A Mixed Markov Model for Change Detection in Aerial Photos with Large Time Differences

Csaba Benedek[1,2] and Tamás Szirányi[1]
[1]Distributed Events Analysis Research Group, Computer and Automation Research Institute
H-1111, Budapest, Kende utca 13-17, Hungary, bcsaba@sztaki.hu, sziranyi@sztaki.hu
[2]Currently with Ariana Project-Team INRIA/CNRS/UNSA, B.P. 93, 06902 Sophia Antipolis, France

Abstract

In the paper we propose a novel multi-layer Mixed Markov model for detecting relevant changes in registered aerial images taken with significant time differences. The introduced approach combines global intensity statistics with local correlation and contrast features. A global energy optimization process simultaneously ensures optimal local feature selection and smooth, observation-consistent classification. Validation is given on real aerial photos.

1 Introduction

Automatic evaluation of aerial photo 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. The paper deals with change detection in aerial images taken with many years time difference, probably in different seasons and in different lighting conditions. Therefore, 'simple frame differencing' [2, 7] techniques cannot be used, since the observed pixel levels may be significantly different even in the unchanged areas. We focus on standard optical images, unlike [4], which uses multi temporal SAR imagery exploiting its insensitivity to atmospheric conditions. We only assume that the database contains registered orthophotos. On the other hand, most previous models monitor purely natural [6] or urban [8] territories, or they are dedicated to a specific task like detecting new built-in areas [1] or destructions due to earthquakes [5]. As Fig. 1 and 3 show, the current photos include both built-in and unpopulated regions, including forests, fields and agricultural lands as well, presenting various types of differences. Our goal is detecting changes which are 'unusual' in a statistical manner. In a related PCA-based model [8], the authors assumed that the 'unimportant' differences are caused by alteration of illumination and camera settings. Since the above effects influence the observed sensor values in a multiplicative or additive fashion, they modelled the relationship of the corresponding pixel levels within the unchanged regions by a globally constant linear transform. Similar approaches can be also found in [7]. However, these models disregard that the scene may 'regularly' alter as well, primarily due to the seasonal vegetation changes. Moreover, in agricultural areas which follow crop rotation, the shape and arrangement of the neighboring tracks of a plough-land may be changed significantly. We will show that the regularity of these changes also can be measured in a statistical way, although they may cause significant deviations from the estimated linear approach.

In the paper we propose a robust multi-layer Mixed Markov model [3] to tackle the above change detection problem. We identify the changes through complementary features: global intensity statistics and local correlation. A contrast based selection process is responsible for locally choosing the more reliable feature in the different image regions, while a smooth change map is ensured using local connectivity constraints.

2 Image model and feature extraction

Let G_1 and G_2 be the two registered images which we wish to compare. G_1 and G_2 have an identical pixel lattice S. The gray values are denoted by $g_1(s)$ and $g_2(s)$ for a pixel $s \in S$ of G_1 and G_2, respectively. Our first task is to extract local features at each $s \in S$ which give us information for classifying s as a *changed* (ch) or *background* (bg) i.e. unchanged surface point. Taking a probabilistic approach, we consider the ch/bg classes as random processes generating the features according to different distributions.

We start our investigations in the joint intensity domain of the two images. Here, instead of prescribing a global linear transform between $g_1(s)$ and $g_2(s)$ for the background areas [8], we give a multi modal description of the observed data. We approximate the 2-D histogram of the

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

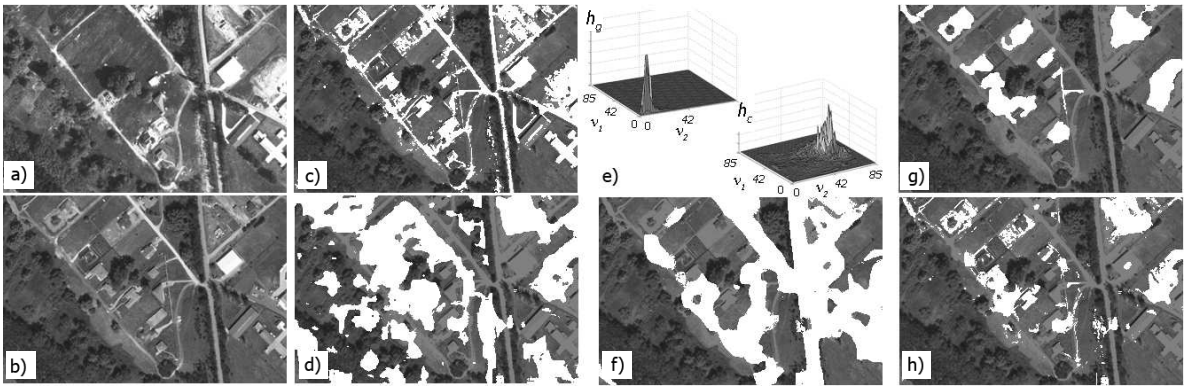

Figure 1. Feature selection: a) G_1 image, b) G_2 image, c) intensity based change detection, d) correlation based change detection, e) h_g and h_c histograms f) local contrast based segmentation, g) ground truth h) change detection with combined features, without considering local connectivity

$\overline{g}(s) = [g_1(s), g_2(s)]^T$ vectors by a mixture of Gaussians distribution. In this way, we 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: $P\big(\overline{g}(s)\big|\mathrm{bg}\big) = \sum_{i=1}^{K} \kappa_i \cdot \eta\big(\overline{g}(s), \overline{\mu}_i, \Sigma_i\big)$, where $\eta(.)$ denotes a two dimensional Gaussian density function with $\overline{\mu}_i$ mean vector and Σ_i covariance matrix, while the κ_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: $P\big(\overline{g}(s)\big|\mathrm{ch}\big) = u$. Next, we demonstrate the limitations of this feature. We derive the segmentation in Fig. 1(c) as the maximum likelihood (ML) estimate, where the label of s is $\phi_g(s) = \mathrm{argmax}_{\psi\in\{\mathrm{ch,bg}\}} P\big(\overline{g}(s)\big|\psi\big)$. One can observe that this multi-Gaussian intensity based approach (MGI) erroneously marks several unaltered regions as changes compared to the ground truth [Fig. 1(g)]. However, the missclassifications are mainly 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.

We derive the next feature, $c(s)$ as the correlation between the rectangular $v \times v$ 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. Our experiments showed that the $P(c(s)|\mathrm{bg})$ and $P(c(s)|\mathrm{ch})$ probabilities can be approximated by different Gaussian distributions. Note that in itself, a simple ML classification based on $c(.)$ results in a fairly poor segmentation ϕ_c [see Fig. 1(d)]. However, we can see that $\overline{g}(s)$ 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 $\overline{g}(s)$.

We formulate the *contrast based feature selection* in a probabilistic manner. Let $\nu_i(s)$ be the local contrast at s in G_i $(i = 1, 2)$, that is measured by the variance of the gray levels in the neighborhood of s. Let be $\overline{\nu}(s) = [\nu_1(s), \nu_2(s)]^T$. Using a few manually segmented training images, one can derive a 2-D histogram h_g which statistically measures the reliability of the decision based on $\overline{g}(s)$ as a function of $\overline{\nu}(s)$. We calculate h_c in a similar manner, for the $c(.)$ feature. As Fig. 1(e) shows, the normalized h_g and h_c histograms can be efficiently approximated by 2-D Gaussian density functions: $P\big(\overline{\nu}(s)|h_\psi\big) = \eta\big(\overline{\nu}(s), \overline{\mu}_\psi, \Sigma_\psi\big)$, $\psi \in \{\mathrm{g}, \mathrm{c}\}$. Thereafter, we create a contrast map as $\phi_\nu(s) = \mathrm{argmax}_{\psi\in\{\mathrm{g,c}\}} P\big(\overline{\nu}(s)|h_\psi\big)$ [see Fig. 1(f): class 'c' marked with white], and the combined map ϕ_*, where $\phi_*(s) = \phi_g(s)$ if $\phi_\nu(s) =$'g'; $\phi_*(s) = \phi_c(s)$ if $\phi_\nu(s) =$'c' [Fig. 1(h)]. Let observe that ϕ_* is a more improved approximation of the ground truth, however, it is still quite noisy. Therefore, based on the above investigations, we introduce a robust segmentation model in the following.

3 A Mixed Markov Random Field image segmentation model

Mixed Markov models [3] 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 $\overline{\nu}(s)$ feature plays a particular role: it may locally switch ON and OFF the $\overline{g}(s)$ respectively $c(s)$ features into the integration procedure. We consider our task as a composition of four interactive segmentation processes [analogously to Fig. 1(c), (d), (f) and (h)]. Thus we map the

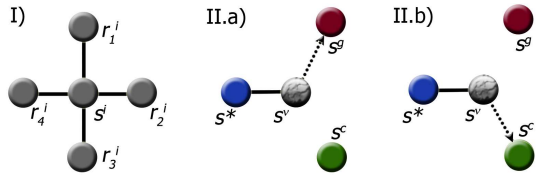

Figure 2. (I) intra- and (II.a,II.b) inter-layer connections in $\mathcal{G}$ (edges denoted by continuous lines, address pointers by dotted arrows).

problem to a graph $\mathcal{G}$ with four layers: S^g, S^c, S^ν 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 $s^c \in S^c$, $s^\nu \in S^\nu$ and $s^* \in S^*$ similarly.

We introduce a labeling random process, which assigns a label $\omega(.)$ to all nodes of $\mathcal{G}$. As usual, graph edges express direct dependencies between the corresponding node labels. Our approach exploits that Mixed Markov models distinguish two types of processing units, called *regular* and *address* nodes [3]. The S^g, S^c, and S^* layers contain *regular nodes*, where the label denotes a possible ch/bg segmentation class: $\forall s \in S, i \in \{g, c, *\} : \omega(s^i) \in \{\mathrm{ch}, \mathrm{bg}\}$. For each s, $\omega(s^g)$ resp. $\omega(s^c)$ corresponds to the segmentation based on the $\overline{g}(s)$ resp. $c(s)$ feature; while the labels at the S^* layer present the final change mask.

On the other hand, the S^ν layer contains *address nodes*, where for $s^\nu \in S^\nu$ the label $\omega(s^\nu)$ 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: $\tilde{\omega}(s^\nu) := \omega(\omega(s^\nu))$ is the label of the (*regular*) node addressed by s^ν, and $\underline{\omega} = \{\omega(s^i) | s \in S,\ i \in \{g, c, \nu, *\}\}$ 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 $\overline{g}(s)$, $\overline{\nu}(s)$ and $c(s)$ local features extracted at pixel s. By definition of Mixed Markov models [3], (static) edges may link any two nodes, and the a posteriori probability of a given global labeling $\underline{\omega}$ is given by:

$$P(\underline{\omega}|\mathcal{F}) = \alpha \prod_{C \in \mathcal{C}} \exp\Big(- V_C\left(\omega_C, \omega_C^\nu, \mathcal{F}\right)\Big), \qquad (1)$$

where $\mathcal{C}$ is the set of cliques in $\mathcal{G}$. For $C \in \mathcal{C}$: $\omega_C = \{\omega(q) | q \in C\}$ and $\omega_C^\nu = \{\tilde{\omega}(s^\nu) \big| s^\nu \in S^\nu \cap C\}$. V_C is a $\mathcal{C} \to \mathbb{R}$ *clique potential function*, which has a 'low' value if the labels within the set $\omega_C \cup \omega_C^\nu$ are semantically consistent, while V_C is 'high' otherwise. Scalar α 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 $\overline{g}(.)$ respectively $c(.)$ values, while the labels in S^* have no direct links with these measurements. For this reason, let be $V_{\{s^g\}} = -\log P\big(\overline{g}(s) \big| \omega(s^g)\big)$, $V_{\{s^c\}} = -\log P\big(c(s) \big| \omega(s^c)\big)$ and $V_{\{s^*\}} \equiv 0$. Note that the above distributions were already defined in Section 2, and $V_{\{s^\nu\}}$ 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 (see Fig. 2-I). Denote the set of the resulting *intra-layer* cliques by $\mathcal{C}_2$. The prescribed potential function of a clique in $\mathcal{C}_2$ will penalize neighbouring nodes having different labels. Assuming r and s to be neighbouring pixels on S, the potential of the doubleton clique $C_2 = \{r^i, s^i\} \in \mathcal{C}_2$ for each $i \in \{g, c, \nu, *\}$ is calculated using a constant $\varphi^i > 0$ as:

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

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 $\nu(s)$ feature. Hence, we put an edge among s^* and s^ν, and prescribe that s^ν should point either to s^g or to s^c (Fig. 2-II.a and II.b). As for the singleton potentials in the S^ν layer, if s^ν points to $s^\psi|_{\psi \in \{\mathrm{g,c}\}}$, let be $V_{\{s^\nu\}} = -\log P\big(\overline{\nu}(s) \big| h_\psi\big)$. On the other hand, we get the potential of the *inter-layer clique* $C_3 = \{s^*, s^\nu\}$ with a fixed $\rho > 0$ as

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

Finally, based on (1), 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:

$$\widehat{\underline{\omega}} = \mathrm{argmin}_{\underline{\omega} \in \Omega} \sum_{s \in S;\ i} V_{\{s^i\}}\big(\omega(s^i), \mathcal{F}_s\big) +$$

$$+ \sum_{\{s,r\} \in \mathcal{C}_2;\ i} V_{C_2}\big(\omega(s^i), \omega(r^i)\big) + \sum_{s \in S} V_{C_3}\big(\omega(s^*), \tilde{\omega}(s^\nu)\big) \qquad (2)$$

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

4 Experiments

The evaluations are done through manually generated ground truth masks using different optical aerial image pairs with 1.5m/pixel resolution. The photos were provided by the Hungarian Institute of Geodesy, Cartography and Remote Sensing and Google Earth. The model parameters are

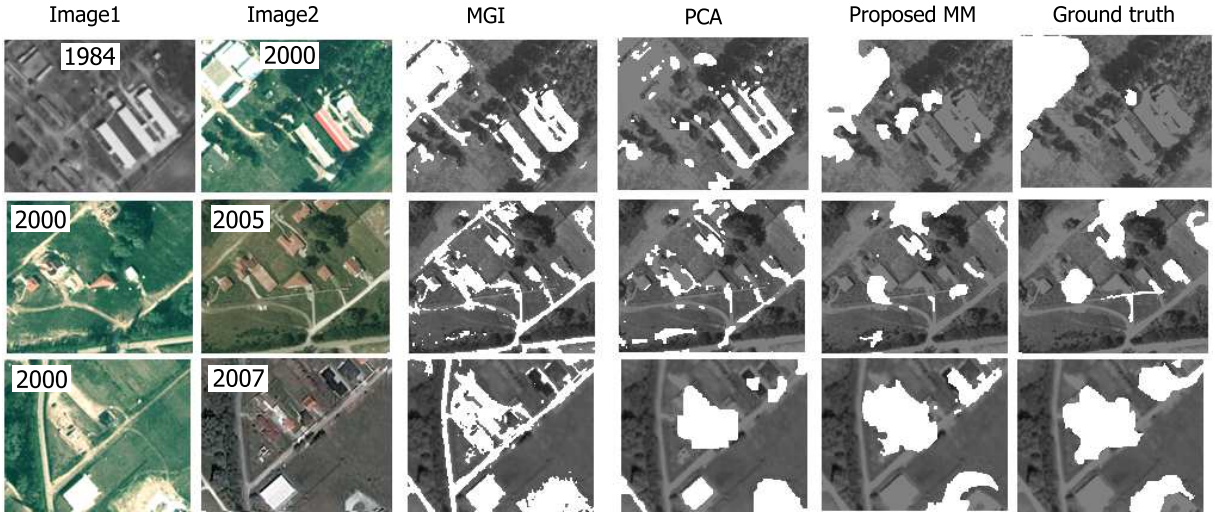

Figure 3. Detected changed regions (with white) in three sample image pairs, using the MGI approach, the PCA-method [8], the proposed Mixed Markov model (MM), and ground truth.

	MGI	PCA	**Prop. MM**
F-measure	0.478	0.605	0.844

Table 1. Quantitative validation results

estimated over a set of training images and we examine the quality of the segmentation on different test pairs. To find a good suboptimal labeling according to (2), we use the modified Metropolis optimization method like in [1, 2].
We have compared the results of the proposed Mixed Markov model to the multi-Gaussian intensity (MGI) based change detection (see Sec. 2), and to the PCA model [8]. The numerical evaluation metric is the F measure, which is the harmonic mean of precision and recall of the detected changed pixels compared to the ground truth. Results are in Table 1. Comparative segmentations of three selected image pairs from the test database are shown in Fig. 3. The experiments show the superiority of the proposed model.

5 Conclusion

This paper has addressed the detection of statistically unusual changes in aerial image pairs taken with a significant time difference. A novel Mixed Markov model has been proposed, which integrates the information from three different observations. The efficiency of the method has been validated through real-world aerial images, and its behavior versus two reference methods has been quantitatively and qualitatively evaluated. The authors would like to thank Josiane Zerubia from INRIA for her kind advices, and the MUSCLE Shape Modeling E-Team for financial support.

References

[1] Cs. Benedek and T. Szirányi. Markovian framework for structural change detection with application on detecting built-in changes in airborne images. In *Proc. SPPRA*, 2007.
[2] Cs. Benedek, T. Szirányi, Z. Kato, and J. Zerubia. A multilayer MRF model for object-motion detection in unregistered airborne image-pairs. In *Proc. IEEE ICIP*, 2007.
[3] A. Fridman. Mixed Markov models. *Proc. National Academy of Sciences of USA*, 100(14):8092–8096, July 2003.
[4] P. Gamba, F. Dell'Acqua, and G. Lisini. Change detection of multitemporal SAR data in urban areas combining feature-based and pixel-based techniques. *IEEE Trans. GRS*, 44(10):2820–2827, 2006.
[5] Y. Kosugi, M. Sakamoto, M. Fukunishi, T. Wei Lu Doihara, and S. Kakumoto. Urban change detection related to earthquakes using an adaptive nonlinear mapping of high-resolution images. *IEEE GRSL*, 1(3):152–156, 2004.
[6] G. Perrin, X. Descombes, and J. Zerubia. 2D and 3D vegetation resource parameters assessment using marked point processes. In *Proc. ICPR*, Hong-Kong, 2006.
[7] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam. Image change detection algorithms: A systematic survey. *IEEE Trans. IP*, 14(3):294–307, 2005.
[8] R. Wiemker. An iterative spectral-spatial bayesian labeling approach for unsupervised robust change detection on remotely sensed multispectral imagery. In *Proc. CAIP*, volume LNCS 1296, pages 263–270, 1997.