A Two-Layer Marked Point Process Framework for Multilevel Object Population Analysis

Csaba Benedek[⋆]

Distributed Events Analysis Research Laboratory, Computer and Automation Research Institute,
H-1111, Kende utca 13-17 Budapest, Hungary
benedek.csaba@sztaki.mta.hu

Abstract. In this paper we introduce a probabilistic approach for extracting object ensembles from various digital images used by machine vision applications. The proposed framework extends conventional Marked Point Process models by allowing corresponding entities to form coherent object groups, by a Bayesian segmentation of the population. A global optimization process attempts to find the optimal configuration of entities and entity groups, considering the observed data, prior knowledge, and local interactions between the neighboring and semantically related objects. The proposed method is demonstrated in three different application areas: built in area analysis in remotely sensed images, traffic monitoring on airborne Lidar data and optical inspection of printed circuit boards.

Keywords: Marked point process, object population analysis

1 Introduction

Object based interpretation of digital images is a crucial step in several vision applications. Due to the quick progress of imaging equipments, we can witness a significant improvement of the available image resolution in many fields. Nowadays in a single image, one can usually detect multiple effects on different scales, calling for recognizer algorithms which perform hierarchical interpretation of the content [7].

Marked Point Processes (MPP) [6] provide an efficient Bayesian tool to characterize object populations, through jointly describing individual objects by various data terms, and using information from entity interactions by prior geometric constraints. However, conventional MPP-based models [4] focus purely on the object level of the scene, as they extract configurations which are composed of similarly shaped and sized entities such as flamingos [4], or buildings [2] in aerial images. Simple prior interaction constraints such us non-overlapping or parallel alignment are also utilized there to refine the accuracy of detection, but in this way only very limited amount of high level structural information can be exploited from the global scenario. The Multi-MPP framework proposed by [6] offers extensions of MPP models regarding two issues: (i) to simultaneously detect variously shaped entities, it jointly samples different types of geometric objects, (ii) by a statistical type and alignment analysis of the extracted nearby entities

local texture representation of the different image regions is obtained. Although this approach fits well to bottom-up exploration tasks of the unknown imaged scene content, it is not straightforward in many vision applications, how to efficiently segment the object population in this framework based on domain specific top-down knowledge.

Up to now, only highly tasks specific attempts have been conducted to model the object encapsulation [1] or the Bayesian object group management [3] issues within the MPP schema. In this paper, as an extension of [3], we introduce a general MPP framework, which enables us to handle a wide family of applications. For avoiding the limitations of using pairwise object interactions only, we propose here a Two-Layer MPP (L^2MPP) model, which partitionates the complete entity population into object groups, called configuration segments, and extracts the objects and the optimal segments simultaneously by a joint energy minimization process. Object interactions are differently defined within the same segment and between two different segments, implementing adaptive object neighborhoods. In this way, we can use in parallel strong alignment or spectral similarity constraints within a group, but the coherent segments may even have irregular, or thin, elongated shapes. We demonstrate the applicability of the proposed L^2MPP model in three different application areas: built-in area analysis in remotely sensed images, traffic monitoring from airborne Lidar point clouds and optical circuit inspection based on line-scanned images with a resolution of a few μm.

2   Problem formulation and notations

The input of the proposed framework is a digital image over a discrete 2D pixel lattice S. Let $u \in \mathcal{H}$ be an object candidate in the image, represented by a plane figure from a preliminary fixed shape library, such as rectangles and ellipses. Each object is described by the shape type attribute $\mathrm{tp}(u)$, its center pixel, global orientation, and a geometry dependent parameter set containing the perpendicular side lengths for rectangles, or the major and minor axes for ellipses. We also use a proximity relation $\sim$ in $\mathcal{H}$: $u \sim v$ if and only if the distance of the object centers is smaller than a threshold. Next, we define the object groups: a global population ω is a set of k *configuration segments*, $\omega = \{\psi_1, \ldots, \psi_k\}$, where each segment ψ_i $(i = 1 \ldots k)$ is a configuration of n_i objects, $\psi_i = \{u_1^i, \ldots, u_{n_i}^i\} \in \mathcal{H}^{n_i}$. Here we prescribe that $\psi_i \cap \psi_j = \emptyset$ for $i \neq j$, while the k set number and $n_1, \ldots, n_k$ set cardinality values may be arbitrary (and initially unknown) integers. We mark with $u \prec \omega$ if u belongs to any ψ in ω, i.e. $\exists \psi_i \in \omega : u \in \psi_i$. Ω denotes the space of all the possible ω global configurations:

$$\Omega = \cup_{k=0}^{\infty} \left\{ \{\psi_1, \ldots, \psi_k\} \in \left[\cup_{n=1}^{\infty} \Psi_n \right]^k \right\} \quad \text{where } \Psi_n = \{\{u_1, \ldots, u_n\} \in \mathcal{H}^n\}$$

Let us denote by $\mathcal{N}_u(\omega)$ the proximity based neighborhood of $u \prec \omega$, which is independent of the group level: $\mathcal{N}_u(\omega) = \{v \prec \omega : u \sim v\}$.

3   Two Layer Marked Point Process Model

Taking an inverse approach, we define in this section an energy function $\Phi(\omega)$, which can evaluate each $\omega \in \Omega$ configuration based on the observed data and prior knowledge.

The constructed energy formula can be decomposed into a unarY term (Y) and an Interaction term (I): $\Phi(\omega) = \Phi_Y(\omega) + \Phi_I(\omega)$. In the following, we introduce the $\Phi_Y(\omega)$ and $\Phi_I(\omega)$ components.

3.1   Unary object appearance terms

Each object u is associated with a *unary* energy term $\varphi_Y(u)$, which characterizes u depending on the local image data, but independently of other objects of the population.

First we define different $f_i(u) : \mathcal{H} \to \mathbb{R}$ features ($f_1 \ldots f_k$) which evaluate an object hypothesis for u in the image, so that 'high' $f(u)$ values correspond to efficient object candidates.

In the *second step*, we construct $\varphi_f(u)$ *data driven* energy subterms for each feature f, by attempting to satisfy $\varphi_f(u) < 0$ for real objects and $\varphi_f(u) > 0$ for false candidates. For this purpose, we project the feature domain to $[-1, 1]$ with a monotonously decreasing nonlinear function [2]: $\varphi_f(u) = \mathcal{Q}(f(u), d_0^f)$ where $\mathcal{Q}(.) = 1 - 1/f(u)$ if $f(u) < d_0^f$, otherwise: $\mathcal{Q}(.) = \exp(-f(u) + d_0^f) - 1$. Here d_0^f is the object acceptance threshold for feature f, which can be set based on manually annotated training data in a straightforward way. Once we obtained the $\varphi_f(u)$ feature energy subterms, the joint data energy of object u is derived by combining averaging, max and min operators, with using the following strategies. From the $\varphi_f(u)$ primitive terms, first we construct object prototypes. For each prototype we can prescribe the fulfillment of one or many feature constraints whose φ_f-subterms are connected with the max operator in the joint energy term of the prototype, which implements the logical AND in the *inverse fitness* domain. Alternatively, we can use averaging methods. Additionally, several object prototypes can be detected simultaneously in a given image, if the prototype-energies are joined with the min (logical OR) operator. Thus the final object energy term is derived by a logical function, which expresses some prior knowledge about the image and the scene, and it is chosen on a case-by-case basis, examples will be shown in Sec. 4.

The data term of the whole configuration is obtained as the sum of the individual object energies: $\Phi_Y(\omega) = \sum_{u \prec \omega} \varphi_Y(u)$.

3.2   Interaction terms

The interaction terms implement geometric or feature based interaction between different objects and object groups of ω. The following formula is used:

$$\Phi_I(\omega) = \sum_{u \prec \omega} I(u, \omega) + \sum_{u \prec \omega, \psi \in \omega} A(u, \psi) \tag{1}$$

The $I(u, \omega)$ term is derived through classical pairwise interaction constraints, and penalizes overlapping objects within the ω configuration:

$$I(u, \omega) = \sum_{\substack{u, v \prec \omega \\ u \sim v}} \frac{\text{Area}\{R_u \cap R_v\}}{\text{Area}\{R_u \cup R_v\}},$$

where $R_u \subset S$ denotes the pixels covered by the geometric figure of u.

On the other hand, with the $A(u, \psi)$ energies, we can define various constraints between the object group level and the object level of the scene. To measure if an object u appropriately matches to a population segment ψ, we define a distance measure $d_\psi(u) \in [0, 1]$, where $d_\psi(u) = 0$ corresponds to a high quality match. In general, we prescribe that the segments are spatially connected, therefore, we use a constantly high difference factor, if u has no neighbors within ψ w.r.t. relation $\sim$. Thus we derive a modified distance:

$$\hat{d}_\psi(u) = \begin{cases} 1 & \text{if } \nexists v \in \psi \setminus \{u\} : u \sim v \\ d_\psi(u) & \text{otherwise} \end{cases}$$

We define the $A(u, \psi)$ arrangement term of (1) in the following way. We slightly penalize population segments which only contain a single object: with a small $0 < c \ll 1$ constant $A(u, \psi) = c$ iff $\psi = \{u\}$. Otherwise, large $\hat{d}_\psi(u)$ is *penalized* if $u \in \psi$; and *favored* if $u \notin \psi$:

$$A(u, \psi) = \mathbf{1}_{u \in \psi} \cdot \hat{d}_\psi(u) + \mathbf{1}_{u \notin \psi} \cdot (1 - \hat{d}_\psi(u))$$

where $\mathbf{1}_E \in \{0, 1\}$ is an indicator function of event E.

3.3 Optimization

The optimal ω can be obtained by minimizing the previously defined energy function $\Phi(\omega)$. Since the complexity of the problem is exponential, we have proposed an approximating solution, called Multilevel Multiple Birth-Death-Maintenance (M^MBDM) algorithm. This iterative technique extends the well established MBD [4] optimization strategy with an object group management component. The steps are as follows.

I) Initialization: start with an empty population $\omega = \emptyset$, set the birth rate b_0, initialize the inverse temperature parameter $\beta = \beta_0$ and the discretization step $\delta = \delta_0$.

II) Main program: alternate the following three steps:

1. *Birth step*: Visit all pixels on the image lattice S one after another. At each pixel s, with probability δb_0, generate a new object u with center s, random type and random geometric parameters. For each new object u, with a probability

$$p_u^0 = \mathbf{1}_{\omega = \emptyset} + \mathbf{1}_{\omega \neq \emptyset} \cdot \min_{\psi_j \in \omega} \hat{d}_{\psi_j}(u),$$

generate a new ψ empty segment (i.e. object group), add u to ψ and ψ to ω. Otherwise, add u to an existing segment $\psi_i \in \omega$ with a probability

$$p_u^i = (1 - \hat{d}_{\psi_i}(u)) / \sum_{\psi_j \in \omega} (1 - \hat{d}_{\psi_j}(u)).$$

2. *Death step*: Consider the actual configuration of all objects within ω and sort it by decreasing values depending on $\varphi_Y(u) + A(u, \psi)\big|_{u \in \psi}$. For each object u taken in this order, compute $\Delta\Phi_\omega(u) = \Phi_\mathcal{D}(\omega / \{u\}) - \Phi_\mathcal{D}(\omega)$, derive the *death rate* $p_\omega^d(u)$ as

$$p_\omega^d(u) = \Gamma(\Delta\Phi_\omega(u)) = \frac{\delta \exp(-\beta \cdot \Delta\Phi_\omega(u))}{1 + \delta \exp(-\beta \cdot \Delta\Phi_\omega(u))}, \tag{2}$$

and delete object u with probability $p_\omega^d(u)$. Remove empty population segments from ω, if they appear.

3. *Group re-arrangement*: Consider the objects of the current ω population, one after another. For each object u of segment ψ we propose an alternative object u', so that the shape type of u', $\text{tp}(u')$, may be different from $\text{tp}(u)$, and the geometric parameters of u' are derived from the parameters of u by adding zero mean Gaussian random values. The next step is selecting a group candidate for u'. For this reason, we randomly choose a v object from the proximity neighborhood of u ($v \in \mathcal{N}_u(\omega)$), and assign u' to the group of v, denoted by ψ'. Then, we estimate the energy cost of exchanging $u \in \psi$ to $u' \in \psi'$:

$$\Delta\varphi(\omega, u, u') = \varphi_Y(u') - \varphi_Y(u) + I(u', \omega \backslash \{u\}) - I(u, \omega) + A(u', \psi') - A(u, \psi)$$

The *object exchange rate* is calculated using the $\Gamma(.)$ function defined by (2):

$$p_\omega^e(u, u') = \Gamma\Big(\Delta\varphi(\omega, u, u')\Big)$$

Finally with a probability $p_\omega^e(u, u')$, we replace u with u'.

III) Convergence test: if the process has not converged yet, increase β and decrease δ with a geometric scheme, and go back to the birth step.

4 Applications

In this section, we introduce three different applications of the proposed Two-Layer MPP model. In each application, we have to define the domain specific f features and feature integration rules to obtain the $\varphi_Y(u)$ unary terms (Sec. 3.1), and we should define the grouping constraints through the definition of the $d_\psi(u)$ object-segment distance term (Sec. 3.2).

4.1 Built-in area analysis in aerial and satellite images

Analyzing built-in areas in aerial and satellite images is a key issue in several remote sensing applications, among others in cartography, GIS data management and updating, or disaster recovery. Most existing techniques focus on the extraction of individual
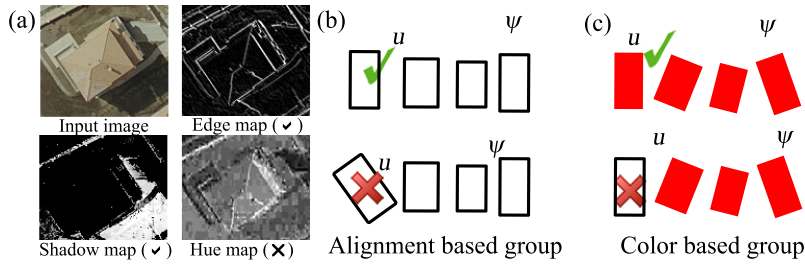
Fig. 1. Building analysis a) Data term features: efficient edge and shadow maps, weak color information b)-c) Favored ($\checkmark$) and penalized ($\times$) sub-configurations within a building group

buildings or building segments from the images [2], however, as pointed out in [5] finding the groups of corresponding buildings (e.g. a residential housing district) has also a great interest in urban environment planning or detecting illegally built objects which do not fit the regular environment. For demonstrating the adaption of the L^2MPP model for urban area analysis, we have chosen a test region of Budapest, Hungary, which is partially displayed in Fig. 4. We assume that the footprint of each building can be approximated by a rectangle or by a couple of slightly overlapping rectangles.

First, we derive the $\varphi_Y(u)$ energy function, which integrates feature-information about roof color, roof edge and shadow [2]. On one hand, *red* roofs can be detected in color images using the hue components of the corresponding pixel values. The color term favors objects which contain mostly roof colored pixels inside the rectangle of u and background pixels around u, features are filling factors in the internal and external regions. For non-red roofs we can rely on the gradient and shadow maps exploiting that under the roof edges 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 (see Fig. 1). In our analysis (Fig. 4) we use two prototypes: the first one prescribes in parallel the edge (eg) and shadow (sh) constraints, while the second one considers the roof color only (co), thus the joint energy is calculated as: $\varphi_Y(u) = \min \left\{ \max \left\{ \varphi_{\text{eg}}(u), \varphi_{\text{sh}}(u) \right\}, \varphi_{\text{co}}(u) \right\}$.

Second, for enabling built-in region segmentation, we construct the object-group distance function $d_\psi(u)$. In our test area, we have observed two different grouping constraints. On one hand, we find several distinct building groups which are formed by regularly aligned, parallel buildings. On the other hand, we can also see a large building group (top left part of Fig. 4(a)), where the orientation of the houses is irregular, but the roof colors are uniform. For this reason, we distinguished two types of groups: if ψ is an alignment based group (Fig. 1(b)), $d_\psi(u)$ is proportional to the angle difference between u and the mean angle within ψ. Otherwise, if ψ is a color group (Fig. 1(c)), $d_\psi(u)$ measures how the color histogram of u matches to the ψ group's expected color distribution, which is set by training samples during the system configuration.

4.2   Traffic monitoring based on remotely sensed Lidar data

Automatic traffic monitoring analysis needs a hierarchical modeling approach: first *individual vehicles* should be detected, then we need to extract *coherent traffic segments*,
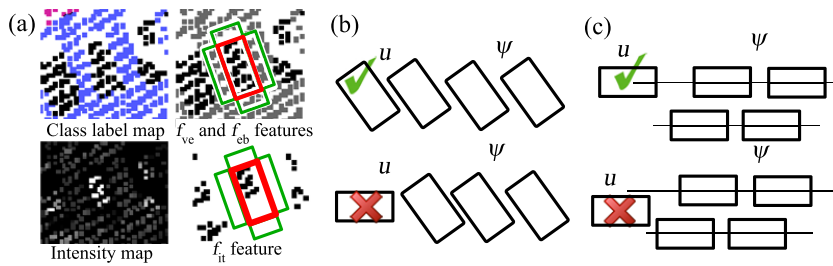
Fig. 2. Traffic monitoring application a) Calculation of the data model features b)-c) Favored ($\checkmark$) and penalized ($\times$) sub-configurations within a traffic segment

by identifying groups of corresponding vehicles, such as cars in a parking lot, or a vehicle queue waiting in front of a traffic light. In [3] a sequential method was introduced relying on airborne LIDAR data, which contains point position, and reflection intensity information. Firstly, the 3D point set is segmented into vehicle and background classes. Then the points with the corresponding class labels and intensity values are projected to the ground plane, where the optimal vehicle and traffic segment configuration is modeled by a rectangle configuration in the projected 2D image.

Three features are exploited here to obtain the φ_Y *unary term* (see Fig. 2). The *vehicle evidence* (f_{ve}) respectively *intensity* (f_{it}) features are calculated as the covering ratio of vehicle classified pixels in the label respectively intensity maps within the proposed rectangle of u. The *external background* (f_{eb}) feature is the rate of background classified pixels in neighboring regions around the proposed u object. Finally the joint data energy of object u is derived as $\varphi_Y(u) = \max(\min(\varphi_d^{\mathrm{it}}(u), \varphi_d^{\mathrm{ve}}(u)), \varphi_d^{\mathrm{eb}}(u))$, where we consider that not all vehicles appear as bright blobs in the intensity map.

The $d_\psi(u)$ distance is the average of two terms: the *first* one is the normalized angle difference between u and the mean angle within ψ (see Fig. 2(b)), *second* with using RANSAC, we fit one or a couple of parallel lines to the object centers within ψ, and calculate the normalized distance of the center of u from the closest line (Fig. 2(c)).

4.3 Automatic optical inspection of printed circuit boards

Automatic optical inspection (AOI) is a widely used approach for quality assessment of Printed Circuit Boards (PCBs). Automated layout-template-free approaches are especially useful for verifying uniquely designed circuits. In the PCBs usually connected groups of similarly shaped and oriented Circuit Elements (CEs) implement a given function, therefore interpretation of the board content need to segment the CE population.

In the considered PCB image data set [1] the CEs can be modeled as bright *rectangles* or *ellipses* surrounded by darker background. To evaluate the contrast between the CEs and the board, we calculate the Bhattacharya [4] distance $d_B(u)$ between the pixel intensity distributions of the internal CE regions and their boundaries. Then the $\varphi_Y(u)$ unary term is derived by $\mathcal{Q}$ mapping of $d_B(u)$ (Sec. 3.1).

Within a CE group, we prescribe that the elements must have similar shape and must follow a strongly regular alignment. Therefore $d_\psi(u) = 1$ if the type of u, $\mathrm{tp}(u)$
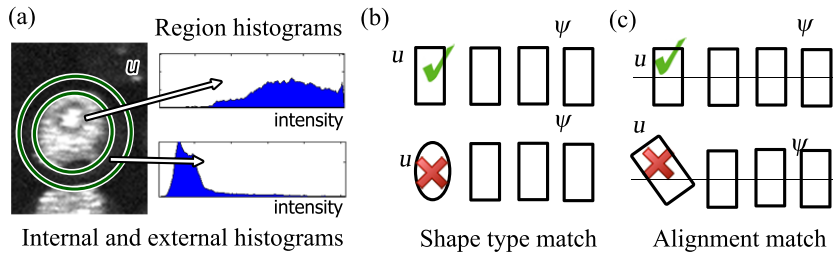
Fig. 3. Circuit inspection a) Data term feature b) Favored ($\checkmark$) and penalized ($\times$) sub-configurations within a CE group, w.r.t. the *shape type match* and *alignment match* constraints

Fig. 4. Built-in area analysis results with the SMPP and the proposed L²MPP approaches, groups are marked with different colors. Errors are annotated "O" refers to object "G" to group artifacts.

is not equal to the type of the ψ group, otherwise $d_\psi(u)$ is the maximum of the angle difference and symmetry distance terms defined in Sec. 4.2 by the traffic monitoring application.

5   Experiments and conclusion

We evaluated our method in real datasets regarding each application, sample results are shown in Fig. 4-6. The parameters of the method were set based on a limited number of training samples, similarly to [2]. For accurate Ground Truth (GT) generation, we have developed an accessory program with graphical user interface, which enables us to manually create and edit a GT configuration of various geometric objects, and assign them to different object groups. The obtained GT configuration can be compared to the output of the algorithm. We have performed quantitative evaluation both at object and at pixel levels, results are shown in Table 1. At object level, we have counted the number of true positive (TP), false positive (FP) and false negative (FN) objects. We have also counted the objects with False Group labels (FG) among the true positive samples, considering GT classification of human observers. To enable automated evaluation, we need to make first a non-ambiguous assignment between the detected and GT object samples, which has been performed with the Hungarian algorithm. At pixel level, we compared the object silhouette masks to the GT mask, and as the harmonic mean of Precision (Pr) and Recall (Rc), we calculated the F-rate (Fr) of the match [2].
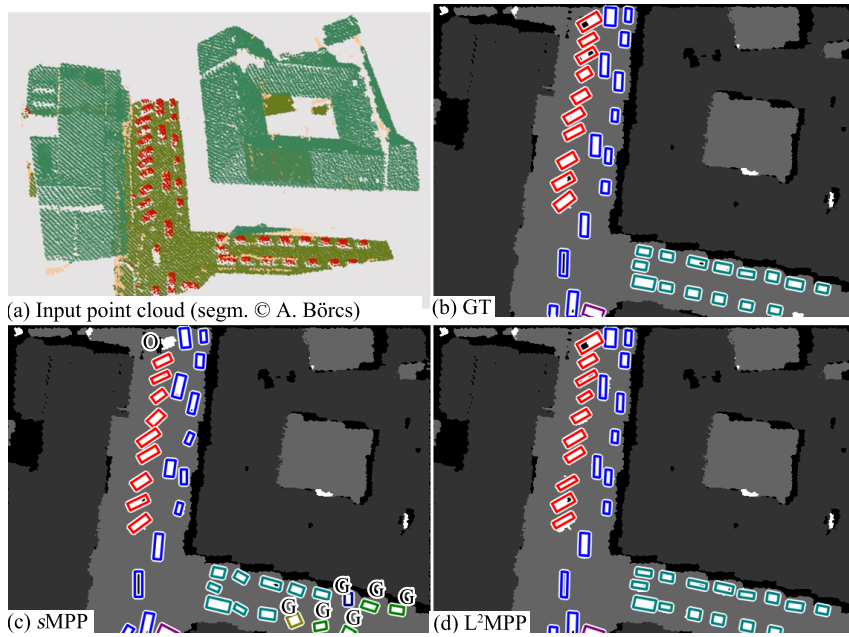
Fig. 5. Traffic monitoring results, "O" refers to object "G" to group level artifacts.
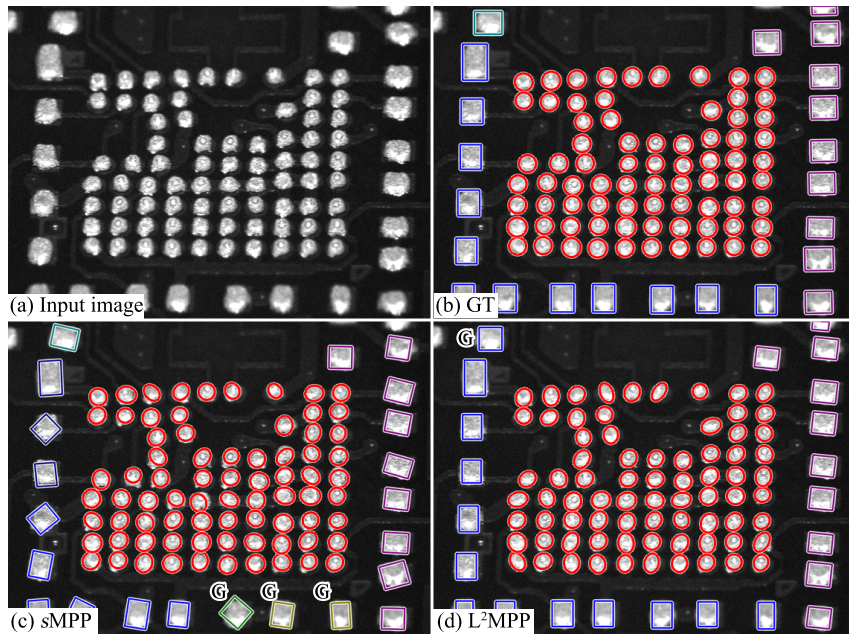

Fig. 6. PCB inspection results, "O" refers to object "G" to group level artifacts.

As a baseline for comparison, we used a sequential technique, which extracts first the object population by a *s*ingle layer MPP model (*s*MPP), using the same unary terms as the proposed L^{2}MPP approach, but regarding the prior terms, only the $I(u, \omega)$ intersection component is considered, similarly to [1, 2]. Thereafter, grouping is performed in post processing by a recursive floodfill-like segmentation of the population. Results of the baseline *s*MPP detection are also displayed in Fig. 4-6 and in Table 1.

We can observe that the introduction of the L^{2}MPP model has resulted in particular gain in the pixel based quality factors (obtained object shapes are more accurate) and notably decreased the objects with False Groups (FG). We note that with using conventional pairwise [6] orientation smoothing terms, it may also be possible to obtain regularly aligned object groups, however, the proposed model offers a higher degree of freedom for simultaneously considering various group level features and exploit interaction between corresponding, but not necessarily closely located objects. As future work, we intend to extend the model for further applications and develop methods for automatic parameter estimation and robustness analysis.

Dataset parameters					Evaluation results							
			Obj	Group		Object & group				Pixel level %		
Applicat.	Input	Resolution	num	num	Method	**TP**	FP	FN	FG	Rc	Pr	**Fr**
Building analysis	Aerial image	0.5 m/pix	44	4	*s*MPP	42	1	3	6	76	75	**76**
					L^{2}MPP	44	1	1	3	79	87	**83**
Traffic monitoring	Lidar points	8 pts/m^2	39	4	*s*MPP	38	0	1	6	82	87	**84**
					L^{2}MPP	39	0	0	0	85	92	**89**
Circuit inspection	AOI image	6 μm/pix	99	4	*s*MPP	98	0	0	3	83	92	**87**
					L^{2}MPP	99	0	0	1	86	98	**92**

Table 1. Object, group and pixel level comparison of the *s*MPP and the proposed L^{2}MPP models

References

1. Benedek, C.: Detection of soldering defects in printed circuit boards with hierarchical marked point processes. Pattern Recognition Letters 32(13), 1535 – 1543 (2011)
2. Benedek, C., Descombes, X., Zerubia, J.: Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics. IEEE Trans. Pattern Anal. Mach. Intell. 34(1), 33–50 (2012)
3. Börcs, A., Benedek, C.: Urban traffic monitoring from aerial Lidar data with a two-level marked point process model. In: International Conference on Pattern Recognition (ICPR). pp. 1379–1382. Tsukuba City, Japan (2012)
4. Descombes, X., Minlos, R., Zhizhina, E.: Object extraction using a stochastic birth-and-death dynamics in continuum. J. Math. Imag. Vision 33, 347–359 (2009)
5. Kovács, A., Szirányi, T.: Orientation based building outline extraction in aerial images. ISPRS Annals of Photogram., Remote Sens. and Spatial Inf. Sci. I-7, 141–146 (2012)
6. Lafarge, F., Gimel'farb, G., Descombes, X.: Geometric feature extraction by a multimarked point process. IEEE Trans. Pattern Anal. Mach. Intell. 32(9), 1597 –1609 (2010)
7. Scarpa, G., Gaetano, R., Haindl, M., Zerubia, J.: Hierarchical multiple Markov chain model for unsupervised texture segmentation. IEEE Trans. Image Proc. 18(8), 1830–1843 (2009)