MARKOVIAN FRAMEWORK FOR STRUCTURAL CHANGE DETECTION WITH APPLICATION ON DETECTING BUILT-IN CHANGES IN AIRBORNE IMAGES

Csaba Benedek
Department of Information Technology
Pázmány Péter Catholic University
H-1083 Budapest, Práter utca 50/A
email: bcsaba@sztaki.hu

Tamás Szirányi
Distributed Events Analysis Research Group
Computer and Automation Research Institute
H-1111 Budapest, Kende u. 13-17
email: sziranyi@sztaki.hu

ABSTRACT

In the paper we address the problem of change detection in airborne image pairs taken with significant time difference. In reconnaissance and exploration tasks, finding the slowly changing areas through a long tract of time is disturbed by the temporal parameter changes of the considered clusters. We introduce a new joint segmentation model, containing two layers corresponding to the same area of different far times and the detected change map. We tested this co-segmentation model considering two clusters on the photos: built-in and natural/cultivated areas. We propose a Bayesian segmentation framework which exploits not only the noisy class-descriptors in the independent images, but also creates links between the segmentation of the two pictures, ensuring to get smooth connected regions in the segmented images, and also in the change mask. The domain dependent part of the model is separated, therefore the proposed structure can be used for significantly different descriptors and problems also.

KEY WORDS
Change detection, MRF

1   Introduction

Change detection is an important precursor in several computer vision applications. The corresponding methods can be divided into different groups. In [1] object silhouettes are extracted on video sequences recorded by fixed cameras, where the background objects are static, while the illumination properties may change in time. Meanwhile, moving cast shadows are removed. In [8] camera trembling and periodic motion in the background (e.g. waving river) are considered. Another important issue is motion detection in images captured by moving cameras. If a long video sequence is available, it is possible to detect and track the objects [9]. On the other hand, if we have only two frames to compare [2], the images should be registered and it is necessary to discriminate the registration errors from the real object displacements. All of the previously mentioned methods are based on comparing the gray or color values of the corresponding pixels[1]. It is more difficult to define changes in situations, where the images, which we compare, were taken with significant time difference. Due to the illumination changes and altering shadow effects the appearance of corresponding territories may be much different. In these cases, we have to carefully define what kind of differences we are looking for, while irrelevant changes should be ignored.

2   Basic goals and notes

In the presented model we search for changes in image pairs from the same areas with respect of given properties. In aspect of these properties, we segment the images using K pixel-clusters: $(Q_0, Q_1, \ldots Q_{K-1})$, and mark the connected image regions whose clusters have changed. For example, in the demonstrating application, a binary segmentation ($K = 2$) is achieved: built-in (Q_0) and unpopulated natural/cultivated (Q_1) areas are discriminated in airborne photos. The test-database contains a huge number of preliminary registered images whose manual checking would be cumbersome and time-consuming.

In the resulting segmented images and change-masks, we expect smooth connected regions corresponding to the different clusters, which can be ensured via Markov Random Fields (MRFs) [3]. However, we must expect noisy cluster descriptors, which may alter by time, moreover, the exact borders of the clusters in the images may be ambiguous, similarly to the case of built-in and unpopulated areas. 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. Therefore, in this case, the result of simple local identity checking on the segmented images is corrupted by several artifacts corresponding to the different segmentations instead of real structural changes[2]. To solve this problem, during the seg-

---

[1]Some of them[8] use a probabilistic interpretation for the pixel correspondency.

[2]We show some corresponding experimental results in Section 6.

mentation procedure of the first image we must consider the second one and vice versa. Hence, we segment the images 'together' forcing the corresponding regions to have the same segmentation-masks regarding the two images.

In this paper, we give a Bayesian approach on the above problem. Here, we derived features describing the different class-memberships of a given image point through a simple textural feature and we have developed a MRF model to perform the common segmentation. We emphasize that our model framework may work together with more sophisticated features [5] and for significantly different problems [e.g. trees, rivers]. However, the improved segmentation versus earlier methods segmenting the images separately can be already observed with this problem and feature selection. For simpler notation, we use only two clusters ($K = 2$) in the following descriptions, since it is appropriate for the selected problem, and the generalization for arbitrary number of segmentation-classes is straightforward. The sketch of our method is as follows: first, we map the change detection problem to a Potts-MRF [6] lattice structure, which has the same size as the input images. We can assign a label to each site of the MRF-lattice, and a field energy corresponds to each global labeling of the model. Next, we find the optimal (or at least, a good suboptimal) global labeling on the above model with respect of the previous energy term. Finally, we map the resulting labeling back to the segmentation problem. The appropriate construction of the field energy operator is responsible for getting appropriate segmentation with respect of the above mentioned notes. The key point in our model is that a label of a given image point is a three dimensional vector. The first and second components indicates whether the given pixel corresponds to the Q_0 (built-in) or Q_1 (unpopulated) cluster in the first and second images, respectively. The third component gives the 'changed'/'unchanged' result.

3 Image model and feature extraction

3.1 Image model

Denote X_1 and X_2 the two frames to compare above the same pixel lattice S. A pixel is defined by a two dimensional vector containing its x-y coordinates: $s = [s_x, s_y]^T$, $s_x = 1...M$, $s_y = 1...N$. We define a 4-neighborhood system on the lattice:

$$\forall s \in S : \ \Phi_s = \{r \in S : ||s - r||_{L1} = 1\}, \quad (1)$$

where we determine the distance between two pixels by the Manhattan (L1) distance.

3.2 Feature selection

Built-in areas usually contain several sharp edges near the borders of houses and roads, while in the fields and forests the density of edges is lower. In the experiments, we found
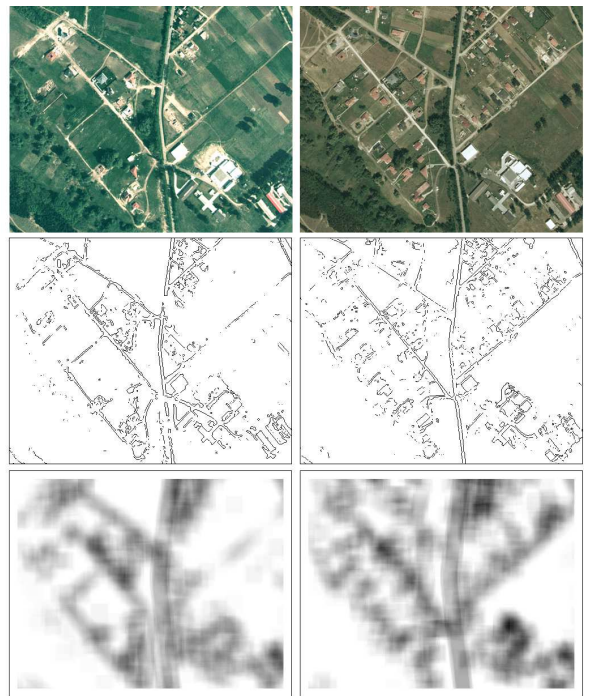

Figure 1. **Feature extraction.** Row 1: images (X_1 and X_2), Row 2: Prewitt edges (E_1 and E_2), Row 3: edge density images (T_1 and T_2; dark pixel correspond to higher edge densities)

the texture descriptor of Rosenfeld and Troy [7] as a good indicator for discriminating these areas. Namely, if $E(s)$ is the element corresponding to pixel s in the binary (Prewitt) edge image of X, the edge density descriptor T is defined by:

$$T(s) = \frac{1}{(2W + 1)^2} \sum_{\substack{r \in S \\ ||s-r|| \leq W}} E(r).$$

Let T_1 and T_2 be the edge density images of X_1 and X_2, respectively.

4 MRF segmentation model

In this section we introduce a Markov Random Field model on the image lattice. First, we define two label sets $L_s \triangleq \{Q_0, Q_1\}$, $L_c \triangleq \{+, -\}$; and a labeling operator:

$$\Omega : S \to L_s \times L_s \times L_c$$

$$\Omega(s) = [\omega_1(s), \omega_2(s), \omega_*(s)]$$

where $\omega_1(s)$ and $\omega_2(s)$ labels define the Q_0/Q_1 segmentation classes of pixel s in the first and second images, respectively[3]. Change label $\omega_*(s)$ indicates whether there was built-in change (+), or not (-) at pixel s. The output

---

[3]Note: it was defined earlier that Q_0 means 'built-in', Q_1 indicates unpopulated regions.

of the change detector consists of the change labels of the different pixels. However, we show in the following that during the optimizing procedure, the segmentation labels play also important roles to get smooth and consistent solution.

A global labeling $\underline{\Omega}$ is defined on the MRF model:

$$\underline{\Omega} = \{[s, \Omega(s)] \mid s \in S\},$$

Θ denotes the set of all the possible global labelings. We define the observation process by the following:

$$\mathcal{F} = \{[s, f(s)] \mid s \in S\},$$

where

$$f(s) = [T_1(s), T_2(s)].$$

We use a maximum a posteriori (MAP) estimator for the label field, namely, the goal is to find the global labeling $\widehat{\underline{\Omega}}$, where:

$$\widehat{\underline{\Omega}} = \mathrm{argmax}_{\underline{\Omega} \in \Theta} P(\underline{\Omega}|\mathcal{F}) =$$
$$= \mathrm{argmin}_{\underline{\Omega} \in \Theta}\{-\log P(\mathcal{F}|\underline{\Omega}) - \log P(\underline{\Omega})\}. \quad (2)$$

Based on the Hammersley-Clifford theorem [3] $P(\underline{\Omega}|\mathcal{F})$ follows Gibbs distribution:

$$P(\underline{\Omega}|\mathcal{F}) = \frac{\exp\left(-U(\underline{\Omega}, \mathcal{F})\right)}{Z} = \frac{\prod_{C \in \mathcal{C}} \exp\left(-V_C(\underline{\Omega}_C, \mathcal{F}_C)\right)}{Z},$$

where U is an energy function, $\mathcal{C}$ is a set containing cliques of sites, $\underline{\Omega}_C$ is the subset of $\underline{\Omega}$ corresponding to a given clique $C \in \mathcal{C}$:

$$\underline{\Omega}_C = \{[q, \Omega(q)] \in \underline{\Omega} \mid q \in C\}.$$

We define $\mathcal{F}_C$ similarly to $\underline{\Omega}_C$ as a subset of $\mathcal{F}$.
V_C is the clique potential function, while Z is a normalizing factor ensuring to present a valid density function. Z is independent of $\underline{\Omega}$.
We can rewrite eq. (2):

$$\widehat{\underline{\Omega}} = \mathrm{argmin}_{\underline{\Omega} \in \Theta} \sum_{C \in \mathcal{C}} V_C(\underline{\Omega}_C, \mathcal{F}_C). \quad (3)$$

We search for the optimal (or reasonable suboptimal) solution of eq. (3) with the Modified Metropolis Dynamic [4]. The proposed model is determined by the cliques and their corresponding clique potential functions. We class the cliques in two groups: we define singletons ($\mathcal{C}_1$) and multisite cliques ($\mathcal{C}_2$). $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2$. The exact definitions are given in Section 4.1 and 4.2, respectively.
To make the outline of the model easier, we visualized the structure in Fig. 4, where we gave examples how singleton and multi-site clique potential can be calculated considering the given labelings at two neighboring sites.

4.1 Singletons

The set of singleton cliques is defined by
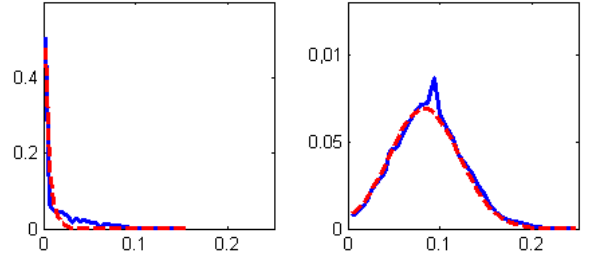
$$\mathcal{C}_1 = \{\{s\} \mid s \in S\}.$$

Figure 2. Left: Histogram (blue continuous line) of the occurring T(.) values regarding manually marked 'unpopulated' (Q_1) pixels and the fitted Beta density function (with red dashed line). Right: Histogram for 'built-in' (Q_0) pixels and the fitted Gaussian density.

The potential of the singleton cliques expresses that the $\omega_1(s)$, $\omega_2(s)$ label components should be consistent with the $T_1(s)$ and $T_2(s)$ observation values (parts of $-\log P(\mathcal{F}|\underline{\Omega})$ in U), while the $\omega_*(s)$ 'change label' should be equal with the 'xor' result on $\omega_1(s)$ and $\omega_2(s)$ in 'most cases'. Therefore,

$$V_{\{s\}} = -\log P(f(s)|\Omega(s)) + \psi(\Omega(s)). \quad (4)$$

We begin the description with the observation-dependent term:

$$P(f(s)|\Omega(s)) = P(T_1(s)|\omega_1(s)) \cdot P(T_2(s)|\omega_2(s)),$$

which expresses that the textural feature processes are conditionally independent from each other in the two layers, given their class labels. E.g. $P(T_1(s)|\omega_1(s) = Q_1)$ is the probability of the fact that the Q_1 class process generates the observation $T_1(s)$ at pixel s.
Our next task is to define an appropriate probabilistic description of the occurring observation values generated by the Q_0/Q_1 classes. First, we performed experiments: regarding different image pairs, we plot the histograms of the occurring $T_1(s)$ and $T_2(s)$ values corresponding to manually marked 'built-in' and 'unpopulated' region points in the input images. Fig. 2 contains the histograms generated for the second image from Fig 1. We observed, that regarding the distribution of the Q_1-classed $T(s)$ values, a Beta density function, $B(., \alpha, \beta)$, was an appropriate approximation, while the values in 'built-in' areas followed Gaussian distribution $N(., \mu, \sigma)$. With these notations:

$$P(T_1(s)|\omega_1(s) = Q_1) = B(T_1(s), \alpha_1, \beta_1),$$

$$P(T_2(s)|\omega_2(s) = Q_1) = B(T_2(s), \alpha_2, \beta_2),$$

$$P(T_1(s)|\omega_1(s) = Q_0) = N(T_1(s), \mu_1, \sigma_1),$$

$$P(T_2(s)|\omega_2(s) = Q_0) = N(T_2(s), \mu_2, \sigma_2).$$

Here we note that the only application-dependent part of the segmentation model is defining the above a posteriori probabilities. Other features and distributions may be used

for other problems.

Next, we introduce the second term in eq. (4), which is responsible for forcing the desired relationship between the parts of the label vector. Usually, the change label of a given pixel is '+' (change), if and only if its segmentation labels are different. However, we consider that noise or segmentation artifacts may also cause erroneous different segmentation labels. Therefore, we give only penalty if the label vector is not consistent, but do not exclude theses cases.

We introduce the following indicator function for $i \in \{1, 2, *\}$:

$$I_i : S \rightarrow \{0, 1\},$$

where

$$I_i(q) = \begin{cases} 1 & \text{if } \omega_i(q) \in \{Q_0, +\} \\ 0 & \text{if } \omega_i(q) \in \{Q_1, -\} \end{cases}$$

With this notation:

$$\psi(\omega_*(s)) = \begin{cases} -\rho & \text{if } I_*(s) = I_1(s) \oplus I_2(s) \\ +\rho & \text{otherwise.} \end{cases}$$

where $\oplus$ means modulo 2 addition.

4.2 Multi-site cliques

The multi-site cliques are responsible for getting smooth connected regions of sites with the same label both during the built-in/unpopulated segmentation of the inputs and also in the change mask. The smoothness is ensured by forcing the neighboring sites to have usually the same labels. Therefore, multi-layer cliques are defined:

$$\mathcal{C}_2 = \{ \{s, r\} \mid r \in \Phi_s; r, s \in S \}.$$

The clique potentials follow the Potts constraint [6]. If $C_2 = \{s, r\} \in \mathcal{C}_2$:

$$V_{C_2} = \sum_{i=1,2,*} \delta^i J(\omega_i(s), \omega_i(r))$$

where for $i \in \{1, 2, *\}$: $\delta^i > 0$ and

$$J(\omega_i(s), \omega_i(r)) = \begin{cases} -1 & \text{if } \omega_i(s) = \omega_i(r) \\ +1 & \text{if } \omega_i(s) \neq \omega_i(r) \end{cases}$$

5 Parameter settings

The free parameters of the method can be classified into different groups. W determines the size of the window, where the edge density texture is collected. We used $W = 5$ for images of size 320×256.

5.1 Parameters of the observation dependent term

We determined the 'built-in' class' Gaussian parameters μ_1, σ_1, μ_2, σ_2 and the unpopulated areas' Beta parameters α_1, β_1, α_2, β_2 with supervision, using manually marked training images.
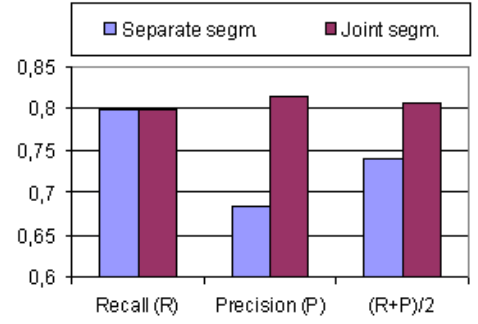

Figure 3. Comparison of the *Recall*, the *Precision* rates, and their average regarding the 'separate segmentation' and the proposed 'joint segmentation' methods.

5.2 Parameters of the clique regularization terms

The parameters of the intra-layer clique potential functions, δ^1, δ^2 and δ^* influence the size of the connected blobs in the segmented images, while ρ relates to the strength of the constraint between the segmentation labels and the 'change label' corresponding to a given site. We set these parameters to 1.

6 Results

We tested our method on registered airborne image pairs captured with 5-20 years time differences. We emphasize, that the primary goal of the test was the validation of the proposed co-segmentation framework, not the appropriateness of the edge density feature as built-in area detector. Therefore, we generated the results for comparison in the following ways:

1. *Joint segm:* We segmented the images and derived the change mask by the proposed model.

2. *Separate segm:* We segmented the images individually and used a simple xor operation to derive the change mask. More precisely, in the proposed framework, we ignored the $\psi(\omega^*(s))$ change mask regularization term ($\rho = 0$), otherwise we optimized the MRF model with the same parameters as before. Finally, we set the change term to fulfill

$$I_*(s) = I_1(s) \oplus I_2(s).$$

The evaluations were done through manually generated ground truth masks. Segmentation results with the two methods for three different image pairs are in Fig. 5.

Regarding the numerical evaluation, denote the number of correctly identified changed pixels of the evaluation images by TP (*true positive*). Similarly, we introduce FP for misclassified not-changed points, and FN for misclassified changed points. The evaluation metrics consists of
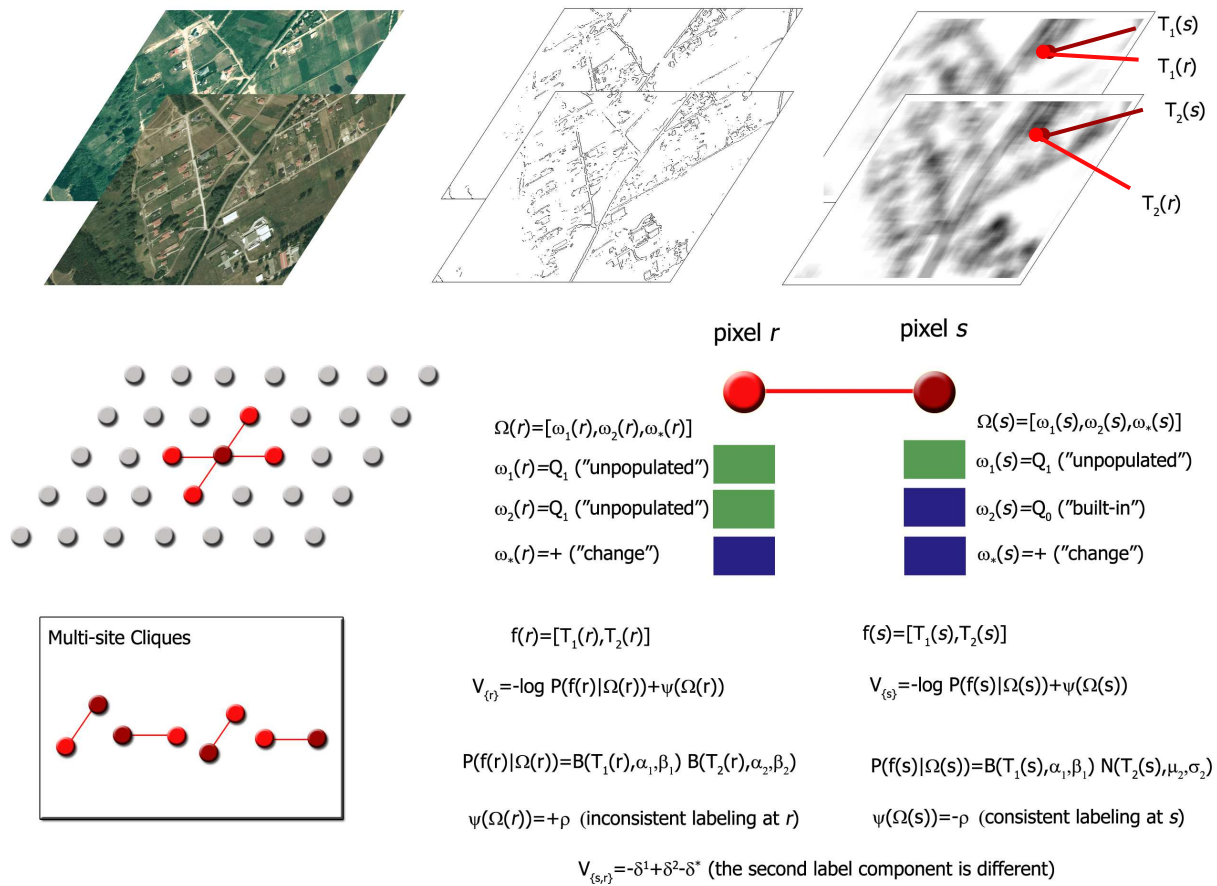
Figure 4. Summary of the proposed model structure and examples how different clique-potentials are defined there. Assumptions: r and s are neighboring sites, while $\Omega(r) = [Q_1, Q_1, +]$ and $\Omega(s) = [Q_1, Q_0, +]$. The calculation of $V_{\{r\}}$, $V_{\{s\}}$ and $V_{\{r,s\}}$ potential terms are demonstrated.
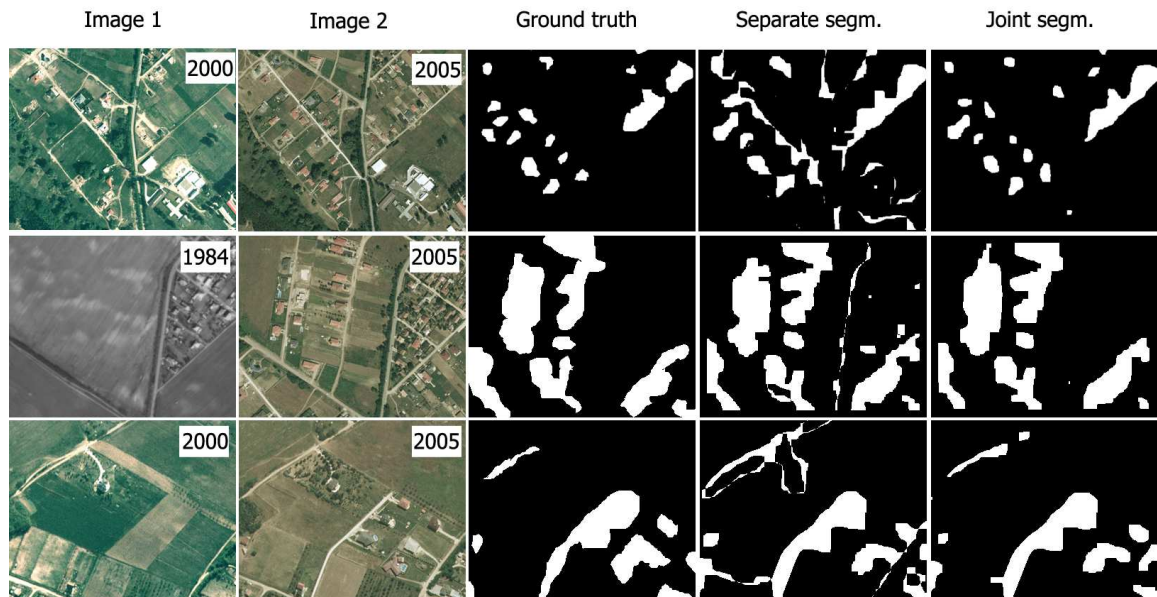

Figure 5. **Validation.** Col. 1 and 2: inputs (with the year of the photos), Col. 3: Ground truth for built-in change detection Col. 4. Change-result with 'separate segmentation'. Col. 5. Change-result with the proposed 'joint segmentation' model.

Figure 6. Illustration of the segmentation results after optimization of the proposed MRF model. Left and middle: marking built-in areas in the first and second input images, respectively. Right: marking the built-in changes in the second photo.

the *Recall* rate and the *Precision* of the change detection.

$$\text{Recall} = \frac{TP}{TP + FN} \qquad \text{Precision} = \frac{TP}{TP + FP}$$

The results are in the diagram of Fig. 3. We can observe that although the *Recall* rates with the two methods are very similar, the *Precision* of the joint segmentation significantly better, since the proposed model is able to eliminate the slightly different segmentations' artifacts.

Finally, we note that the proposed model presents also the 'built-in'/'unpopulated' segmentation of the input images by considering the ω_1 and ω_2 label components, respectively (Fig. 6).

7   Conclusion

In this paper, we addressed the problem of change detection in image pairs taken with significant time difference. We introduced a general co-segmentation model and illustrated its advantages versus segmenting the images separately via a selected application: detecting built-in area changes in airborne photos.

8   Acknowledgement

References

[1] Cs. Benedek, and T. Szirányi, Markovian Framework for Foreground-Background-Shadow Separation of Real World Video Scenes, *Proc. Asian Conference on Computer Vision, LNCS 3851*, pp. 898-907, 2006.

[2] D. Farin and P. With, Misregistration Errors in Change Detection Algorithms and How to Avoid Them, *Proc.* *International Conference on Image Processing*, vol. 2, pp. 438–441, 2005.

[3] 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.

[4] Z. Kato, J. Zerubia, and M. Berthod, Satellite Image Classification Using a Modified Metropolis Dynamics, *Proc. International Conference on Acoustics, Speech and Signal Processing,* vol. 3, pp. 573–576, 1992.

[5] A. Lorette, X. Descombes and J. Zerubia: Texture analysis through a Markovian modelling and fuzzy classification: Application to urban area Extraction from Satellite Images, *International Journal of Computer Vision*, vol. 36, No. 3, pp. 221–236, 2000.

[6] R. Potts, Some generalized order-disorder transformation, *Proceedings of the Cambridge Philosophical Society,* vol. 48, pp. 106, 1952.

[7] A. Ronsenfeld and E. B. Troy, Visual Texture Analysis, *Proc. UMR-Mervin J. Kelly Communications Conference*, Section 10-1, 1970.

[8] Y. Sheikh and M. Shah, Bayesian Modeling of Dynamic Scenes for Object Detection, *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 27, No. 11, pp. 1778–1792, 2005.

[9] A. Yilmaz, X. Li and M. Shah, Contour Based Object Tracking Using Level Sets, *Proc. Asian Conference on Computer Vision*, 2004.