

EFFICIENT BUILDING CHANGE DETECTION IN SPARSELY POPULATED AREAS USING COUPLED MARKED POINT PROCESSES

Csaba Benedek

Distributed Events Analysis Research Group, Computer and Automation Research Institute
H-1111, Budapest, Kende utca 13-17, bcsaba@sztaki.hu

ABSTRACT

In this paper we introduce new model for efficient building extraction and change detection in sparsely populated remotely sensed image pairs. In the two time layers interacting marked point processes recognize the buildings while a priori constraints and low level similarity information are involved in the model. The exploration is accelerated through a non-uniform object birth process, which proposes objects in the built-in regions with higher probability based on global statistical image features.

Index Terms— Change detection, building recognition

1. INTRODUCTION

Building change detection is a key task in several applications of remote sensing image analysis, such as monitoring the evolution of built-in areas [1] or extracting damaged buildings after a natural disaster [2]. In this paper, we focus on multitemporal aerial or satellite image pairs from sparsely populated areas. Without topographic building databases from the earlier time instances [1], the task requires automatic building extraction in each image to be compared. Most previous solutions handle the building identification and change detection tasks in a sequential way. In the Post Detection Comparison (PDC) methods [2], independent building detection processes are applied in the earlier and latter images which is followed by object level comparison. Methods using the Post Change Classification (PCC) approach rely on a change map based on low-level similarity features [3] (like difference image or texture comparison), and then search for new/demolished buildings only in the changed regions. As a drawback, in these two-step techniques both the change detection and object extraction parts can be a bottleneck, e.g. in PDC, a missing object resulted by the failure of the object detector cannot be recovered later in the process.

Another important issue is related to building modeling. Direct techniques consist of quick image operations, such as junction detection in the edge map [4] or floodfill based segmentation [5]. However in large and heterogenous databases,

the building and background classes may strongly overlap in the feature domains, making the separation less accurate. Robust methods can be obtained by applying inverse approaches [6], where a fitness value is assigned to each possible object configuration and an optimization process based on the hypothesis generation-acceptance schema attempts to find the configuration with the highest confidence. However, as the processing time is proportional to the size of the images [6], this approach is inefficient in sparse areas, where the most part of the computational power can be wasted to check irrelevant object hypotheses.

The method introduced here attempts to overcome the aforementioned problems. Building extraction is achieved by two interacting marked point processes [6] corresponding to the two time layers, where building candidates are characterized in the inverse modeling approach, meanwhile low level change information affects the detection directly through inter-layer constraints. The building population is evolved by an iterative birth-death dynamics, while the computational cost is significantly decreased by applying a non-uniform object birth process, which proposed relevant object with higher frequency based on joint intensity statistics of the input images.

2. NOTATIONS AND PROBLEM DEFINITION

Let G_1 and G_2 be the two co-registered aerial or satellite images which we wish to investigate. G_1 and G_2 have an identical pixel lattice S . The color values are denoted by $g_1(s)$ and $g_2(s)$ for a pixel $s \in S$ of G_1 and G_2 , respectively.

We consider each building to be constructed from one or many rectangular building segments, which we aim to extract by the following model. Let u be a rectangular building segment candidate. To incorporate the change information, we assign to u a binary indicator flag $\xi(u) \in \{\text{ch}, \text{ua}\}$, where ch means changed object. Let be $R_u \subset S$ the set of pixels corresponding to u . Rectangle R_u is described by its center pixel $c \in S$, and three scalar parameters: e_L , e_l side lengths and $\theta \in [-90^\circ, +90^\circ]$ orientation. We aim to extract a population of building segments in each time layer: $\omega_1 = \{u_{11}, u_{12}, \dots, u_{1m}\}$ in G_1 and $\omega_2 = \{u_{21}, u_{22}, \dots, u_{2n}\}$ in G_2 , where the numbers of objects m and n are also un-

¹This work was supported by the Hungarian Research Fund "Coherent attributes in the interpret. and percept. of the visual world" (OTKA-80352).

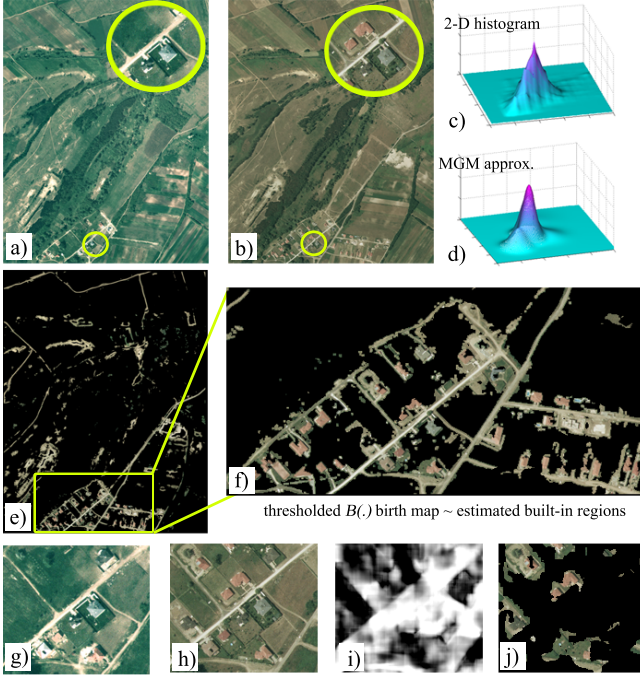


Fig. 1. (a)-(f) birth map generation: (a),(b) input images, (c) 2-D joint histogram $h[g_1^I, g_2^I]$ (d) MGM approximation of h , (e) thresholded $B(s)$ birth map (f) estimated built-in regions **(g)-(j) correlation based change estimation** (g),(h) input image parts (i) Ψ -map (dark region \leftarrow high correlation) and (j) estimation of the changed built-in regions

known. We refer to the whole two-layer configuration as $\hat{\omega} = \{\omega_1; \omega_2\}$

In the proposed model, low level (see Sec. 3) and object level (Sec. 4) features are utilized parallelly. Low level descriptors are mainly exploited by the optimization process (Sec. 6) to estimate roughly where the buildings and changed built-in regions *can* be located. Based on them, we will generate relevant objects with higher probability, which can significantly speed up the exploration phase. On the other hand, object level features are applied to verify the proposed building candidates and keep only the appropriate ones (Sec. 5).

3. BUILT-IN AREA AND CHANGE ESTIMATION

The rural *background* regions at the countryside and rarely populated areas can be usually well described by a few dominant color values [3], e.g. the mean colors of different types of forests, plough lands, grass etc. (see Fig. 1(a),(b)). These colors appear with high weight in the global image histograms, thus the remaining regions, which are expected to contain the built-in territories as well, can be obtained by outlier filtering. Although for one image layer, experimental evidences show that overlap between the background and

built-in-area color domains is usually too large for efficient separation, we can improve the performance by considering color co-occurrences in the two images. For simpler demonstration, we illustrate this step with grayscale inputs. We create first a 2-D histogram h which assigns a frequency value to each intensity pair, e.g. $h[g_1^I, g_2^I]$ is the number of pixels whose gray value is g_1^I in the first image, G_1 , and g_2^I in the second one, G_2 [Fig. 1(c)]. In this way we measure, which intensity pairs occur often together in the two images, therefore they characterize unchanged background regions with a higher confidence. As for colored inputs, a similar schema can be adopted with 6-D histograms. For the density estimation of this histogram we have chosen a Mixture of K 2-D (resp. 6-D) Gaussians approach which is demonstrated in Fig. 1(d). Mixture components with high weight correspond to the frequent color pairs (e.g. typical color of forest in image G_1 and G_2), K is a parameter of the process. Denoting by $\bar{g}(s) = [g_1(s); g_2(s)]^T$ the concatenated color vector of s , we calculate the $P_{bg}(s) = \sum_{i=1}^K \kappa_i \cdot \eta(\bar{g}(s), \bar{\mu}_i, \bar{\Sigma}_i)$ pixel-by-pixel background-probabilities. Thereafter, we derive a normalized building ‘birth map’, which assigns higher birth value for pixels with lower background likelihoods:

$$B(s) = \frac{\max_{r \in S} P_{bg}(r) - P_{bg}(s)}{\max_{r \in S} P_{bg}(r) - \min_{r \in S} P_{bg}(r)}$$

Fig. 1(e) and (f) shows that the thresholded B -map offers an efficient estimation of the regions of buildings

We continue with *change-area* extraction. We derive the next feature, $c(s)$ as the cross correlation between the rectangular $l \times l$ neighborhoods of s in G_1 and in G_2 . Side length l is chosen so that the window can cover an average building narrowly. As pointed out in [3], $c(s)$ is a reliable similarity measure in the high-contrasted built-in regions. We derive a similarity map as $\Psi(s) = \text{sigm}(c(s))$, where $\text{sigm}(\cdot)$ is a sigmoid function which projects the correlation values to $[0, 1]$ in a monotonous, non-linear manner [see Fig. 1(g)-(j)].

4. SINGLE VIEW BUILDING MODEL

The building model assigns to each proposed object candidate an energy term $\varphi^{(i)}(u)$, which evaluates the hypothesis that u is a building in the i^{th} image (hereafter we ignore the i superscript). $\varphi(u)$ is interpreted as the negative building fitness value and rectangles with $\varphi(u) < 0$ are the potentially relevant objects [6]. The energy function can integrate different feature-information such as roof color, roof edge and shadow [7]. On one hand, *red* roofs can be detected in color images using the hue components of the corresponding pixel values [5]. Thereafter with notation \tilde{R}_u for the dilated mask of u ’s rectangle, the color term favours objects which contain in majority roof colored pixels inside R_u and background pixels in $\tilde{R}_u \setminus R_u$. For non-red roofs we can rely on the gradient and shadow maps exploiting that under the roof edges

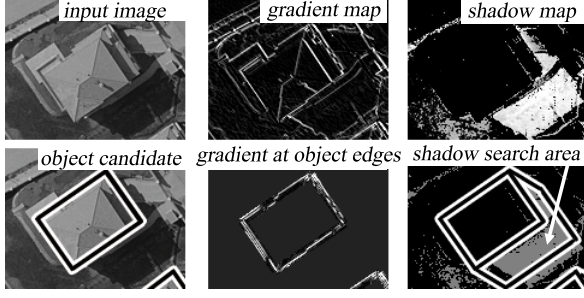


Fig. 2. Edge and shadow features

strong intensity changes should be observed in the images, while in sunny weather dark shadow blobs are present next to the buildings in the shadow direction [4] (See Fig. 2).

5. COUPLED POINT PROCESSES FOR MULTITEMPORAL OBJECT DETECTION

Let be \mathcal{H} the space of u objects. Using a Borel set $H \in \mathcal{H}$, the configuration space is defined as $\Omega = \Omega_1 \times \Omega_2$, where

$$\Omega_i = \bigcup_{n=0}^{\infty} \Omega_i^n, \quad \Omega_i^n = \{\{u_1, \dots, u_n\} \in H^n\}, \quad i \in \{1, 2\}$$

As before, we denote by $\omega_i \in \Omega_i$ an arbitrary object-configuration in the i^{th} image, $i \in \{1, 2\}$. We also define a \sim neighborhood relation in \mathcal{H} : $u \sim v$ if their rectangles R_u and R_v intersect.

We introduce a non-stationary data-dependent energy function on the configuration space as:

$$\Phi_{\mathcal{D}}(\{\omega_1; \omega_2\}) = \sum_{i=1}^2 \sum_{u \in \omega_i} A_{\mathcal{D}}^i(u) + I(u, \omega_i) + C(u, \omega_{i \oplus 1})$$

Here $A_{\mathcal{D}}(u)$ is the data term while $I(u, \omega_i)$ and $C(u, \omega_{i \oplus 1})$ are the intra-layer, resp. inter-layer, prior interaction terms, \oplus denotes mod 2 addition. The optimal configuration estimate can be obtained as $\arg \min_{\omega_1 \in \Omega_1, \omega_2 \in \Omega_2} [\Phi_{\mathcal{D}}(\omega_1, \omega_2)]$.

Unary potentials characterize a given building segment candidate $u = \{c, e_L, e_l, \theta, \xi\}$ as a function of the local image data in the images, but independently of other object of the population:

$$A_{\mathcal{D}}^i(u) = \varphi^i(u) + \frac{\gamma_{\xi}}{\#R_u} \sum_{s \in R_u} \mathbb{I}_{[\xi(u)=ua]} + (-1)^{\mathbb{I}_{[\xi(u)=ua]}} \Psi(s)$$

where $\mathbb{I}_{[E]} \in \{0, 1\}$ is the indicator function of event E , and $\Psi(\cdot)$ is the low level similarity map between the two time layers (Sec. 3). The last term penalizes unchanged objects in the regions of textural differences, and new/demolished buildings in *changeless* areas.

Interaction potentials enforce prior geometrical constraints. Denote by $\Xi(u, v)$ the intersection-coefficient between the u and v object rectangles calculated as $\#(R_u \cap R_v) / \#(R_u \cup R_v)$. Within a time layer, we penalize overlapping objects, thus for $u \in \omega_i$, the intra-layer potential is got by $I(u, \omega_i) = \gamma_I \max_{v \in \omega_i} \Xi(u, v)$. On the other hand, for an unchanged object u in layer i , we expect an unchanged couple object in layer $i \oplus 1$, with high overlapping:

$$C(u, \omega_{i \oplus 1}) = \gamma_C \mathbb{I}_{[\xi(u)=ua]} \left(1 - \max_{v \in \omega_{i \oplus 1}} \mathbb{I}_{[\xi(v)=ua]} \Xi(u, v) \right)$$

where γ_I and γ_C are weighting constants. Observe that to calculate the $I(u, \cdot)$ and $C(u, \cdot)$ we should check only the \sim neighborhood of u instead of the whole population.

6. OPTIMIZATION

We estimate the optimal object configuration by the Multiple Birth and Death Algorithm [6] as follows:

Initialization: calculate the $B(\cdot)$ birth and $\Psi(\cdot)$ similarity maps; start with empty populations $\omega_1 = \emptyset$ and $\omega_2 = \emptyset$.

Main program: initialize the inverse temperature parameter $\beta = \beta_0$ and the discretization step $\delta = \delta_0$, and alternate birth and death steps.

1. *Birth step:* at each pixel $s \in S$, either run the *twin birth* process with probability $\delta B(s) \cdot \Psi(s)$; or run the *single birth* process with probability $\delta B(s)$, as follows:

Single birth generates a new object u with center s , and sets its the change flag, side length and orientation parameters randomly. Finally, u is added *either* to ω_1 *or* to ω_2 .

Twin birth generates an object $u = \{s, e_L, e_l, \theta, ua\}$ with random size and orientation; and a 'twin' object $v = \{s+s_e, e_L+e_L, e_l+e_l, \theta+\epsilon_\theta, ua\}$ by cloning u and slightly perturbing the rectangle parameters. Finally u is added to ω_1 and v is added to ω_2 .

2. *Death step:* Consider the current configuration $\hat{\omega} = \{\omega_1; \omega_2\}$ and create a list of the $u \in \hat{\omega}$ objects sorted from the highest to the lowest $A_{\mathcal{D}}^i(u)$ values. For each object u taken in this order, compute $\Delta \Phi_{\hat{\omega}} = \Phi_{\mathcal{D}}(\hat{\omega}/\{u\}) - \Phi_{\mathcal{D}}(\hat{\omega})$, derive the *death rate* as follows:

$$d_{\hat{\omega}}(u) = \frac{\delta a_{\hat{\omega}}(u)}{1 + \delta a_{\hat{\omega}}(u)}, \quad \text{with} \quad a_{\hat{\omega}}(u) = e^{-\beta \cdot \Delta \Phi_{\hat{\omega}}(u)}$$

and remove u from $\hat{\omega}$ with probability $d_{\hat{\omega}}(u)$.

Convergence test: if the process has not converged yet, increase the inverse temperature β and decrease the discretization step δ with a geometric scheme, and go back to the birth step. The convergence is obtained when all the objects added during the birth step, and only these ones, have been killed during the death step.



Fig. 3. Detection results. Red denotes changed, blue unchanged building.

SZADA	#CH	#UCH	MO	FO	MC	FC	DA
PDC	31	12	6	7	1	9	0.66
CMPP	31	12	3	0	3	0	0.75
BUDAPEST	#CH	#UCH	MO	FO	MC	FC	DA
PDC	20	21	5	3	3	7	0.72
CMPP	20	21	0	3	0	3	0.76

Table 1. Quantitative evaluation results. #CH and #UCH denote total the number of changed resp. unchanged buildings in the test sets SZADA and BUDAPEST. CMPP is the proposed method, further notations are defined in Sec. 7.

7. EXPERIMENTS

We evaluated the proposed Coupled Marked Point Process (CMPP) method on real aerial datasets SZADA [3] and BUDAPEST [7]. Qualitative results are shown in Fig. 3. The comparative validation focused on showing the advantages of the two key contributions of CMPP: the interacting multi-temporal point processes and the non-uniform birth procedure based on Global Intensity Statistics (GIS). We set the CMPP method first against the sequential PDC approach [2] detailed earlier. As quantitative metrics, we measured the number of missing and falsely detected objects (MO and FO), missing and false change alarms (MC, FC), and the pixel-level accuracy of the detection (DA) described by the F-measure [7]. Results in Table 7 show the superiority of CMPP versus PDC. The later method caused false and missing alarms mostly in the first image which had a lower color quality. Moreover since the CMPP uses separate marked objects for the different time instances of the buildings, the model can handle 2-D registration errors more robustly than [7] where the old and new building footprints were described by the same rectangle. Using the GIS birth map does not modify the final results, but

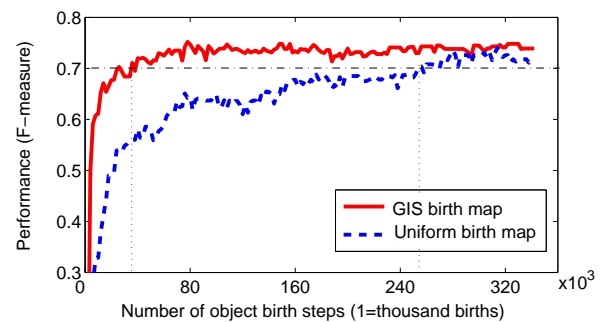


Fig. 4. Evolution of the detection performance over the iteration steps during the optimization. Pixel level DA rate is given as a function of the applied birth steps.

affects the speed of optimization significantly as shown in Fig. 4. Observe that to reach the 70% DA rate, we need to generate 260,000 objects with the uniform map, and only 36,000 building candidates with the GIS map.

8. CONCLUSION

This paper has addressed the detection of buildings and building changes in remotely sensed image pairs using an energy minimization approach. We have shown that the proposed coupled point process schema outperforms the conventional layer-by-layer techniques, while the optimization can be speeded up by a non-uniform, GIS based birth map.

9. REFERENCES

- [1] N. Champion, "2D building change detection from high resolution aerial images and correlation digital surface models," in *IAPRSSIS*, 2007.
- [2] S. Tanathong, K.T. Rudahl, and S.E. Goldin, "Object oriented change detection of buildings after the Indian ocean tsunami disaster," in *IEEE Intl. Conf. ECTI*, 2008, pp. 65–68.
- [3] C. Benedek and T. Szirányi, "Change detection in optical aerial images by a multi-layer conditional mixed markov model," *IEEE Trans. Geosc. Remote Sens.*, vol. 47, no. 10, pp. 3416–3430, 2009.
- [4] B. Sirmacek and C. Unsalan, "Building detection from aerial imagery using invariant color features and shadow information," in *IEEE ISICIS*, Istanbul, Turkey, 2008.
- [5] S. Muller and D.W. Zaum, "Robust building detection in aerial images," in *ISPRS CMRT*, 2005, pp. 143–148.
- [6] X. Descombes, R. Minlos, and E. Zhizhina, "Object extraction using a stochastic birth-and-death dynamics in continuum," *J. Math. Imaging and Vision*, vol. 33, pp. 347–359, 2009.
- [7] C. Benedek, X. Descombes, and J. Zerubia, "Building extraction and change detection in multitemporal remotely sensed images with multiple birth and death dynamics," in *IEEE WACV*, Snowbird, USA, 2009.