HIERARCHICAL IMAGE CONTENT ANALYSIS WITH AN EMBEDDED MARKED POINT PROCESS FRAMEWORK

Csaba Benedek

Institute for Computer Science and Control, H-1111, Kende u. 13-17 Budapest, Hungary

ABSTRACT

In this paper we introduce a probabilistic approach for extracting complex hierarchical object structures from digital images. The proposed framework extends conventional Marked Point Process models by (i) admitting object-subobject ensembles in parent-child relationships and (ii) allowing corresponding objects to form coherent object groups. The proposed method is demonstrated in three application areas: optical circuit inspection, built in area analysis in aerial images, and traffic monitoring on airborne Lidar data.

Index Terms— marked point process, hierarchy

1. INTRODUCTION

Nowadays various imaging technologies, from remote sensing data acquisition until microscopic imaging, provide very high resolution visual data. As a result a single digital image may encapsulate multi-scale information from the scene, enabling us to simultaneously analyze the crowds of entities at a macro level, and small details of the individual field objects.

Marked Point Processes (MPP) [1, 2, 3] have recently been widely used for analyzing object populations, however they usually implement a single layer scene model, supporting the extraction of configurations of similar entities such as birds [4], or buildings [5] 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.

Previous attempts for multi-level image understanding followed either region based [6], object based [7, 8] or hybrid [9] approaches. However, the above models were suited to a specific application areas with specific inputs: remotely sensed optical images [6, 9] or Lidar point clouds [8], and Automatic Optical Inspection (AOI) of Printed Circuit Boards (PCB), using μm resolution images [7]. Experiences show that for such complex, application dependent models, the adaption to another application domain is rarely straightforward, needing a significant modeling and implementation work. Following a reverse approach, we introduce in this paper a novel general three-level MPP framework which can handle a wide family of applications. The structure elements and the energy optimization algorithm of the complex model are defined and implemented at the abstract level, while we keep focus on ensuring very simple interfaces to the different applications, enabling efficient domain adaption. Key contributions of the proposed methodology are as follows:

(i) We describe the hierarchy between objects and object parts as a parent-child relationship embedded into the MPP framework. The model of a child is affected by its parent entity, considering geometrical and spectral constraints.

(ii) We partition the (parent) entity population population into object groups, called configuration segments, and extract the objects and the optimal segments simultaneously by a joint energy minimization process. We create adaptive object neighborhoods by segment driven object interactions.

In this paper, we propose a composite three-layer Embedded MPP (EMPP) model, which extends our earlier two-layer approach [10] with embedding the subobject (*child*) layer. We introduce a three-level modification of the Multiple Birth and Death (MBD) optimization algorithm [3, 4], and demonstrate that the proposed technique finds efficient configuration in the increased dimensional populations space. Finally, we show three different applications from the remote sensing and AOI domains, which can use the advantages of the EMPP model.

2. PROBLEM FORMULATION AND NOTATIONS

To model the hierarchical scene content, the proposed Embedded Marked Point Process (EMPP) framework has a multilayer structure, as shown in Fig. 1. At the top, we have a super node, called the *population* or the *configuration*, which is a high-level model of the imaged scene. The population consists of an arbitrary number of object groups, where each group is a composition of one or many super (or parent) objects. Finally, the super objects may encapsulate any number of subobjects (or child objects).

The input of EMPP is an image over a pixel lattice S. Let u be a parent object candidate of the scene, which is represented by a plane figure from a preliminary fixed shape library, such us ellipses and rectangles. For each object, we define the coordinates of a reference point, the global orien-

This work was supported by the Government of Hungary through a European Space Agency (ESA) Contract under the Plan for European Cooperating States (PECS), and by the Hungarian Research Fund (OTKA #101598).

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

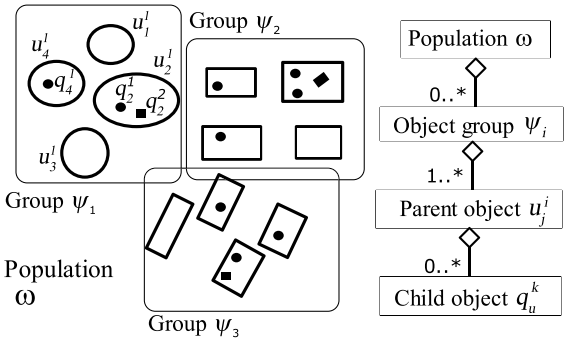

Fig. 1. A sample EMPP population with three object groups, and various object shapes both at parent and child layers.

tation, and further geometric parameters such as axes or side lengths. Each parent object u may contain a set of child objects $Q_u = \{q_u^1 \dots q_u^{m(u)}\}$ where $m(u) \leq m_{\max}$, and each child is a sample from the previously defined geometric figure library. $Q_u = \emptyset$ marks that u has no child.

We continue with the object grouping process. A given population ω is a set of k *object groups* or (also referred later as *configuration segments*), $\omega = \{\psi_1, \dots, \psi_k\}$, where each group ψ_i ($i = 1 \dots k$) is a configuration of n_i objects: $\psi_i = \{u_1^i, \dots, u_{n_i}^i\}$. Here we prescribe that $\psi_i \cap \psi_j = \emptyset$ for $i \neq j$, while the k set number and n_i set cardinality values may be arbitrary integers. We mark with $u \prec \omega$ if u belongs to any ψ in ω, and let $\mathcal{N}_u(\omega)$ be the neighborhood of $u \prec \omega$, using a $u \sim v$ proximity relation. Finally, we denote by Ω the space of all the possible global configurations, considering that each population $\omega \in \Omega$ may include any number of groups composed of any number of objects and child objects.

3. EMPP ENERGY MODEL

The EMPP framework uses an energy function $\Phi(\omega)$, which can evaluate each $\omega \in \Omega$ configuration based on the observed data and prior knowledge. Therefore, the energy can be decomposed into a unarY term (Y) and an Interaction term (I): $\Phi(\omega) = \Phi_Y(\omega) + \Phi_I(\omega)$, and the optimal $\widehat{\omega}$ configuration is obtained by minimizing $\Phi(\omega)$ over Ω.

3.1. Unary object appearance terms

We use an energy term $\varphi_Y(u)$ which characterizes u depending on the local image data, but independently of other objects. $\varphi_Y(u)$ is decomposed into a parent term $\varphi_Y^p(u)$ and for each child object q_u a child term $\varphi_Y^c(u, q_u)$. The child term may depend on both the image and the geometry of the parent (e.g. an intensity histogram within the parent region).

At *parent level*, first we define different $f(u)$ fitness features, which evaluate an object hypothesis for u in the image. Then we construct $\varphi_{Y,f}^p(u)$ *data driven* energy subterms for each feature f, so that we project the feature domain to $[-1, 1]$ with a monotonously decreasing nonlinear $\mathcal{M}(f, d_0^f)$ function [5]: $\varphi_f(u) = \mathcal{M}(f(u), d_0^f)$ where $\mathcal{M}(.) = 1 - 1/f(u)$ if $f(u) < d_0^f$, otherwise: $\mathcal{M}(.) = \exp(-f(u) + d_0^f) - 1$. d_0^f is the object acceptance threshold for feature f, which can be set based on annotated training data in a straightforward way.

The $\varphi_Y^p(u)$ parent energy of u is calculated from the $\varphi_{Y,f}^p(u)$ subterms. *First* we construct object prototypes, prescribing the fulfillment of one or many feature constraints, whose φ_f-subterms are connected with the $\max$ operator in the prototype energy term (logical AND in the negative likelihood domain). Several object prototypes can be detected simultaneously in a given image, if the prototype-energies are joined with the $\min$ (logical OR) operator. Thus $\varphi_Y^p(u)$ is derived by a logical function, which expresses application dependent knowledge, chosen on a case-by-case basis.

The construction of the *child's unary term* $\varphi_Y^c(u, q_u)$ is based on similar principles: it is obtained using different features mapped by the $\mathcal{M}$ function. The complete unary term of u is the sum of the parent level terms and the child level terms: $\varphi_Y(u) = \varphi_Y^p(u) + \sum_{q_u \in Q_u} \varphi_Y^c(u, q_u)$. 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 constraints between the elements of ω:

$$\Phi_I(\omega) = \sum_{\substack{u,v \prec \omega \\ u \sim v}} I(u, v) + \sum_{u \prec \omega} J(u, Q_u) + \sum_{u \prec \omega, \psi \in \omega} A(u, \psi).$$

The $I(u, v)$ terms provide classical pairwise interaction constraints, in our later examples they penalize overlapping objects within the ω configuration: $I(u, v) = \frac{\text{Area}\{u \cap v\}}{\text{Area}\{u \cup v\}}$.

The $J(u, Q_u)$ terms model interactions between the corresponding parent a child objects, and interactions between different child objects corresponding to the same parent. For example, we can prescribe that the children of a given parent (i.e. *siblings*) should not overlap with each other, and not overhang the parent, or the siblings should have same shape, similar color, size, orientation etc.

Finally, with the $A(u, \psi)$ energies, can define various constraints between the object group level and the (parent) 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 constant high difference factor, if u has no neighbors within ψ w.r.t. relation $\sim$, so that $d_\psi(u) \stackrel{\text{DEF}}{=} 1$, if $\nexists v \in \psi \backslash \{u\}$: $u \sim v$.

By definition of $A(u, \psi)$, we slightly penalize population segments which only contain a single object: with a small $0 < c$ constant $A(u, \psi) = c$ iff $\psi = \{u\}$. For segments with multiple objects, large $d_\psi(u)$ distances are penalized within

a group, but they are favored between groups, i.e. if $u \in \psi$: $A(u,\psi) = d_\psi(u)$; if $u \notin \psi$: $A(u,\psi) = 1 - d_\psi(u)$.

4. OPTIMIZATION

To estimate the optimal object configuration, we have proposed a three-level modification of the MBD algorithm [3, 4]:

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

Main program: alternate the following three steps:

• *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 and random geometric parameters. For each new object u, *either* generate a new ψ empty configuration segment, add u to ψ and ψ to ω; or add u to an existing segment from its neighborhood, as detailed in [10].

• *Death step*: Consider the actual configuration of all objects within ω and sort it by decreasing values depending on $\varphi_Y(u) + J(u, Q_u) + A(u,\psi)\big|_{u\in\psi}$. 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* $d_\omega(u)$ as

$$d_\omega(u) = \Gamma(\Delta\Phi_\omega(u)) = \frac{\delta \exp(-\beta\cdot\Delta\Phi_\omega(u))}{1+\delta \exp(-\beta\cdot\Delta\Phi_\omega(u))},$$

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

• *Group re-arrangement*: Propose randomly group merge, group split and vehicle re-clustering moves. For each proposed move M, calculate the corresponding energy cost $\Delta\Phi^{\mathrm{M}}_\omega$, and apply the move with a probability $\Gamma(\Delta\Phi^{\mathrm{M}}_\omega)$.

• *Child Maintenance*: For each $u \prec \omega$ object: (i) add new child objects to Q_u randomly (ii) sort Q_u by decreasing values depending on the $\varphi^c_d(u, q_u)$ values (iii) for each child object $q_u \in Q_u$ taken in this order, compute the child removal rate $d^c_u(q_u)$ similarly to the parent level, but considering only the child level unary and interaction terms. (iv) remove q_u from Q_u with a probability $d^c_u(q_u)$.

Test: if the process has not converged yet, increase β and decrease δ with a geometric scheme, and go back to *Birth*.

5. APPLICATIONS

In this section, we introduce three different applications of the proposed EMPP model. In each application, we have to define the domain specific f features and feature integration rules to obtain the parent level $\varphi^p_Y(u)$ and child level $\varphi^c_Y(u)$ unary terms (Sec. 3.1), we should set up the $J(u, Q_u)$ parent-child interaction rules and define the grouping constraints through the definition of the $d_\psi(u)$ object-segment distance (Sec. 3.2).

5.1. Built-in area analysis in aerial and satellite images

Model elements: parent objects are rectangular buildings or building parts. Child objects are tall structure elements on the roofs, such us chimneys and satellite dishes, also modeled by rectangles. Configuration segments are groups of corresponding buildings (eg. residential housing district, Fig. 2a).

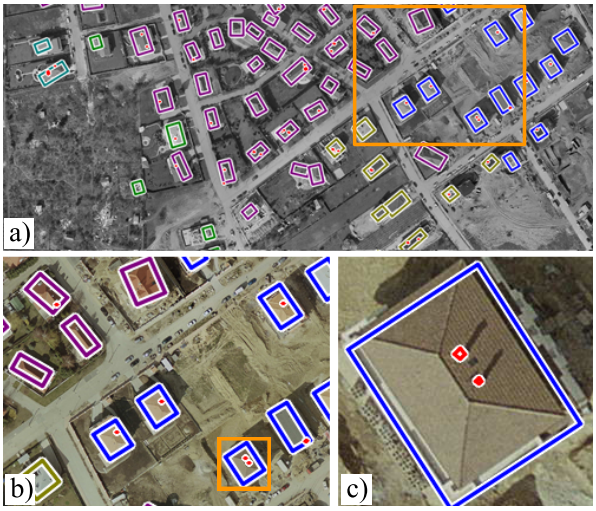

Fig. 2. Results of built-in area analysis, displayed at three different scales. Building groups are distinguished with different colors (purple: *red roofs' district*, others: orientation based groups); red markers denote the detected chimneys

Parent unary terms (φ^p_Y): two object prototypes, based on features prescribing either high image gradients under building edges and shadows next to the buildings; or salient (typically red) roof colors separable from the background [5].

Child unary terms (φ^c_Y): chimneys et al. differ from the roof in color, and cast shadows on the roof (Fig. 2c).

Parent-child terms $J(u, Q_u)$: Non-overlapping siblings with similar orientation. Children figures are encapsulated by the parent rectangles (Fig. 2c).

Object-segment distance $d_\psi(u)$: groups are formed either based on similar (salient) roof color, or based on similar orientation [10]. $d_\psi(u)$ is the normalized color/orientation distance between u and the mean value within ψ (Fig. 2a,b).

Application: urban environment planning or detecting illegally built objects which do not fit the regular environment. Detecting illegal or irregular chimneys.

5.2. Traffic monitoring based on aerial Lidar data

Preprocessing: the Lidar point set is segmented into vehicle and background classes, and the labels and the intensity values of the points are projected to the ground plane [8].

Model elements: parent objects are vehicles, child objects are windshields (both rectangles). Configuration segments are formed by corresponding vehicles, such as cars in a parking lot, or a vehicle queue in front of a traffic light (Fig. 3a).

Parent unary terms (φ^p_Y): covering ratio of vehicle points within u's rectangle based on geometric and intensity based separation. Covering ratio of background points around u [8].

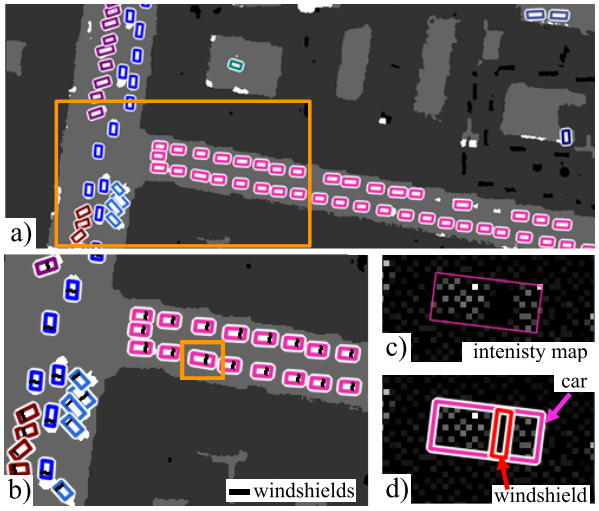

Fig. 3. Results of traffic analysis: a) cars and traffic segments b) selected region with the detected windshields c) intensity map of a selected car, d) detection result for c)

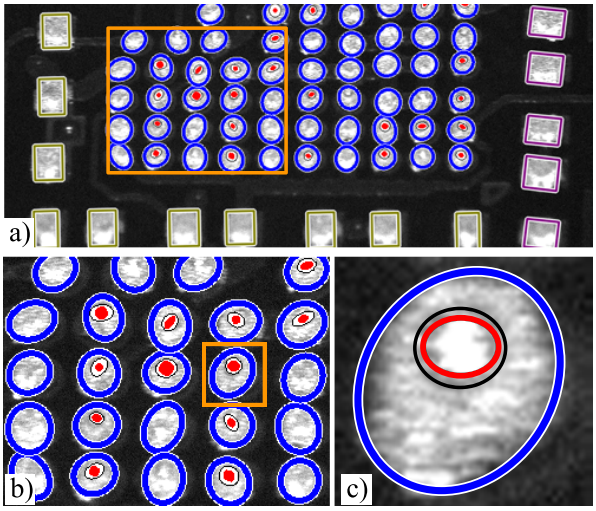

Fig. 4. Results of PCB analysis. CEs are grouped by shape and orientation, scoops are extracted within the CEs

Child unary terms (φ_Y^c)*:* due to their glass material, windshield regions are composed of missing points, or points with salient low intensities within the car's rectangle (Fig. 3c,d).

Parent-child terms $J(u, Q_u)$*:* the windshield is encapsulated by the car's figure, and the orientation is perpendicular to the car's main axis (Fig. 3b,d).

Object-segment distance $d_\psi(u)$*:* orientation distance between u and the mean orientation within $\psi(u)$. For correct grouping of a vehicle queue in a curved road, orientation can be calculated relatively to the closest road side as in [8].

Application: automatic traffic monitoring and control, surveillance. Windscreen configuration can be used for classifying vehicle types, estimating vehicle direction (Fig. 3b).

5.3. Automatic optical inspection of printed circuit boards

Goal: shape extraction and grouping of Circuit Elements (CEs) in uniquely designed PCBs, detecting special soldering errors called *scooping* [7].

Model elements: parent objects are CEs of various shapes, child objects are scoops, modeled by pairs of concentric ellipses [7]. Groups are formed by CEs which likely have similar functionalities [10] (Fig 4a).

Parent unary terms (φ_Y^p)*:* CEs have bright figures surrounded by darker background, used feature is the Bhattacharya [3] distance between the pixel intensity distributions of the internal CE regions and their boundaries.

Child unary terms (φ_Y^c)*:* dominant brightness value of the scoop central region, contrast between the central region and the median ring, resp. the median ring and the external ring (Fig 4c) [7].

Parent-child terms $J(u, Q_u)$*:* each parent CE may have at most one child, whose figure cannot overhang its parent.

Object-segment distance $d_\psi(u)$*:* within a CE group, the elements must have similar shape and must follow a strongly regular alignment. Therefore $d_\psi(u) = 1$ if the type of u, is not equal to the type of the ψ group, otherwise $d_\psi(u)$ is angle difference between u and the mean value in ψ.

Application: automatic interpretation and quality assessment of uniquely designed PCBs by AOI systems.

6. EXPERIMENTS AND CONCLUSION

We tested our method on real datasets for each application, sample results are shown in Fig. 2-4. The parameters of the method were set based on a limited number of training samples [10]. For evaluation, we have counted the number of true positive, false positive and false negative objects both at parent and child levels, and calculated the F-rate of detection (harmonic mean of precision and recall). We have also counted the objects with False Group labels among the true positive samples, using classification of human observers.

The *built-in area* dataset contained 69 buildings with 66 chimneys or antennas. Detection rate was 95% at parent, 73% at child level, Correct Grouping Rate (CGR) was 91%.

In the *traffic* dataset, we measured a 92% detection rate and a 93% CGR among the 170 observable vehicles, the detected windshild position was in 82% correct.

Finally in the *PCB* dataset, all the 98 circuit elements were correctly detected and classified, while the child level scooping detection rate was 89%.

The above experiments confirm at a proof-of-concept level, that the proposed EMPP model is able to handle real world tasks from significantly different application domains, providing an expandable Bayesian framework for multi-level image content interpretation. Future work will focus on robustness analysis and automated parameter estimation.

7. REFERENCES

[1] F. Chatelain, X. Descombes, F. Lafarge, C. Lantuejoul, C. Mallet, R. Minlos, M. Schmitt, M. Sigelle, R. Stoica, and E. Zhizhina, *Stochastic geometry for image analysis*, Digital Signal and Image Processing. Wiley-ISTE, 2011.

[2] A. Gamal Eldin, X. Descombes, and J. Zerubia, "A novel algorithm for occlusions and perspective effects using a 3D object process," in *IEEE International Conf. on Acoustics, Speech and Signal Processing*, Prague, Czech Republic, 2011, pp. 1569 – 1572.

[3] X. Descombes, R. Minlos, and E. Zhizhina, "Object extraction using a stochastic birth-and-death dynamics in continuum," *Journal of Mathematical Imaging and Vision*, vol. 33, pp. 347–359, 2009.

[4] S. Descamps, X. Descombes, A. Bechet, and J. Zerubia, "Flamingo detection using a multiple birth and death process," in *IEEE International Conf. on Acoustics, Speech and Signal Processing*, Las Vegas, NV, 2008, pp. 1113–1116.

[5] C. Benedek, X. Descombes, and J. Zerubia, "Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 33–50, 2012.

[6] G. Scarpa, R. Gaetano, M. Haindl, and J. Zerubia, "Hierarchical multiple Markov chain model for unsupervised texture segmentation," *IEEE Trans. on Image Processing*, vol. 18, no. 8, pp. 1830–1843, 2009.

[7] C. Benedek, O. Krammer, M. Janóczki, and L. Jakab, "Solder paste scooping detection by multi-level visual inspection of printed circuit boards," *IEEE Trans. on Industrial Electronics*, vol. 60, no. 6, 2013.

[8] A. Börcs and C. Benedek, "Urban traffic monitoring from aerial LIDAR data with a two-level marked point process model," in *International Conference on Pattern Recognition (ICPR)*, Tsukuba City, Japan, 2012, pp. 1379–1382, Extended version submitted to IEEE Trans. Geosci. Rem. Sens.

[9] J. Porway, Q. Wang, and S. C. Zhu, "A hierarchical and contextual model for aerial image parsing," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 254–283, 2010.

[10] C. Benedek, "A two-layer marked point process framework for multilevel object population analysis," in *International Conference on Image Analysis and Recognition (ICIAR)*, vol. 7950 of *Lecture Notes in Computer Science*, pp. 160–169. Póvoa de Varzim, Portugal, 2013.