

Segmentation of remote sensing images using similarity measure based fusion-MRF model

Tamas Sziranyi, *Senior Member, IEEE* and Maha Shadaydeh

Abstract—Classifying segments and detecting changes in terrestrial areas are important and time-consuming efforts for remote-sensing image analysis tasks, including comparison and retrieval in repositories containing multi-temporal remote image samples for the same area in very different quality and details. We propose a multi-layer fusion model for adaptive segmentation and change detection of optical remote sensing image series, where trajectory analysis or direct comparison is not applicable. Our method applies unsupervised or partly supervised clustering on a fused image series by using cross-layer similarity measure, followed by a multi-layer MRF segmentation. 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. The significant benefit of the proposed method has been numerically validated on remotely sensed image series with ground-truth data.

Index Terms—Image segmentation, Remote sensing, Fusion-MRF, Similarity measure, Change detection, Cluster Reward Algorithm

I. INTRODUCTION

EARTH observation based on aerial and satellite image series, including high-resolution remote sensing image time-series (RSITS) [1], results in a large data volume containing several known and maybe unknown details at different sampling-time rates. The sample series (of large image repositories) can have long and irregular revisit times [2], where usual time-series' evaluation methods of learning and retrieving spatio-temporal structures [1] cannot be applied.

The definition of changes is usually related to some supervised segmentation method, where preliminary statistical or semantic information is considered. For the semantic level reasoning, in [3] a model is shown to map heterogeneous pixels with similar intermediate-level semantic meaning into land cover classes of various mapping products. When different land classes can be well characterized by different statistical models, energy optimization based segmentation methods can be applied for discriminating green and urban areas [2] in a Markov Random Field (MRF) [4]-[5] framework. As the above examples show, very different levels of information may help in defining terrestrial details of semantic meaning into

land cover classes. However, in these cases a kind of human interaction defining semantic labels (e.g. in [3]) is needed, considering that the clustering is steady over the time. If new details related to new clusters occur then it should be detected and the cluster schema should be updated.

Another class of problems occurs when the definition of possible classes is uncertain, and we have more image samples in time, but the rare sampling rate makes it impossible to find spatio-temporal structures. Exploring these rarely scanned RSITS, the images in the series are usually different in light, weather, season, traffic, flooding or blooming conditions. For this reason a sort of prior segmentation, based on object (building roofs [6]), structure (urban and green areas, roads in [7]) or pixel connection models [8] should precede the comparison of the different time-layers. Since comparison of time-layers needs a common basis for clusters, the segmentation of one time-layer should be parametrized by some preliminary cluster-consensus method (as an unsupervised training), where the map of clusters related to similar areas on different time-layers are assigned to the common multi-layer categories.

For MRF segmentation unsupervised labeling is often used [9][10]. Hierarchical models or tree structures are also applicable for Markov chain models for unsupervised texture segmentation [7][11]. For remote sensing tasks a color segmentation method, using the unsupervised TS-MRF algorithm [12], can be successfully used, but this method is divisive, meaning largely unbalanced clusters. This problem is partly solved in [7] by a graph based representation related to neighborhood relationships, by measuring the context similarity. In the above examples statistical or structural investigation in the neighborhood or multiscale hierarchy gives solutions for unsupervised or semi-supervised methods to avoid direct human annotation.

In the paper, 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 multi-layer MRF [13] 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 we can find robust in-layer and cross-layer descriptors. For this reason we add block-wise similarity measures to the stacking of the layers' pixel/microstructure information; we propose to use Cluster Reward Algorithm (*CRA*) [14] in the multi-layer fusion calculated between layer pairs in the series. The

This work has been supported by the National Research Fund, OTKA 106.374.

Authors are with the Distributed Events Analysis Research Laboratory, Institute for Computer Science and Control, Hungarian Academy of Sciences, H-1111 Kende utca 13-17, Budapest, Hungary. E-mail: {sziranyi,maha}@sztaki.mta.hu.

Digital object identifier:

novelties of our 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 *multi-layer labeling*;
- Multi-layer 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.

II. MULTI-LAYER FUSION-MRF MODEL

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

$$\bar{x}_s^{L_i} = [\bar{x}_{C(s)}^{L_i}, \bar{x}_{M(s)}^{L_i}]^T \quad (1)$$

where $\bar{x}_{C(s)}^{L_i}$ contains the pixel's color values, and $\bar{x}_{M(s)}^{L_i}$ 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 RSITS are characterized by the stack $\bar{x}_s^{L_{i_1 \dots i_n}}$ of these vectors for a reasonable set of them, $n \leq N$:

$$\bar{x}_s^{L_{i_1 \dots i_n}} = \{\bar{x}_s^{L_{i_1}}, \bar{x}_s^{L_{i_2}}, \dots, \bar{x}_s^{L_{i_n}}\} \quad (2)$$

A. Fusion-MRF: multi-layer segmentation and change detection

For MRF segmentation, more details can be found in [2], [4], [5], [9]. Once feature vectors are generated, the six steps of the algorithm proposed here are applied. This 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 [14]. In case of professional data suppliers orthonormed and geographically registered images are given; no further registration is needed. In our 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 ($\bar{x}_s^{L_{i_1 \dots i_n}}$) and calculating the cluster parameters (mean and covariance of the conditional term in eq. 6) for the fusion based "multi-layer 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 (see Subsection II-C) on the fused layer data ($\bar{x}_s^{L_{i_1 \dots i_n}}$) containing the cross-layer measures (eq. 5), and the multi-layer cluster parameters, resulting in a multi-layer labeling $\Omega_{L_{i_1 \dots i_n}}$;
- 4) Single-layer training: the map of multi-layer labeling $\Omega_{L_{i_1 \dots i_n}}$ 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 multi-layer 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: Ω_{L_i} ;
- 6) The consecutive image layers $(\dots, (i-1), (i), \dots)$ are compared to find the changes among the different label maps to get the $\delta_{i-1, i}$ change map:

$$\delta_{i-1, i}(\cdot) = \left[\left(\Omega_{L_i}(\cdot) \neq \Omega_{L_{i-1}}(\cdot) \right) = TRUE \right] \quad (3)$$

B. Cross-layer CRA similarity measures in the fusion

Different similarity measures have been considered in the preliminary tests, such as distance to independence, mutual information, *CRA* [14], Kullback Leibler divergence (see [15] and references therein).

CRA(I, J) between two images I and J is calculated using the joint histogram of the two images: p_{IJ} and the marginal histograms: p_I, p_J as follows [15]:

$$CRA(I, J) = \frac{\sum_{i,j} p_{IJ}^2(i, j) - \sum_i p_I^2(i) \cdot \sum_j p_J^2(j)}{\sqrt{\sum_i p_I^2(i) \cdot \sum_j p_J^2(j) - \sum_i p_I^2(i) \cdot \sum_j p_J^2(j)}} \quad (4)$$

The value of *CRA*(I, J) is large when there is high correlation between the two images or the joint histogram has little dispersion. The *CRA* similarity measure is chosen as it gives better segmentation and change detection results than other similarity measures such as correlation and mutual information. This is due to the fact that joint histogram estimation noise has weak influence on the *CRA* values and thus smaller window size can be used [15], which in turn enables detection of changes in small areas. In the proposed segmentation algorithm a multi-layer 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 our presentation here we used three consecutive images only, however the algorithm can be easily extended to more layers. The stack of feature vectors $\bar{x}_s^{L_{1 \dots 3}}$ 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; The window size can be varied according to the required scale of change detection. Each *CRA* image is then normalized to have values in the range $[0, 1]$. 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:

$$\bar{x}_s^{L_{i, i+1, i+2}} = [x_s^{L_i} + \alpha CRA_s(i, i+1), x_s^{L_{i+1}} + \alpha CRA_s(i+1, i+2), x_s^{L_{i+2}} + \alpha CRA_s(i, i+2)]^T \quad (5)$$

where α is a positive normalizing scalar ensuring the same range of the two different terms.

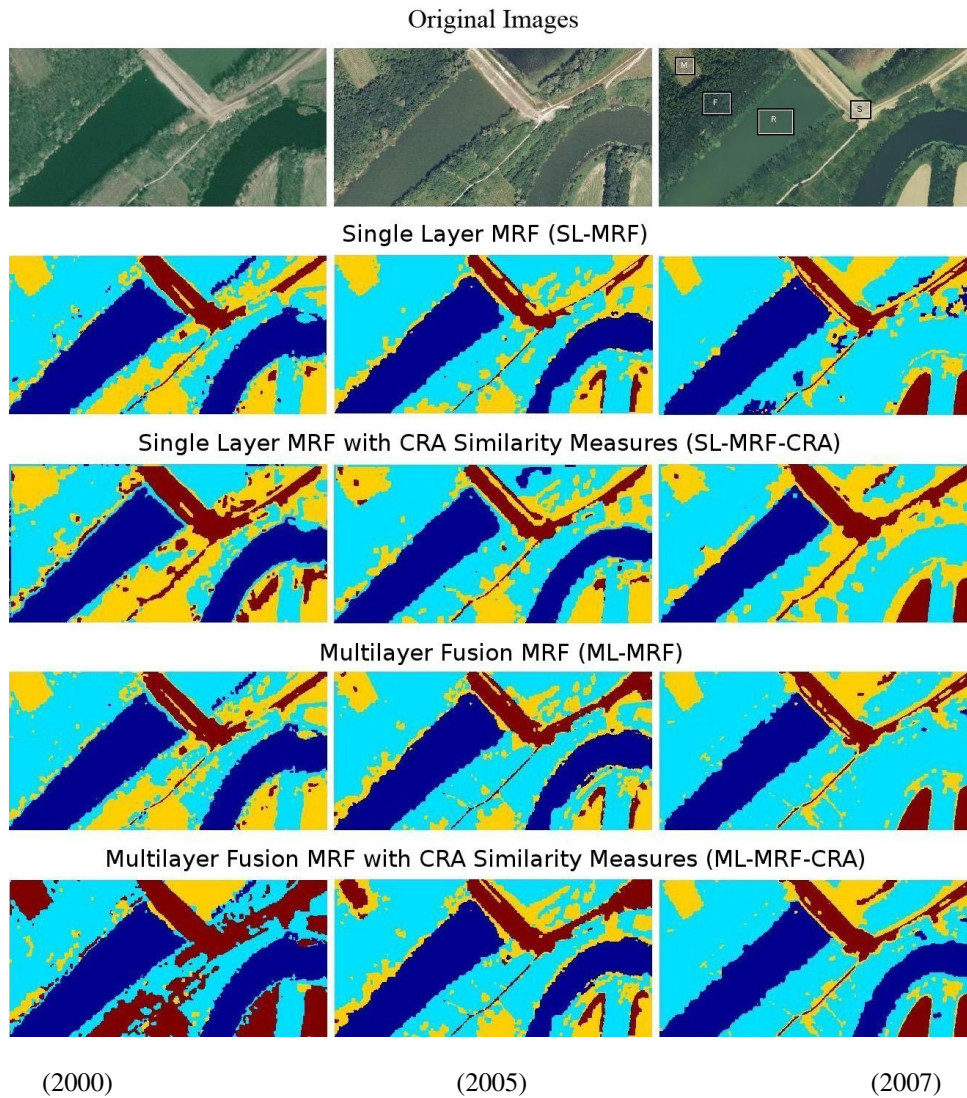


Fig. 1. Row (1): aerial photos around the Tiszadob oxbow area (Hungary, photos by FÖMI) from the years 2000, 2005 and 2007. Rows (2-5): segmentation results for the proposed new ML-MRF-CRA algorithm compared the performance to that of the conventional SL-MRF method, its improvement with *CRA* (SL-MRF-CRA), and the proposed ML-MRF without *CRA*. Training areas used in the segmentation process are shown in the upper right image, Meadow (M), Forest (F), Sand (S), and River (R).

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 eq. (5) means lower dimensionality than using these features as two separate values as in eq. (1). 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.

C. MRF optimization

Let $\mathcal{S} = \{s_1, s_2, \dots, s_H\}$ denote the image pixels, and $\Omega = \{\omega = (\omega_{s_1}, \dots, \omega_{s_H}) : \omega_{s_i} \in \Lambda, 1 \leq i \leq H\}$ be the set of all possible labels assigned to the image classes. We use a Maximum A Posteriori (MAP) estimator for the label field. The MAP estimator is realized by combining a conditional independent random field of observed data $P(\bar{x}_s | \omega_s)$ and an unconditional Potts model [16] for forming smooth connected regions, eq. (7). The optimal segmentation corresponds to the

global labeling $\hat{\Omega}$ and is defined by the energy minimum:

$$\hat{\Omega} = \operatorname{argmin}_{\Omega} \left[\sum_{s \in \mathcal{S}} -\log P(\bar{x}_s | \omega_s) + \sum_{r, s \in \mathcal{S}} \Theta(\omega_r, \omega_s) \right] \quad (6)$$

where the minimum is searched over all the possible segmentations (Ω) of the input, and the $\Theta(\omega_r, \omega_s)$ neighborhood-energy term is zero if s and r are not neighboring pixels, otherwise Θ can be modified by applying the β homogeneity weight:

$$\Theta(\omega_r, \omega_s) = \begin{cases} 0 & \text{if } \omega_r = \omega_s \\ +\beta & \text{if } \omega_r \neq \omega_s \end{cases} \quad (7)$$

In our application we set $\beta = 10.0$ and we used a graph cut based α -expansion algorithm for energy minimization of MRF, with the adherent implementation of [5].

III. EXPERIMENTAL RESULTS AND DISCUSSION

Experiments have been tested on aerial scanned images of Tiszadob area from years 2000, 2005, 2007; the spatial

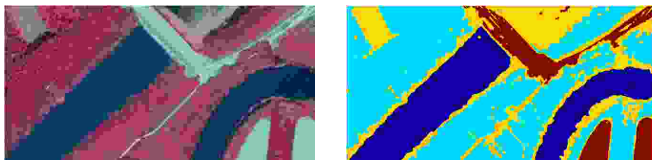


Fig. 2. Ground-truth result by a one layer fusion with infrared image to find the class variants of different water covers. Infrared image from the year 2007 (Left) and the segmentation results for 2007 (Right) using MRF segmentation on CIE *Lab* color values (Fig.1 top right input layer) fused by the infrared image. Compare the results to the 3rd column of Fig.1.

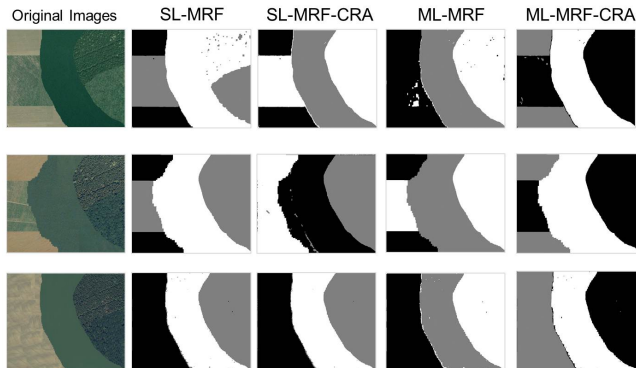


Fig. 3. Unsupervised segmentation results for the multi-layer image segmentation for meadow, river and three variants of forest/bush categories; the proposed similarity measure based fusion-MRF model gives the best result with automatically synchronized cluster definition.

resolution is 0.5m/pixel, and all the images have been aligned as orthophoto. 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). The multi-layer fusion-MRF and the CRA cross-layer measure proposed in Section II are validated on the above aerial images of very different scanning time instants and seasonal conditions, having complex (multiple pattern) classes. First, a multi-class (having 4 labels) segmentation is performed to find meaningful regions in a partly supervised multi-layer segmentation procedure. Then, in a second experiment, we show results evaluated on a ground-truth image series, which is artificially composed of different parts taken from the Tiszadob image series. In the experiment MRF segmentation (Sect. II-A) on the fused images is applied following the unsupervised K-means clustering of the fused data. The labeling resulting from the fused segmentation is then fed into each single layers for training the Gaussian models [2], [9] of the inlayer clusters. In both experiments, we compare the performance of our *new multi-layer MRF* segmentation against independently processed single layer labeling [9],[10]. The comparison is carried out with and without including the *CRA* values in the feature vector to highlight the effect of the *proposed* fusion of *CRA* images. That is, we compare the performance of the *best* proposed ML-MRF-CRA with the following three methods:

- Single Layer MRF on CIE *Lab* color values: SL-MRF;
- SL-MRF with CRA similarity measures: SL-MRF-CRA;
- The proposed multi-layer fusion-MRF on *Lab* color values only: ML-MRF.

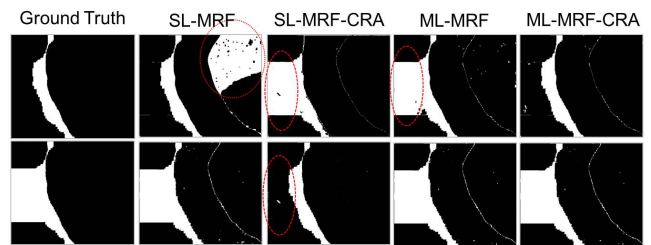


Fig. 4. Change detection results for unsupervised multi-layer image segmentation methods in Fig.3; ML-MRF-CRA method performs the best. Circled areas denote miss-classified regions. See the numerical results in Table I. First row: changes 200-2005, Second row: changes 2005-2007

TABLE I
MISCLASSIFIED PIXELS' RATES FOR FIG.1 AND FIG.4, AND THE EXECUTION TIME (MATLAB, 2.67GHZ) OF THE THREE LAYERS IN FIG.1

| Method | Sup.Segment. 2007, Fig.1 | Unsup.Change Det, Fig.4 2003-2005 | 2005-2007 | Time sec |
|------------|-----------------------------|--------------------------------------|-----------|-------------|
| SL-MRF | 19% | 16.4% | 10.4% | 43 |
| SL-MRF-CRA | 20% | 11.3% | 11.2% | 293 |
| ML-MRF | 21% | 11.04% | 1.0% | 52 |
| ML-MRF-CRA | 10% | 0.65% | 1.0% | 290 |

A. Supervised multi-class segmentation of image time series using cross-layer CRA

In this experiment three 800×400 aerial image sections - shown in Fig. 1 (Row 1) - are evaluated in the comparison. The images are from different scanning time conditions. They consist of four main classes: meadow, forest, river, and sand areas. We can check that the small island that appears in year 2000 input image, does not exist in the other two images. The rivers contain dense vegetation that is very different from one year to the other; this makes it difficult to identify the water class based on color or texture alone. Since only one variant per class is trained (see selected training areas in Fig. 1, upper right image), other class-variants (e.g. differently colored oxbow sections of different vegetations in 2007) can be recognized by the help of similarity measure among layers.

Segmentation results for the three layers are shown in Fig. 1; we can see that only ML-MRF-CRA can detect the different oxbow sections with success. For ground-truth illustration purpose, we further run a two-layer MRF optimization for the fusion of the color and infrared images of 2007. The infrared image and resulted labeling of this experiment are shown in Fig. 2. The use of infrared image helped in identifying correctly the water class and other details. Compared to Fig. 1 (2007, right), these results show that oxbow water-surface variants cannot be identified by the methods SL/ML-MRF, but only with ML-MRF-CRA. By using Fig. 2 (right) as ground-truth reference, we can numerically evaluate the labeling methods; rates for 2007 can be found in Column 2 of Table I. These results show that the use of fusion segmentation with *CRA* values has improved the segmentation accuracy significantly from that of the color data only, without using IR source. We can see here that more information can be exploited from the single layers by using cross-layer and fusion data.

B. Unsupervised segmentation and change detection

In [13] we have successfully tested a two-class multi-layer change detection method. Now, numerical evaluation of a three-class unsupervised segmentation and change detection experiment is shown, where one of the classes is assigned by 3 variations. A ground-truth image series has been generated: three 910×750 images composed of original textures (meadow, forest, and river) from the Tiszadob aerial image series. MRF segmentation (Steps 1-3 in II-A) on the fused images is applied following a K-means clustering of the fused data with three classes. The segmented fusion labeling is then fed into each single layer for training (keeping together the class-variations) and then MRF segmentation on each layer is run separately (Steps 4-5 in II-A); segmentation results are shown in Fig. 3; Change detection results (see Sect. II-A, Step 6) are shown in Fig. 4, and in Columns 3&4 of Table I. This experiment shows that the use of the fusion-MRF model with CRA results in proper change detection for compound classes.

C. Discussion

From the above experiments we can conclude that:

- The use of cross-layer similarity helps to better identify some classes where radiometric values are dubious: in Fig. 1 the river class variants for 2007, and in Fig. 3 the meadow, river and forest/bush (three variants) classes;
- The proposed method relies on the $\delta_{i-1,i}$ change map in cross-layer labeling (eq. 3) and does not depend on any threshold values; thus it performs better than similarity measure based change detection algorithms which use thresholds for change/no-change classification [14];
- Since the outcome classes of the multi-layer segmentation is later used in the training of each single layer, similar classes are automatically given similar labels in all layers.
- No specific information of image sources is considered, which makes it possible to use the method for any other image sources or modalities.

The improved performance of the proposed algorithm for both experiments comes along with additional computational load due to the calculation of the CRA images, see Column 5 of Table I. It depends on the used window size and also on the estimation method of the joint histograms. However, the projection of the multi-layer labeling (Step 4 of Sect. II-A) to each single layer makes the MRF on single layers (Step 5 of Sect. II-A) converge faster compared to SL-MRF.

In the above experiments, we used 7×7 -pixel estimation window for the calculation of CRA local histograms; testing the evaluation at different window sizes, this size was found to give the best segmentation as well as change detection of the small island in evaluation of Fig. 1. The choice of the window size depends on the resolution of the images and the scale of the desired change detection. The detection of small changes requires small window size; however larger window size gives better estimation of the CRA similarity measures. Finding the optimum CRA window size for each point adaptively, along with the definition of the scale of change detection and the image resolution, needs further research on the local scale characteristics, see more in [11] or [17].

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. This problem needs further research on the quality of images and the dimensionality of the feature vectors.

IV. CONCLUSION

We have shown that a fused segmentation using cross-image featuring may result in better segmentation or change detection for sparsely sampled remote sensing image series with classes having more undefined variants. The unsupervised or partly supervised (only one class-variant in training) method results in getting the common classes on the different layers, and its projection to the single layers works as an adaptive training.

REFERENCES

- [1] P. Heas and M. Datcu, "Modeling trajectory of dynamic clusters in image time-series for spatio-temporal reasoning," *IEEE Trans. Geosci. and Remote Sensing*, vol. 43, pp. 1635–1647, 2005.
- [2] C. Benedek and T. Sziranyi, "Change detection in optical aerial images by a multilayer conditional mixed Markov model," *IEEE Trans. Geosci. and Remote Sensing*, vol. 47, no. 10, pp. 3416–3430, October 2009.
- [3] D. Bratasanu, I. Nedelcu, and M. Datcu, "Bridging the semantic gap for satellite image annotation and automatic mapping applications," *IEEE J. Selected Topics in Applied Earth Observation and Remote Sensing*, vol. 4, pp. 193–204, 2011.
- [4] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, pp. 721–741, 1984.
- [5] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "A comparative study of energy minimization methods for Markov Random Fields," in *9th ECCV*, vol. 2, 2006, pp. 16–29.
- [6] C. Benedek, X. Descombes, and J. Zerubia, "Building development monitoring in multitemporal remotely sensed image pairs," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 34, pp. 33–50, 2012.
- [7] R. Gaetano, G. Scarpa, and T. Sziranyi, "Graph-based analysis of textured images for hierarchical segmentation," in *British Machine Vision Conference (BMVC)*. BMVA, 2010, p. #318.
- [8] R. Stoica, X. Descombes, and J. Zerubia, "A Gibbs point process for road extraction in remotely sensed images," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 121–136, 2004.
- [9] Z. Kato, J. Zerubia, and M. Berthod, "Unsupervised parallel image classification using Markovian models," *Pattern Recognition*, vol. 32, pp. 591–604, 1999.
- [10] T. Sziranyi and J. Zerubia, "MRF image segmentation using Cellular NN," *IEEE Trans. Circuits and Systems I*, vol. 44, pp. 86–89, 1997.
- [11] G. Scarpa, R. Gaetano, M. Haindl, and J. Zerubia, "Hierarchical multiple Markov chain model for unsupervised texture segmentation," *IEEE Trans. Image Processing*, vol. 18, pp. 1830–1843, August 2009.
- [12] C. D'Elia, G. Poggi, and G. Scarpa, "A tree-structured Markov Random Field model for bayesian image segmentation," *IEEE Trans. Image Processing*, vol. 12, no. 10, pp. 1259–1273, October 2003.
- [13] T. Sziranyi and M. Shadaydeh, "Improved segmentation of a series of remote sensing images: a fusion MRF model," in *CBMI*. IEEE, 2013.
- [14] J. Inglada and A. Giros, "On the possibility of automatic multisensor image registration," *IEEE Trans. Geosci. and Remote Sensing*, vol. 42, pp. 2104–2120, 2004.
- [15] V. Alberga, "Similarity measures of remotely sensed multi-sensor images for change detection applications," *IEEE Trans. Geosci. and Remote Sensing*, vol. 1, pp. 122–143, 2009.
- [16] R. Potts, "Some generalized order-disorder transformation," in *Proceedings of the Cambridge Philosophical Society*, no. 48, 1952, p. 106.
- [17] T. Lindeberg, "Scale selection properties of generalized scale-space interest point detectors," *J. Math. Imaging and Vision*, vol. 46, pp. 177–210, 2013.