|\omega(s^c))$ and $V_{\{s^*\}} \equiv 0$. Note that the above distributions were already defined in Section 4.1, and $V_{\{s^v\}}$ will be later given.

For presenting smooth segmentations, we put connections within each layer among node pairs corresponding to neighboring pixels on the $S$ image lattice. Denote the set of the resulting *intra-layer* cliques by $C_2$. The prescribed potential function of a clique in $C_2$ will penalize neighboring nodes having different labels. Assuming $r$ and $s$ to be neighboring pixels on $S$, the potential of the doubleton clique $C_2 = \{r^i, s^i\} \in C_2$ for each $i \in \{g, c, v, *\}$ is calculated similarly to formula (13) from Section 3.2.

We continue with the description of the *inter-layer* interactions. Based on previous investigations, $\omega(s^*)$ should mostly be equal either to $\omega(s^g)$ or to $\omega(s^c)$, depending on the observed $v(s)$ feature. Hence, we put an edge among $s^*$ and $s^v$, and prescribe that $s^v$ should point either to $s^g$ or to $s^c$. As for the singleton potentials in the $S^v$ layer, if $s^v$ points to $s^\psi|_{\psi \in \{g,c\}}$, let be $V_{\{s^v\}} = -\log P(\bar{v}(s)|h_\psi)$. On the other hand, we get the potential of the *inter-layer clique* $C_3 = \{s^*, s^v\}$ with a fixed $\rho > 0$ as

$$V_{C_3}(\omega(s^*), \tilde{\omega}(s^v)) = \begin{cases} 0 & \text{if } \omega(s^*) = \tilde{\omega}(s^v) \\ \rho & \text{otherwise} \end{cases}$$

Finally, based on (16), the $\widehat{\underline{\omega}}$ maximum a posteriori estimate of the optimal global labeling, which maximizes $P(\widehat{\underline{\omega}}|\mathcal{F})$ (hence minimizes $-\log P(\widehat{\underline{\omega}}|\mathcal{F})$) can be obtained as:

$$\widehat{\underline{\omega}} = \underset{\underline{\omega} \in \Omega}{\arg\min} \sum_{s \in S;\ i} V_{\{s^i\}}\left(\omega(s^i), \mathcal{F}_s\right) + \sum_{\{s,r\} \in C_2;\ i} V_{C_2}\left(\omega(s^i), \omega(r^i)\right)$$
$$+ \sum_{s \in S} V_{C_3}\left(\omega(s^*), \tilde{\omega}(s^v)\right) \tag{17}$$

where $i \in \{g, c, v, *\}$ and $\Omega$ denotes the set of all the possible global labelings. The final segmentation is taken as the labeling of the $S^*$ layer.

## 5. Multitemporal image segmentation with the Fusion-MRF model

In the Fusion-MRF (FMRF) method (Szirányi and Shadaydeh, 2014), remote sensing areas of fused image series are examined in different levels of MRF segmentation; the goal is to automatically detect the category changes of the yearly transmuting areas having rich variations within a category by using more sample layers. The overlapping combination of category variations can be collected in a multilayer MRF segmentation; this supports the layer-by-layer MRF segmentation and change detection later. The definition of change is parallel to the definition of similarity; locations of image time series data that come from different sensors at different lighting and weather conditions can be compared if robust in-layer and cross-layer descriptors could be found. For this reason, in the proposed FMRF method, block-wise similarity measures is added to the stacking of the layers' pixel/microstructure information; the authors propose to use Cluster Reward Algorithm (*CRA*) (Inglada and Giros, 2004) in the multilayer fusion calculated between layer pairs in the series. The novelties of the FMRF approach are discussed in the following:

- Finding clusters on the stack of image-layers results in *aligned cluster-definition* for the different layers.
- Fused segmentation on the stack of image-layers, resulting in *multilayer labeling*.
- Multilayer labeling is used for the *unsupervised* clustering of the *single-layer* labeling; this aligned labeling makes the *change detection unequivocal*.
- A noise-tolerant *cross-layer similarity* measure, *CRA*, is used to better identify some classes where radiometric values are dubious.

## 5.1. Multitemporal image model and similarity feature

In a series of $N$ layers of remote sensing images, let $\overline{x}_s^{L_i}$ denote the feature vector at pixel $s$ of layer $L_i$, $i = 1, 2, \ldots, N$. This feature vector might contain color, texture/micro-structural features, cross layer similarity measures, or mixture of these. Set $X = \{\overline{x}_s | s \in S\}$ marks the global image data. An example of a feature vector would be

$$\overline{x}_s^{L_i} = [\overline{x}_{C(s)}^{L_i}, \overline{x}_{M(s)}^{L_i}]^T \tag{18}$$

where $\overline{x}_{C(s)}^{L_i}$ contains the pixel's color values, and $\overline{x}_{M(s)}^{L_i}$ is the cross layer similarity measures between the image and other two or more images in the series. The cross layer similarity measure might be correlation, mutual information, or *CRA*.

The multiple layers of remotely sensed images time series are characterized by the stack $\overline{x}_s^{L_{i_1 \ldots i_n}}$ of these vectors for a reasonable set of them, $n \leqslant N$:

$$\overline{x}_s^{L_{i_1 \ldots i_n}} = \{\overline{x}_s^{L_{i_1}}, \overline{x}_s^{L_{i_2}}, \ldots \overline{x}_s^{L_{i_n}}\} \tag{19}$$

Different similarity measures have been considered in the preliminary tests, such as distance to independence, mutual information, *CRA* (Inglada and Giros, 2004), Kullback Leibler divergence (see Alberga, 2009 and references therein). The *CRA* similarity measure is chosen as it gives better segmentation and change detection results than other similarity measures such as Kullback Leibler divergence and mutual information.

## 5.2. Fusion-MRF: multilayer segmentation and change detection

The segmentation and change detection procedure contains different levels of MRF optimization in the following main steps:

1. Selecting and registering the image layers; an example is shown in Inglada and Giros (2004). In case of professional data suppliers orthonormed and geographically registered images are given; no further registration is needed. In the discussed method no color-constancy or any shape/color semantic information is needed; the color of the corresponding areas and the texture can differ strongly layer-by-layer.
2. Finding clusters in the set of vectors ($\overline{x}_s^{L_{i_1 \ldots i_n}}$) and calculating the cluster parameters (mean and covariance of the conditional term in (22)) for the fusion based "*multilayer clusters*". This step can be performed either by using unsupervised methods such as the *K*-means algorithm, or by choosing the characteristic training areas manually.
3. Running MRF segmentation on the fused layer data ($\overline{x}_s^{L_{i_1 \ldots i_n}}$) containing the cross-layer measures, and the multilayer cluster parameters, resulting in a multilayer labeling $\Omega_{L_{i_1 \ldots i_n}}$;
4. Single-layer training: the map of multilayer labeling $\Omega_{L_{i_1 \ldots i_n}}$ is used as a training map for each image layer $L_i$: cluster parameters are calculated for each single layer controlled by the label map of multilayer clusters.
5. For each single layer $L_i$ (containing only its color and maybe texture features) a MRF segmentation is processed, resulting in a labeling: $\Omega_{L_i}$.
6. The consecutive image layers $(\ldots, (i-1), (i), \ldots)$ are compared to find the changes among the different label maps to get the $\delta_{i-1,i}$ change map:

$$\delta_{i-1,i}(\cdot) = \left[ (\Omega_{L_i}(\cdot) \neq \Omega_{L_{i-1}}(\cdot)) = TRUE \right] \tag{20}$$

In the proposed segmentation algorithm a multilayer MRF model is applied by contributing the term of the cross-layer *CRA* similarity measure calculated between each pair in a subset of three or more consecutive images. In what follows three consecutive images are used, however the algorithm can be easily extended to more layers. The stack of feature vectors $\overline{x}_s^{L_{1 \ldots 3}}$ is generated as follows

1. For each pair of the three consecutive images $L_i$, $L_{i+1}$ and $L_{i+2}$, the *CRA* image is calculated. In the calculation of the *CRA* image at each pixel, we use $D \times D$-pixel estimation window around this pixel to calculate the local histograms; Let the obtained CRA images be $CRA(i, i+1)$, $CRA(i+1, i+2)$, and $CRA(i, i+2)$.
2. Let $x_s^{L_i}$ denote the luminance value of pixel $s$ in image $L_i$. Construct the stack of feature vectors for pixels $s$ in the three images $L_i$, $L_{i+1}$ and $L_{i+2}$ as follows:

$$\overline{x}_s^{L_{i,i+1,i+2}} = [x_s^{L_i} + \alpha CRA_s(i, i+1), x_s^{L_{i+1}} + \alpha CRA_s(i+1, i+2), x_s^{L_{i+2}} + \alpha CRA_s(i, i+2)]^T \tag{21}$$

where $\alpha$ is a positive normalizing scalar ensuring the same range of the two different terms.

Note that the use of the addition of $x_s^{L_i}$ and $CRA_s(i, i+1)$ in the feature vector as given in (21) means lower dimensionality than using these features as two separate values as in (18). However, with the assumption that $x_s^{L_i}$ and $CRA_s(i, i+1)$ are statistically independent, it can be verified that they will contribute similar terms to the energy of MRF as when they are used as two separate features.

Next, we define the MRF energy model. Let $S = \{s_1, s_2, \ldots, s_H\}$ denote the image pixels, and $\underline{\omega} = \{\omega = (\omega_{s_1}, \ldots, \omega_{s_H}) : \omega_{s_i} \in \Lambda, 1 \leqslant i \leqslant H\}$ be the set of all possible labels assigned to the image classes.

The output labeling is taken again as the MAP estimator of the following energy function:

$$\widehat{\underline{\omega}} = \underset{\underline{\omega} \in \Omega}{\operatorname{argmin}} \left[ \sum_{s \in S} -\log P(\overline{x}_s | \omega(s)) + \sum_{r,s \in S} V_{C_2}(\omega(r), \omega(s)) \right] \tag{22}$$

where similarly to the intra-layer interactions in the L₃MRF and CXM methods, the Potts constraint (Potts, 1952) is used again to obtain smooth connected regions in the segmentation map (see formula (13) from Section 3.2).

## 6. MRF optimization

There is a large variety of MRF optimization algorithms used by the different change detection models, such as the Iterated Conditional Modes (ICM) (Besag, 1986) algorithm, or various simulated annealing techniques. Graph cut based minimization techniques have particularly become popular in the last ten years, since unlike the above iterative approaches, they guarantee to provide the global minimum of the energy in polynomial time. However the energy function must fulfill a number of specific requirements which must be considered during the modeling phase, limiting the possible model structures.

The Multicue MRF segmentation model presented in Section 3.2 provides a binary labeling, where a label $\omega(s^i)$, $s \in S$ and $i \in \{h, c, g\}$, can be regarded as a binary variable taking values 0 and 1. Moreover, the Gibbs energy given in (15) is composed of singleton and doubleton potentials only, i.e. the model has unary and pairwise interactions only. Therefore the Gibbs energy in (15) can be represented as a weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ (Kolmogorov and Zabih, 2004). The set of vertices's $\mathcal{V}$ in the graph $\mathcal{G}$ consists of all the sites of the three-layer MRF model and two terminal nodes: the *source* **s** and the *sink* **t**. The edges $\mathcal{E}$ are present between interacting sites as defined by the MRF model and special edges are connecting each

site with the terminal nodes **s** and **t** (for more details see (Kolmogorov and Zabih, 2004)). Estimating the minimal energy configuration is equivalent to compute the min cut/max flow on the corresponding graph $\mathcal{G}$. The energy function $E(\underline{\omega})$ is composed of unary and pairwise interactions (Kolmogorov and Zabih, 2004):

$$E(\underline{\omega}) = \sum_{s \in S} E_s(\omega(s^i)) + \sum_{(s,r) \in C - \{C_1\}} E_{s,r}(\omega(s^i), \omega(r^j)) \qquad (23)$$

where $i, j \in \{h, c, g\}$; $i = j$ for intra-layer cliques and inter-layer clique's otherwise. The above equivalence between energy minimization of $E$ and the min cut/max flow on the graph is only true, if $E$ also satisfies the *submodularity* constraint. Hence the authors used a standard graph-cut algorithm implemented by Vladimir Kolmogorov (http://pub.ist.ac.at/vnk/software.html) to minimize the energy in (15).

The energy term of (16) in the CXM model can be minimized by conventional iterative techniques, like ICM or simulated annealing (Geman and Geman, 1984). For choosing a good compromise between the quality factor and processing speed, the CXM model adapted the deterministic Modified Metropolis relaxation algorithm (Kato and Zerubia, 2012) to the multilayer model structure, as detailed in Benedek and Szirányi (2009). 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. As the authors noted, due to the fully modular structure of the energy term, the introduced model could be completed straightforwardly with additional sensor information (e.g. color or infrared sensors) or task-specific features depending on availability, in exchange that the iterative optimization takes usually longer time than the graph-cut optimization.

Similarly to the L$_3$MRF model, the FMRF energy function can be optimized by graph cut based techniques. The authors in Szirányi and Shadaydeh (2014) adopted the $\alpha$-expansion algorithm for MRF energy minimization, using the implementation of Szeliski et al. (2006). This relaxation technique ensures the convergence to a solution, where the energy is guaranteed to be under a boundary defined by the global minimal solution multiplied by a constant factor.

## 7. Parameter settings in the different model

The L$_3$MRF and CXM models follow a *supervised* approach of change detection. Hence all the parameters are estimated using the training data provided in each of the three data sets. Regarding the data term parameters, different feature distributions for the changed and unchanged regions should be estimated based on the empirical feature histograms obtained from the labeled training data. The standard Expectation–Maximization (EM) algorithm was applied in both models, however in CXM the EM steps were embedded into an iterative framework due mutual dependency between the parameters (Benedek and Szirányi, 2009).

In the FMRF, calculation of the cluster parameters for the fusion based multilayer clusters can be performed either by using unsupervised methods such as the *K*-means algorithm, or by supervised training by choosing the characteristic training areas manually. In all experiments presented here the unsupervised *K*-means algorithm was used.

### 7.1. L$_3$MRF parameters

The following parameters are used in the proposed model which need to be estimated:

- Parameters for Modified HOG feature selection: $\{dw, nb\}$.
- Parameters of the various pdf's as introduced in (3), (4), (6), (7): $\{a, b, c, u, v, w, a_d, b_d, a_h, b_h\}$.

- Parameters of the intra-layer and inter-layer clique potential functions: $\{\mathcal{K}^i, \rho^{hc}, \rho^{cg}\}$ where $i \in \{h, c, g\}$.

The initial parameters for the HOG feature selection namely the detection window size **dw** and the bin size **nb** were set by evaluating the maximum-likelihood results. By experimentation the desired results were obtained by setting the detection window to $11 \times 11$ and the number of bins to 9 (Singh et al., 2014).

Parameters for the Generalized Gamma Distribution (in (3), (6)) corresponding to the background class (both for the gray level and the Modified HOG features) were learned from the training data provided in each of the data sets. The threshold values for the uniform distribution (in (4), (7)) corresponding to the foreground class for both the gray level and the HOG features were set to the optimal values again for each of the training data belonging to its respective data set.

The intra- and inter-layer clique potentials are independent of the input images hence we set their values by trial and error over a small training dataset and then keep these values for all input image pairs.

### 7.2. CXM parameters

The CXM segmentation model has the following parameters:

- Preliminary parameters of feature calculation (see Section 4.1): $\Lambda = \{K, z\}$.
- Parameters of the probability density functions introduced in Section 4.1: $\Theta = \{\overline{\mu}_i, \Sigma_i | i \in 1 \ldots K\} \cup \{\mu_c, \sigma_c, \overline{\mu}_\psi, \Sigma_\psi\}$.
- Parameters of the intra- and inter-layer potential functions: $\Phi = \left\{\rho, \mathcal{K}^i : i \in \{g, c, *, v\}\right\}$.

The first step is determining the $\Lambda$ preliminary parameters. $K$, which is the number of the Gaussian mixture components in the background's intensity model, was set by a quantitative analysis. We considered the distribution sequence $P^K(\overline{g}(s)|bg)$ for $K = 1, 2, \ldots$ where $P^K$ is obtained by EM estimation from the training data using $K$ mixture components. Thereafter we measured the Bhattacharyya distances between the empirical histogram and the approximated distributions. As Fig. 4 demonstrates in the Szada set, the minimal error has been observed at $K = 5$, but the variance of the measured error rates was under 10% for the different $K$ components between 2 and 10.

On the other hand the size of the correlation window $z$ particularly depends on the image resolution and textureness. Since in the considered images 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
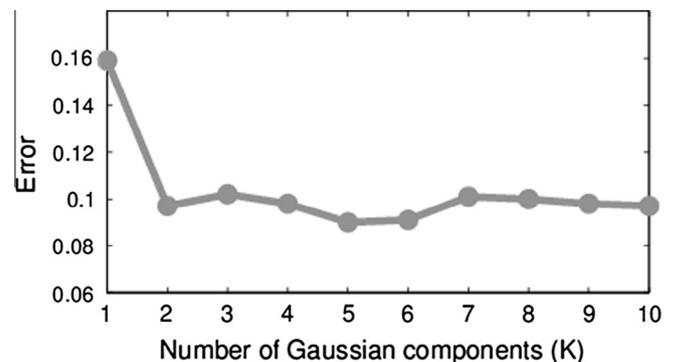


**Fig. 4.** CXM parameter analysis example. Error of the mixture of Gaussians approximation of the joint intensity statistics as a function of the number of mixture components ($K$).

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 $\Lambda$, the $\Theta$ parameters can be obtained automatically from the training image pairs using conventional estimators embedded in an iterative framework as detailed in Benedek and Szirányi (2009). As for the $\Phi$ parameter set, the experiments indicated that the model is not sensitive to a particular setting within a wide range, and these parameters can be estimated a priori. We used $\rho = \mathcal{K}^i = 1$, $i \in \{g, c, *, v\}$.

### 7.3. FMRF parameters

The following parameters are used in the FMRF model.

- Number of image layers $N$ and number of classes $H$.
- Window size $D$ used in the calculation of the CRA similarity measure.
- The homogeneity weight $\mathcal{K}$ used in the Potts term of the energy function.

For the present experiments using three image layers in the comparison gives good results. Using more layers requires dimensionality reduction or larger training areas that assure the presence of sufficient independent samples. Moreover, in such case, the number of possible CRA image combinations is larger than the number of layers. The used CRA images can be selected on the basis of maximal cross-layer information complexity. The number of classes used is based on the change detection application.

In CRA similarity measure calculation, the choice of the window size D, used in the estimation of local histograms, is selected based on the resolution of the images and the scale of the desired change detection. The size of the window should be less than the size of the change to be detected. The detection of small changes requires small window size; however larger window size gives better estimation of the CRA similarity measures. In our experiments, a minimum window size of $5 \times 5$ window is required regardless of image resolution to obtain good estimate of the local histograms used in CRA similarity measure calculation. Finding the optimum window size for each point adaptively, along with the definition of the scale of change detection and the image resolution, needs further research. In an attempt to deal with this problem Shadaydeh and Szirányi (2014) proposed new method for improved similarity measure estimation using weighted local histograms. The weight assigned to each pixel in the histogram estimation window follows an exponential function of its distance from the center of the window and the corresponding pixel value in an initial change map image which is derived from other microstructure or radiometric information. This improved similarity measure benefits from the good detection ability of small estimation window and the good estimation accuracy of large estimation window; hence it can replace the time-consuming multi-scale selection approaches for statistics based similarity measures in remote sensing.

The homogeneity weight $\mathcal{K}$ was again set by trial and error, similarly to the L$_3$MRF model.

## 8. Method comparison

### 8.1. Test databases and experimental conditions

For comparative evaluation of the L$_3$MRF, CXM and FMRF techniques, we used mainly the SZTAKI AirChange Benchmark Set.[1]

---

[1] http://web.eee.sztaki.hu/remotesensing/airchange_benchmark.html.

This benchmark has already been used in Benedek and Szirányi (2009) for an extensive validation of the CXM model, which was compared to the then-state-of-the-art, and it serves well the comparison between CXM and L$_3$MRF. The benchmark contains images from three data sets of optical aerial images provided by the Hungarian Institute of Geodesy Cartography and Remote Sensing (FÖMI) and Google Earth and corresponding Ground Truth change masks. All the images have been aligned as orthophotos.

The *first* data set – called Szada – contains co-registered images taken by FÖMI in years 2000 and 2005. This test set consists of *seven* – also manually evaluated – image pairs, covering in aggregate 9.5 km$^2$ area at 1.5 m/pixel resolution (the size of each image in the test set is $952 \times 640$ pixels). The *second* test set – called Tiszadob – includes *five* image pairs from 2000 and 2007 (6.8 km$^2$) with similar size and quality parameters to Szada. Images of 2000 and 2005 were scanned on photo-films (Hasselblad 500 EL/M) before digital scan. The 2007 image has been originally scanned in digital form (Nikon D3X, with AF-S Nikkor 50 mm 1.4G lens). Finally, in the *third* data set – called Archive – an aerial image taken by FÖMI in 1984 can be compared to a corresponding Google Earth image from around 2007. The latter case is highly challenging, since the image from 1984 has a degraded quality, and several major differences appear due to the 23 years time difference between the two shots. An important additional information is that the image pair from the Archive set has been taken from the same area as the first image pair of the Szada set, therefore in this region four co-registered images available with different time stamps (three of them are partially shown in Fig. 6). The latter fact can be utilized by the Fusion-MRF (FMRF) model (Szirányi and Shadaydeh, 2014), which may consider several images to obtain a robust change mask between two selected time layers.

Considering the above conditions, during the experiments with the FMRF model, three image layers were used on the Szada data set by extracting changes between years 2000 and 2005. The authors used as third (*auxiliary*) image the Google Earth photo of Archive from 2007. As the input of the PCC approach, two segmentation classes were used: one for urban and one for non-urban areas. The feature vector used for the multilayer segmentation contained luminance value and the CRA similarity measure as defined in (21), while the CIE L*a*b* color values were only used in the single layer segmentation step. For change detection between years 1984 and 2007 (experiment Archive), the Szada 2005 image was used as third image. Similarly to the previous case, urban and non-urban regions have been distinguished during the segmentation and the used feature vector included luminance and CRA similarity. For single layer segmentation, luminance values and texture features (output energy of Laws filters) were exploited. All the used features were normalized to have maximum value of 1. Finally a *third* experiment has been conducted in a fully natural (non-urban) region, called Forest (see Fig. 7), which was not included in the AirChange Benchmark. From the selected area, three high resolution images (around 0.5 m/pixel) were available, but only two images were used in the final validation test, as the experiments did not report further improvement by including the third image. For the segmentation, three classes were defined here: meadow, planted meadow and forest. The feature vector contained the luminance and CRA similarity as usual. In single layer segmentation, no color components were used except luminance in addition to the *Laws* features.

### 8.2. Ground Truth generation

Relevant Ground Truth (GT) generation is a difficult and often controversial issue in change detection evaluation. For this reason, we will discuss two different GT approaches in this paper: one for

*Direct* techniques, and one for *Post-Classification Comparison (PCC)* methods, with also showing experiments for cross-validation.

In *Direct methods* the users often define from the application point of view the relevant types of changes, and they directly label the changed regions. The creators of the *AirChange Benchmark Set* followed this approach, and they considered three main prototypes of changes, as displayed in Fig. 5: (a) new built-up regions and building operations, (b) planting of large group of trees of forestation, (c) fresh plough-land or groundworks before building over. Note that this *AirChange GT* does NOT contain change classification, only binary changed-unchanged decision for each pixel.

On the other hand, in *PCC methods*, where the results of different image segmentations are compared, the GT should be generated for the classification step at each time layer, thereafter the change GT can be automatically derived as taking the image regions with altered class labels. This evaluation approach is followed in the *Region PCC GT* generation process of Szirányi and Shadaydeh (2014), where different segmentation classes have been considered for the different image pairs (see Section 8.1), such as urban and non-urban for Szada and Archive; meadow, planted meadow and forest for the Forest image pair.

## 8.3. Results and evaluation

A few qualitative change mask examples obtained by the discussed multilayer Markovian models are shown in Figs. 6 and 7 for selected image regions from the Szada and Forest data sets respectively. We provide the numerical validation results in two parts. First, we compare the two *direct* methods: L₃MRF and CXM. In addition, to demonstrate the efficiency of multilayer models within the state-of-the-art, we also match the obtained results to four previous methods from the literature which focus on similar goals but use different approaches of the problems. Second, we compare the FMRF method (as a PCC technique) to the CXM model, and discuss the experiments.

### 8.3.1. Comparison of direct models

The quantitative evaluation in this paper is primarily based on the *AirChange GT* set, which has already been utilized in Benedek and Szirányi (2009) for comparing the CXM technique to *four* different *direct* change detection models from the literature (Wiemker, 1997; Bruzzone and Fernandez-Prieto, 2002; Ghosh et al., 2007; Castellana et al., 2007), and also in Singh et al. (2014) where some comparative tests have been provided between the CXM and L₃MRF methods. Due to the limitations of the WAFO toolbox (Brodtkorb et al., 2000), the L₃MRF method could have only been tested in a subset of the *AirChange Benchmark*, which contains 5 image pairs from Szada, 2 image pairs from Tiszadob and an image pair from the Archive set. To keep here all validation figures relevant, we limit our forthcoming surveys to this part of the Benchmark, and refer to further results in Benedek and Szirányi (2009). Since both methods are supervised, we used around 20% of the available GT annotated image regions as training data.

The practical differences in the results originate from the differences in the used features and the way of integration. A limitation of standard pixel value differencing is that due to seasonal changes or altered illumination, the observed gray levels may be significantly different even in the corresponding unchanged territories. While using the gray level difference (GLD) feature of L₃MRF the regions are always marked as change whenever the gray levels are not close enough to each other, the intensity co-occurrence statistics (ICS) feature of CXM may correctly classify large homogeneous regions, even if the observed pixel value changes are high. However, ICS is quite sensitive to the calibration of the uniform distribution parameters. Note as well that ICS interprets change as *unusual events* (i.e. rare pixel value occurrences) rather than as *difference*, which assumption is only correct in certain applications. For example, as Fig. 8 shows, the hardly visible fresh plough land has been indicated as a region change with CXM, but not with L₃MRF. In addition, by using ICS a lot of changes in homogeneous regions can mislead the statistical color assignment or if the intensity values of fine but low contrasted structures (such as blurred road or building regions) are identical, they may be marked erroneously as outliers. Since the HOG feature exhibits invariance to geometric and photometric transformations (except for object orientation), the L₃MRF method can handle registration errors and parallax more robustly than correlation. From a structural point of view, in CXM the noise of the variance meta-feature may mean a bottleneck, as it can corrupt the relevant feature selection step (e.g. low contrasted roads, buildings). On the other hand, in L₃MRF both features are simultaneously considered during the integration step.

During the numerical tests, we used first the same metrics as e.g. in Ghosh et al. (2007), where in the change masks, the pixel level false alarms (F-A, number of unchanged pixels which were detected as changes) and missed alarms (M-A, erroneously ignored changed pixels) are counted w.r.t. the GT, then an overall error (O-E) is calculated simply as the sum of the previous two quantities. Numerical results for L₃MRF, CXM and the above four reference methods are listed in Table 2, where the F-A, M-A and O-E rates are given in percent of the checked pixels. We can observe that both multilayer models, (CXM and L₃MRF) outperform the other references and the L₃MRF shows a slight advantage vs. CXM w.r.t. this metrics. We can also notice based on Fig. 6 that the L₃MRF approach produces more homogeneous and smoothing effects than the CXM model.

A limitation of the above evaluation process is that the F-A, M-A and O-E rates do not consider the quantity of changes in the images. If the area of changed regions is very small compared to the image size, even a weak classifier may yield low O-E values. A possible alternative solution is to calculate the Recall (Rc) and Precision (Pr) rates from the point of view of the change class (Benedek et al., 2009), and calculate the F-measure rate (Fr) as the harmonic mean of Pr and Rc. In Table 3 we have also provided the measured Precision–Recall–F-rate triplets regarding the Szada test set. We can see here some advantage of the CXM method due to similar *Precision* but higher *Recall* rates. This feature may



(a) New built-in regions or building operations      (b) Planting a group of trees or forestation      (c) Fresh plough-lands or groundworks
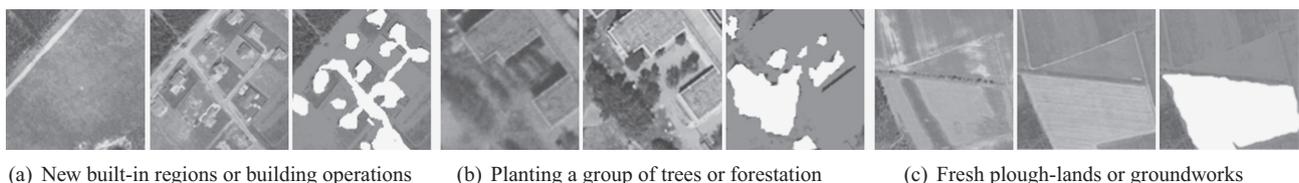
**Fig. 5.** Change prototypes considered for Ground Truth (GT) generation in the SZTAKI AirChange Benchmark Set. Each image triplet displays the first and second images and the highlighted GT changes.

mean benefits in applications, where false alarms can be eliminated later by higher level (like object level) processing modules.

### 8.3.2. Comparison of the FMRF and the CXM models

Since the *AirChange GT* set is only partially relevant for the PCC-based FMRF method, we have not mixed the discussion about FMRF with the direct approaches in the previous subsection. Instead, we selected three relevant image pairs, and applied the quantitative evaluation both using the *AirChange GT* and the *Region PCC GT* masks. The measured Pr, Rc and Fr values are listed in Table 4, and two sample image parts are shown in Figs. 6 and 7. We can observe that for the Szada and Archive images, where FMRF used urban/non-urban image classification, both methods gave quite similar results, and they were less sensitive to the two different GT generation approaches. As also shown in Fig. 6(d) and (e) the two GT versions are not significantly different here, since most sources of the changes were the new urban developments which issues were focused by both GT models. On the other hand,



(a) Image from 2000: $G_1$ reference     (b) Image from 2005: $G_2$ reference     (c) Image from 2007 – only used by FMRF

(d) AirChange Ground Truth ($G_1$ vs. $G_2$)     (e) Region PCC Ground Truth ($G_1$ vs. $G_2$)

(f) L$_3$MRF result ($G_1$ vs. $G_2$)     (g) CXM result ($G_1$ vs. $G_2$)     (h) FMRF result ($G_1$ vs. $G_2$)
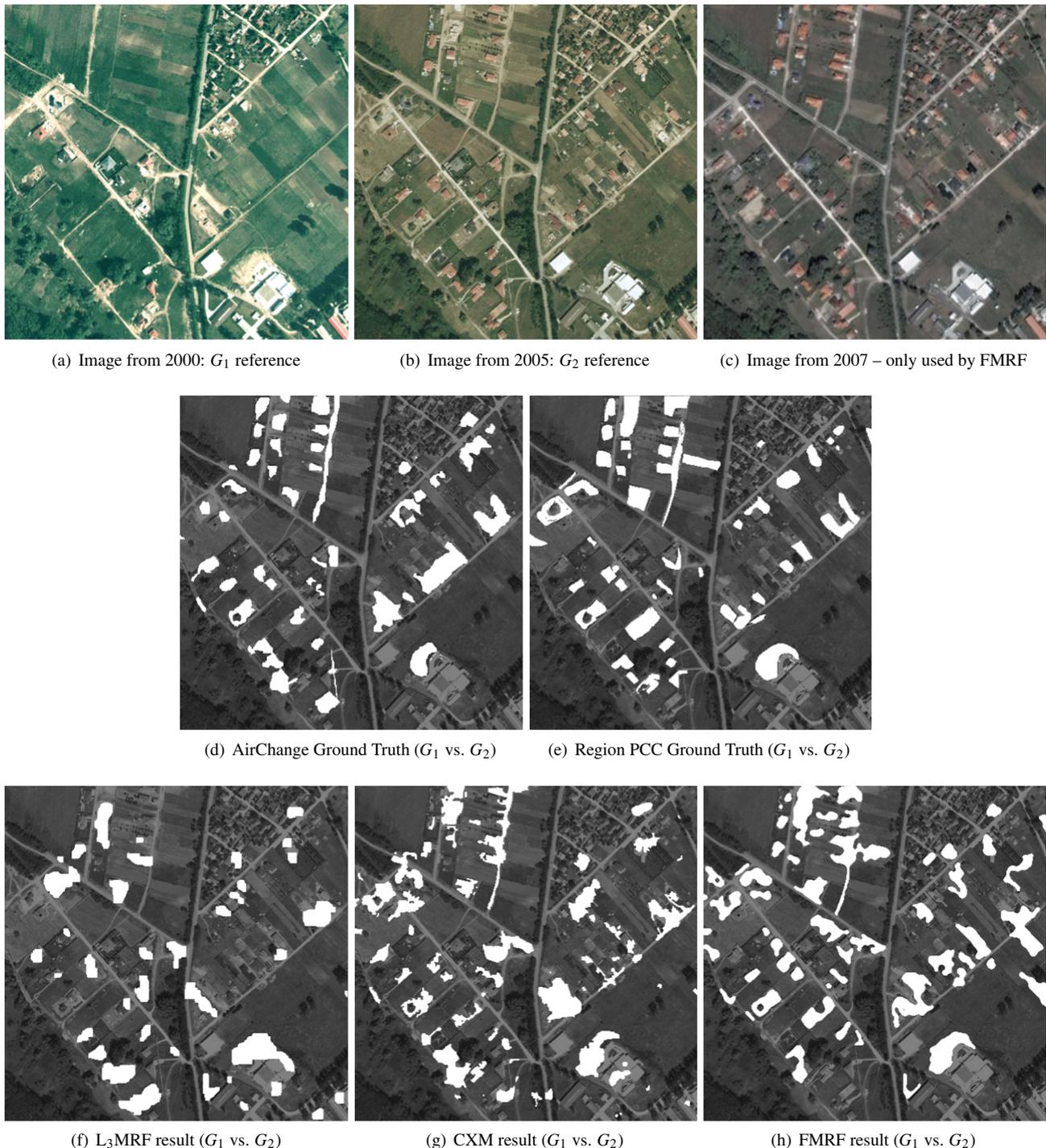
**Fig. 6.** Results by the three discussed multilayer change detection models for an image segment from the Szada data set (see (a) and (b)). Changes are displayed with white color. The FMRF method also used here a third auxiliary input image from 2007 shown in (c). (d) and (e) display the two different Ground Truth approaches, where the *Region PCC Ground Truth* has been generated based on urban/non-urban segmentation of the images. (f)–(h) images show the change detection results by the three methods. (For interpretation of the references to color in this figure legend, the Reader is referred to the web version of this article.)

(a) Image 1: $G_1$    (b) Image 2: $G_2$    (c) AirChange GT    (d) CXM result    (e) Region PCC GT    (f) FMRF result
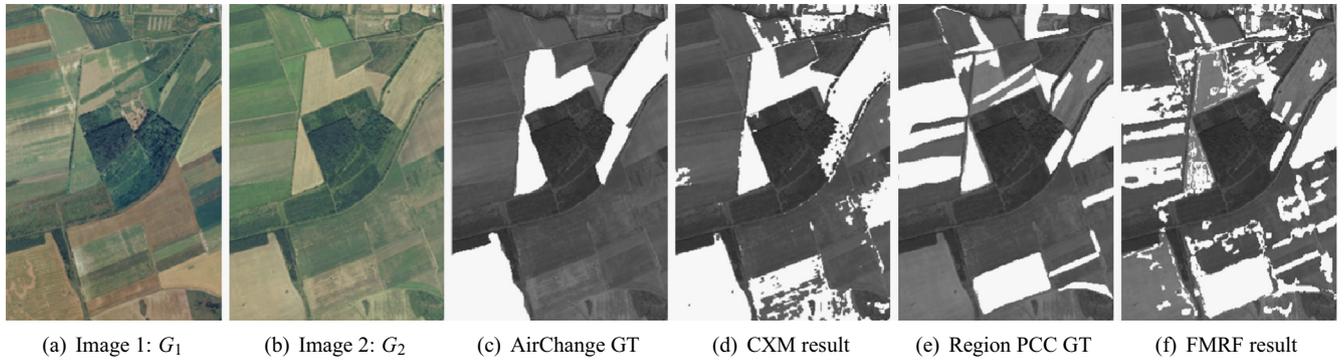
**Fig. 7.** Qualitative comparison of the CXM and FMRF methods and the two different GT types on the FOREST data sample.



(a) Image 1    (b) Image 2    (c) L$_3$MRF    (d) CXM
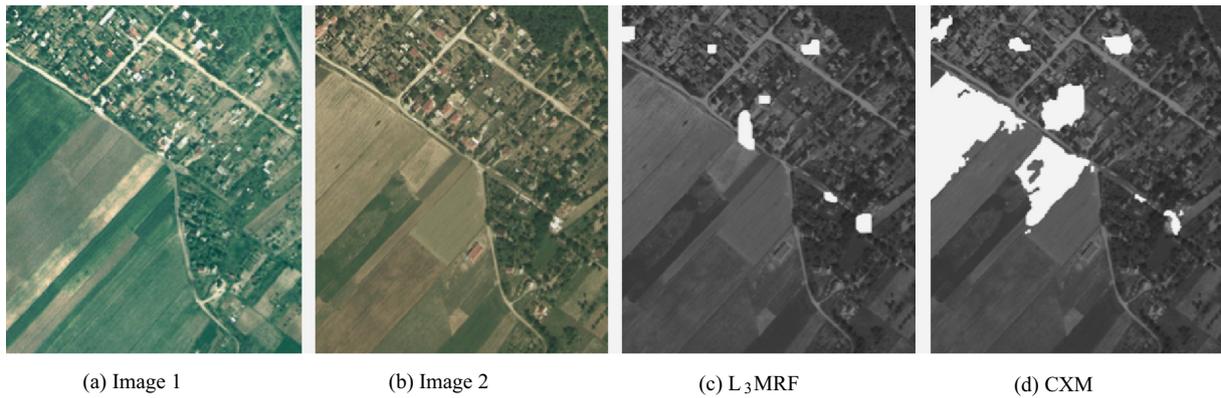
**Fig. 8.** Difference between the obtained change masks: the large change area of CXM corresponds to a hardly visible fresh plough land.

**Table 2**
Method evaluation: False-Alarm (F-A), Missed-Alarm (M-A) and Overall-Error (O-A) rates of the CXM and L$_3$MRF methods, and further reference techniques (best values are typeset with bold). All error rates are given in percentage of the number of processed image pixels.

| Method | SZADA data set | | | TISZADOB data set | | | ARCHIVE data set | | |
|---|---|---|---|---|---|---|---|---|---|
| | F-A | M-A | O-E | F-A | M-A | O-E | F-A | M-A | O-E |
| PCA (Wiemker, 1997) | 4.51 | 1.70 | 6.21 | 3.26 | 3.39 | 6.65 | 2.30 | 6.76 | 9.06 |
| Hopfield (Ghosh et al., 2007) | 4.32 | 1.98 | 6.30 | 4.17 | 3.16 | 7.33 | 8.26 | 5.86 | 14.12 |
| Parzen (Bruzzone and Fernandez-Prieto, 2002) | 4.44 | 1.90 | 6.34 | 4.82 | 3.37 | 8.19 | 4.69 | 7.18 | 11.87 |
| MLP (Castellana et al., 2007) | 3.55 | 2.00 | 5.56 | 2.65 | 3.72 | 6.38 | 5.53 | 5.96 | 11.48 |
| **CXM** (Benedek and Szirányi, 2009) | 2.87 | **1.32** | 4.19 | 2.91 | **1.77** | 4.68 | 7.60 | **3.06** | 10.66 |
| **L$_3$MRF** (Singh et al., 2014) | 1.11 | 2.33 | **3.44** | **0.06** | 3.90 | **3.96** | **2.23** | 6.43 | **8.66** |

**Table 3**
Method evaluation: *Precision*, *Recall* and *F-measure* rates in percent w.r.t. the 'change' class in the SZADA set (higher values are preferred). The discussed multilayer models and the best values are typeset with bold.

| Method | Precision | Recall | F-rate |
|---|---|---|---|
| PCA | 21.7 | 42.4 | 28.7 |
| Hopfield | 18.3 | 32.8 | 23.5 |
| Parzen | 19.1 | 35.5 | 24.8 |
| MLP | 21.1 | 32.2 | 25.5 |
| **CXM** | **36.2** | **55.3** | **43.8** |
| **L$_3$MRF** | 36.0 | 21.1 | 26.6 |

regarding the FOREST (Fig. 7) images, which contain unpopulated regions only, the sensitivity of evaluation against the chosen Ground Truth type has strongly increased, i.e. the CXM showed only efficient performance figures with the *AirChange GT*, while the FMRF proved to be significantly better by the comparison to the *Region PCC GT*. Here we can also observe large differences between the GT images (Fig. 7(c) and (e)): the *AirChange GT* mask only highlighted the forest cut and the fresh plough land areas; the plough lands yielded large homogeneous image regions with a characteristic gray level, appearing outlier in the joint intensity domain. On the other hand, the *Region PCC GT* indicated all differences between unplanted meadows, planted meadows, and forest regions.

### 8.4. Computational time

Although MRF based methods were regarded as computationally expensive techniques in the past, recent improvements in hardware platforms and optimization techniques, such as the introduction of the deterministic MMD algorithm or Graph cut based energy minimization, enabled the wide usage of these models in many practical tasks. Among the discussed three models the L$_3$MRF shows the most efficient computational performance, as processing a sample image pair from the AirChange Benchmark Set needs around 10–15 s with C++ implementation in a standard laptop. Under similar circumstances, the CXM performs the task

**Table 4**
Comparative evaluation between the CXM and the FMRF multilayer models, Precision (Pr), Recall (Rc) and F-measure rates (Fr) are displayed (better values are typeset with bold). *AirChange GT* and *Region PCC GT* mark two different Ground Truth masks as described in Section 8.2.

| Image | AirChange GT | | | | | | Region PCC GT | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | CXM | | | FMRF | | | CXM | | | FMRF | | |
| | Pr | Rc | Fr | Pr | Rc | Fr | Pr | Rc | Fr | Pr | Rc | Fr |
| Szada/1 | **36.5** | **58.4** | **44.9** | 32.6 | 54.3 | 40.8 | 34.5 | 50.3 | 40.9 | **45.6** | **69.0** | **54.9** |
| Archive | **47.0** | 62.7 | **53.7** | 27.9 | **71.2** | 40.1 | 51.4 | 32.7 | 40.0 | **59.1** | **72.2** | **65.0** |
| Forest | **61.7** | **93.4** | **74.3** | 13.4 | 18.5 | 15.6 | 31.3 | 39.1 | 34.8 | **61.5** | **69.6** | **65.3** |

within 25–30 s on a computer with an Intel(R) Core(TM) i7 3.20 GHz CPU: the feature extraction takes around 5 s, and the remaining time in necessary for the simulated annealing based MRF energy optimization. The FMRF – which was originally suited to higher resolution inputs – works with Matlab implementation for around 2 min in an image pair, using a PC with 2.67 GHz CPU. Here almost 80% of the CPU time is related to the CRA similarity image computation step. Since in the introduced change detection examples, we compared images with several years time differences, we can assume that in most practical scenarios the problem can be solved offline, therefore the computational requirements of any of the demonstrated models should not mean bottleneck for the workflow.

## 9. Conclusion

In this paper we gave a comparative survey on three different multilayer MRF solutions, whose main properties are summarized in Table 1. We have observed that with appropriate feature selection and parametrization we may obtain quite similarly efficient change masks with all the three models, and the results are also competitive with other MRF solutions proposed earlier in the literature. The differences should be often evaluated based on the exact needs of the focused applications: PCC methods are preferred if the changes can be interpreted at the level of well defined segmentation classes, while direct approaches can be trained simply by a batch of positive and negative *change* examples. Since multilayer MRFs offer flexible model structures, all the three models can be completed with additional features or accommodated to different types of remote sensing data. The paper has shown that although the definition of image clusters and the focused change prototypes are critical tasks in the detection, the different approaches may lead to different, but effective models with established mathematical background.

## References

Alberga, V., 2009. Similarity measures of remotely sensed multi-sensor images for change detection applications. Remote Sens. 1 (3), 122–143.

Baselice, F., Ferraioli, G., Pascazio, V., 2014. Markovian change detection of urban areas using very high resolution complex SAR images. IEEE Geosci. Remote Sens. Lett. 11 (5), 995–999.

Bazi, Y., Bruzzone, L., Melgani, F., 2005a. An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images. IEEE Trans. Geosci. Remote Sens. 43 (4), 874–887.

Bazi, Y., Bruzzone, L., Melgani, F., 2005b. An unsupervised approach based on the generalized gaussian model to automatic change detection in multitemporal SAR images. IEEE Trans. Geosci. Remote Sens. 43 (4), 874–887.

Benedek, C., Szirányi, T., 2007. Markovian framework for structural change detection with application on detecting built-in changes in airborne images. In: Int. Conf. on Signal Processing, Pattern Recognition and Applications. ACTA, Innsbruck, Austria, pp. 68–73.

Benedek, C., Szirányi, T., 2009. Change detection in optical aerial images by a multi-layer conditional mixed Markov model. IEEE Trans. Geosci. Remote Sens. 47 (10), 3416–3430.

Benedek, C., Szirányi, T., Kato, Z., Zerubia, J., 2009. Detection of object motion regions in aerial image pairs with a multi-layer Markovian model. IEEE Trans. Image Process. 18 (10), 2303–2315.

Benedek, C., Descombes, X., Zerubia, J., 2012. Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth–death dynamics. IEEE Trans. Pattern Anal. Mach. Intell. 34 (1), 33–50.

Besag, J., 1986. On the statistical analysis of dirty images. J. Roy. Stat. Soc. 48, 259–302.

Bovolo, F., Bruzzone, L., Marconcini, M., 2008. A novel approach to unsupervised change detection based on a semisupervised SVM and a similarity measure. IEEE Trans. Geosci. Remote Sens. 46 (7), 2070–2082.

Brodtkorb, P., Johannesson, P., Lindgren, G., Rychlik, I., Rydén, J., Sjö, E., 2000. WAFO – a Matlab toolbox for the analysis of random waves and loads. In: Int. Offshore and Polar Eng. Conf., vol. 3. Seattle, USA, pp. 343–350.

Bruzzone, L., Fernandez-Prieto, D., 2000. Automatic analysis of the difference image for unsupervised change detection. IEEE Trans. Geosci. Remote Sens. 38 (3), 1171–1182.

Bruzzone, L., Fernandez-Prieto, D., 2002. An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images. IEEE Trans. Image Process. 11 (4), 452–466.

Carincotte, C., Derrode, S., Bourennane, S., 2006. Unsupervised change detection on SAR images using fuzzy hidden Markov chains. IEEE Trans. Geosci. Remote Sens. 44 (2), 432–441.

Castellana, L., D'Addabbo, A., Pasquariello, G., 2007. A composed supervised/unsupervised approach to improve change detection from remote sensing. Pattern Recogn. Lett. 28 (4), 405–413.

Chatelain, F., Tourneret, J.-Y., Inglada, J., 2008. Change detection in multisensor SAR images using bivariate gamma distributions. IEEE Trans. Image Process. 17 (3), 249–258.

Chen, Y., Cao, Z., 2013. An improved MRF-based change detection approach for multitemporal remote sensing imagery. Signal Process. 93 (1), 163–175.

Chen, K., Zhou, Z., Lu, H., Huo, C., 2007. Change detection based on conditional random field models. In: WSEAS International Conference on Remote Sensing. Venice, Italy, pp. 93–97.

Clausi, D., Deng, H., 2005. Design-based texture feature fusion using Gabor filters and co-occurrence probabilities. IEEE Trans. Image Process. 14 (7), 925–936.

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1. San Diego, CA, USA, pp. 886–893.

Fernandez-Prieto, D., Marconcini, M., 2011. A novel partially supervised approach to targeted change detection. IEEE Trans. Geosci. Remote Sens. 49 (12), 5016–5038.

Fridman, A., 2003. Mixed Markov models. Proc. Natl. Acad. Sci. USA 100 (14), 8092–8096.

Fung, T., LeDrew, E., 1988. The determination of optimal threshold levels for change detection using various accuracy indices. Photogram. Eng. Remote Sens. 54, 1449–1454.

Gamba, P., Dell'Acqua, F., Lisini, G., 2006. Change detection of multitemporal SAR data in urban areas combining feature-based and pixel-based techniques. IEEE Trans. Geosci. Remote Sens. 44 (10), 2820–2827.

Geman, S., Geman, D., 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. IEEE Trans. Pattern Anal. Mach. Intell. 6 (6), 721–741.

Ghosh, S., Bruzzone, L., Patra, S., Bovolo, F., Ghosh, A., 2007. A context-sensitive technique for unsupervised change detection based on Hopfield-type neural networks. IEEE Trans. Geosci. Remote Sens. 45 (3), 778–789.

Ghosh, A., Subudhi, B., Bruzzone, L., 2013. Integration of Gibbs Markov Random Field and Hopfield-type neural networks for unsupervised change detection in remotely sensed multitemporal images. IEEE Trans. Image Process. 22 (8), 3087–3096.

Hoberg, T., Rottensteiner, F., Heipke, C., 2012. Context models for CRF-based classification of multitemporal remote sensing data. ISPRS Ann. Photogram. Remote Sens. Spatial Inform. Sci. I-7, 129–134.

Hoberg, T., Rottensteiner, F., Feitosa, R., Heipke, C., 2015. Conditional random fields for multitemporal and multiscale classification of optical satellite imagery. IEEE Trans. Geosci. Remote Sens. 53 (2), 659–673.

Hodge, V., Austin, J., 2004. A survey of outlier detection methodologies. Artif. Intell. Rev. 22 (2), 85–126.

Inglada, J., Giros, A., 2004. On the possibility of automatic multisensor image registration. IEEE Trans. Geosci. Remote Sens. 42, 2104–2120.

Jodoin, P., Mignotte, M., Rosenberger, C., 2007. Segmentation framework based on label field fusion. IEEE Trans. Image Process. 16 (10), 2535–2550.

Kalayeh, H., Landgrebe, D., 1986. Utilizing multitemporal data by a stochastic model. IEEE Trans. Geosci. Remote Sens. GE-24 (5), 792–795.

Kato, Z., Pong, T.C., 2006. A Markov random field image segmentation model for color textured images. Image Vision Comput. 24 (10), 1103–1114.

Kato, Z., Zerubia, J., 2012. Markov random fields in image segmentation. Foundations and Trends in Signal Processing. Now Publishers.

Kato, Z., Pong, T.C., Song, G.Q., 2002. Multicue MRF image segmentation: combining texture and color. In: International Conference on Pattern Recognition. Quebec, Canada, pp. 660–663.

Kim, J., Zabih, R., 2002. Factorial Markov random fields. In: European Conference on Computer Vision. Copenhagen, Denmark, pp. 321–334.

Kolmogorov, V., Zabih, R., 2004. What energy functions can be minimized via graph cuts? IEEE Trans. Pattern Anal. Mach. Intell. 26 (2), 147–159.

Kosugi, Y., Sakamoto, M., Fukunishi, M., Wei Lu Doihara, T., Kakumoto, S., 2004. Urban change detection related to earthquakes using an adaptive nonlinear mapping of high-resolution images. IEEE Geosci. Remote Sens. Lett. 1 (3), 152–156.

Kumar, S., Anouncia, M., Johnson, S., Agarwal, A., Dwivedi, P., 2012. Agriculture change detection model using remote sensing images and GIS: study area vellore. In: IEEE Int'l Conf. on Radar, Communication and Computing (ICRCC), pp. 54–57.

Liu, W., Prinet, V., 2006. Probabilistic modeling for structural change inference. In: Asian Conference on Computer Vision. Hyderabad, India, pp. 836–846.

Liu, D., Song, K., Townshend, J.R., Gong, P., 2008. Using local transition probability models in Markov random fields for forest change detection. Remote Sens. Environ. 112 (5), 2222–2231, earth Observations for Terrestrial Biodiversity and Ecosystems Special Issu.

Lorette, A., Descombes, X., Zerubia, J., 2000. Texture analysis through a Markovian modelling and fuzzy classification: application to urban area extraction from satellite images. Int. J. Comput. Vis. 36 (3), 221–236.

Martinis, S., Twele, A., Voigt, S., 2011. Unsupervised extraction of flood-induced backscatter changes in SAR data using Markov image modeling on irregular graphs. IEEE Trans. Geosci. Remote Sens. 49 (1), 251–263.

Melgani, F., Bazi, Y., 2006. Markovian fusion approach to robust unsupervised change detection in remotely sensed imagery. IEEE Geosci. Remote Sens. Lett. 3 (4), 457–461.

Melgani, F., Serpico, S., 2003. A Markov random field approach to spatio-temporal contextual image classification. IEEE Trans. Geosci. Remote Sens. 41 (11), 2478–2487.

Moser, G., Angiati, E., Serpico, S., 2011. Multiscale unsupervised change detection on optical images by Markov random fields and wavelets. IEEE Geosci. Remote Sens. Lett. 8 (4), 725–729.

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

Potts, R., 1952. Some generalized order-disorder transformation. In: Proceedings of the Cambridge Philosophical Society, No. 48, p. 106.

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

Reed, S., Tena Ruiz, I., Capus, C., Petillot, Y., 2006. The fusion of large scale classified side-scan sonar image mosaics. IEEE Trans. Image Process. 15 (7), 2049–2060.

Serpico, S., Moser, G., 2006. Weight parameter optimization by the Ho-Kashyap Algorithm in MRF models for supervised image classification. IEEE Trans. Geosci. Remote Sens. 44 (12), 3695–3705.

Shadaydeh, M., Szirányi, T., 2014. An improved local similarity measure estimation for change detection in remote sensing images. In: IEEE International Conference on Aerospace and Remote Sensing Technology. Yogyakarta, Indonesia, pp. 230–234.

Shuang, W., Leung, H., 2012. A Markov random field approach for sidescan sonar change detection. IEEE J. Ocean. Eng. 37 (4), 659–669.

Singh, P., Kato, Z., Zerubia, J., 2014. A multilayer markovian model for change detection in aerial image pairs with large time differences. In: International Conference on Pattern Recognition. Stockholm, Sweden.

Solberg, A., Taxt, T., Jain, A., 1996. A Markov random field model for classification of multisource satellite imagery. IEEE Trans. Geosci. Remote Sens. 34 (1), 100–113.

Subudhi, B.N., Bovolo, F., Ghosh, A., Bruzzone, L., 2014. Spatio-contextual fuzzy clustering with Markov random field model for change detection in remotely sensed images. Opt. Laser Technol. 57, 284–292, optical Image Processing.

Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., Rother, C., 2006. A comparative study of energy minimization methods for Markov Random Fields. In: European Conference on Computer Vision, vol. 2. Graz, Austria, pp. 16–29.

Szirányi, T., Shadaydeh, M., 2014. Segmentation of remote sensing images using similarity-measure-based fusion-MRF model. IEEE Geosci. Remote Sens. Lett. 11 (9), 1544–1548.

Wang, Y., Wu, Y., Zhang, Q., Zhang, P., Li, M., Lu, Y., 2013. Unsupervised change detection on SAR images using triplet Markov field model. IEEE Geosci. Remote Sens. Lett. 10 (4), 697–701.

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

Xu, R., Wunsch, I., 2005. Survey of clustering algorithms. IEEE Trans. Neural Netw. 16 (3), 645–678.

Xu, Q., Pu, Y., Wang, W., Zhong, H., 2012. Multispectral remote sensing image change detection based on Markovian fusion. In: International Conference on Agro-Geoinformatics. Shanghai, China.

Zhong, P., Wang, R., 2007. A multiple conditional random fields ensemble model for urban area detection in remote sensing optical images. IEEE Trans. Geosci. Remote Sens. 45 (12), 3978–3988.