FACADE INTERPRETATION USING A MARKED POINT PROCESS

Susanne Wenzel, Wolfgang Förstner

Department of Photogrammetry, Institute of Geodesy and Geoinformation, University of Bonn, Germany
susanne.wenzel@uni-bonn.de, wf@ipb.uni-bonn.de

Commission III, WG III/4

KEY WORDS: Facades, Image Interpretation, Marked Point Processes, rjMCMC, Simulated Annealing

ABSTRACT:

Our objective is the interpretation of facade images in a top-down manner, using a Markov marked point process formulated as a Gibbs process. Given single rectified facade images, we aim at the accurate detection of relevant facade objects as windows and entrances, using prior knowledge about their possible configurations within facade images. We represent facade objects by a simplified rectangular object model and present an energy model, which evaluates the agreement of a proposed configuration with the given image and the statistics about typical configurations, which we learned from training data. We show promising results on different datasets and provide a qualitative evaluation, which demonstrates the capability of complete and accurate detection of facade objects.

1. INTRODUCTION

Our objective is the interpretation of facade images by combining evidence from bottom up and prior knowledge from top-down. Given single rectified facade images, we aim at the accurate detection of relevant facade objects as windows, entrances, and balconies.
Object detection in a sliding window approach yields reliable results, but lacks on accuracy and completeness. Therefore, prior knowledge about the arrangement of facade elements needs to be introduced. This might be achieved in different ways. Already on the pixel level, Markov random fields are used to model prior knowledge about the objects layout (Čech and Šára, 2008). On the entity level, rules on the facade's attributes (Ripperda, 2008; Müller et al., 2007) or the object's neighbourhood relations (Tylecek and Sara, 2010) are used. Related to the last one, we incorporate prior knowledge about typical arrangements of facade elements. We denote a set of such elements, describing the facade, a configuration of facade objects. We evaluate spatial interactions between neighbouring objects, such as alignment, size differences, or distances and use them to enforce the detection results to configurations, which fit typical facades. Typical configurations might by modelled manually, but this inherently lacks in generality, needs a huge amount of rules, increasing with the number of objects classes, and needs to be well tuned to each type of facade. We propose to learn spatial interactions for each combination of neighbouring object classes from training images. This might be restricted to a certain type of facades.

The task is to combine evidence from bottom-up, for each single object proposal, given by a suited classifier, with prior knowledge from top-down, given by learned configuration properties. Markov random fields (MRF) are widely used in image processing for such tasks. Instead, we use marked point processes (MPP), which can be seen as natural extension of MRFs, not only due to the freedom of the underlying graph structure and its dimensionality, but also due to the way they handle parametric objects. MPPs have shown competitive results in several object detection problems (Lafarge and Gimel'farb, 2008; Lafarge et al., 2010b; Börcs and Benedek, 2012; Bredif et al., 2013; Ortner et al., 2007, 2008; Tournaire et al., 2007, 2010; Verdie and Lafarge, 2013). While MRFs are restricted to labelling problems in static graphs, MPPs, as introduced by Baddeley and Lieshout (1993), handle parametric objects within dynamic graphs. A MPP is a random variable whose realizations are configurations of parametric objects. Thus, we use the term point synonymously for objects, represented by points in $\mathbb{R}^2$, attached with additional parameters, e.g., width and height. The number of objects is a random variable, too, and needs not to be defined beforehand in contrast to modelling as MRF. Further, we are able to incorporate complex spatial interactions. Typically, this is designed manually (Ortner et al., 2007, 2008; Lafarge et al., 2010b; Börcs and Benedek, 2012; Verdie and Lafarge, 2013). We propose to learn properties of typical configurations in terms of spatial interactions of neighbouring objects from training images. This way, we are able to combine the initial belief from bottom-up image categorization and the prior knowledge we have about typical configurations of facade objects.
The whole process is specified by three key elements:

- The object model. We model objects in images as axis-parallel rectangles, thus, points attached with marks for width and hight. Additionally, each point is given a class label. This leads to the definition of the configuration space, which is the space of all possible arrangements of parametrized objects, fulfilling certain semantic and geometric relations.
- The energy. We model the MPP as Gibbs process, thus, its distribution is given in terms of an energy. The energy validates the quality of a configuration according to the image content and the spatial interaction of objects in a limited neighbourhood.
- An optimization method. We aim at minimizing the energy, which is complex and of varying dimensionality. Therefor, we sample the configuration space by rjMCMC coupled with simulated annealing.

1.1 Related Work

Our field of interest is facade interpretation, thus, we will give a brief synopsis of recent work in this field and review related work in the application of MPPs.
Main objects of interest, when dealing with facade images, are windows. At mid-level, many works deal with window detection (Lee and Nevatia, 2004; Ali et al., 2007; Recky and Leberl, 2010)

or exploit repetitive structures to capture grids of windows (Wenzel et al., 2007; Tylecek and Sara, 2010; Wendel et al., 2010; Park et al., 2010). In contrast to object detection, pixelwise labelling is used to yield a facade segmentation. Fröhlich et al. (2010) combine of a strong pixelwise classification, using random forests (RF), with an unsupervised segmentation. Teboul et al. (2010) use the pixelwise classification as low-level input for a shape grammar. They formulate a constrained generic shape grammar to express special types of buildings and train a RF classificator to determine the relationship between semantic elements of the grammar and the observed image support. Martinović et al. (2012) start from an oversegmentation of a facade and produce probabilistic interpretations for each segment, using Recursive Neural Networks, which they merge with the output of a specialised facade component detectors formulating a MRF. Other works deal with 3D information, either as additional input data or as intermediate or final results, respectively. To model the structure of facade objects, they use grammars, too (Werner and Zisserman, 2002; Dick et al., 2004; Alegre and Dellaert, 2004; Ripperda, 2008). Parameters of the grammar are estimated either with MCMC or directly determined during a recursive splitting and merging procedure. But, all these grammar approaches lack the restricted domain of valid facades that the grammar rules are designed for.
Recently, Cohen et al. (2014) propose to use dynamic programming for efficiently parsing facade images. Instead of using rules of a grammar, they sequentially parse the individual classes according given constraints, which make the process more flexible to different types of facades.
While Teboul et al. (2010) and Cohen et al. (2014) rely on strong architectural constraints, Martinović et al. (2012) and Koziński et al. (2015) avoid strong prior assumption about the structure of facades and introduce weak architectural knowledge, which enforces the final reconstruction to be architecturally plausible and consistent. Koziński et al. (2015) propose to express architectural prior knowledge into a set of hierarchical rules over different semantic classes that specify, which pairs of classes can be assigned to pairs of vertically- and horizontally-adjacent pixels. They transfer those rules into the structure of a MRF. They handle occlusions, and therefore, are able to recover partly occluded facades and to infer their structure.

In our work, we are dealing with MPPs, which were introduced by Baddeley and Lieshout (1993) to the field of stochastic geometry. In conjunction with reversible jump MCMC methods, introduced by Green (1995) and Geyer and Møller (1994), they became popular for several tasks of image processing and interpretation.
To the best of our knowledge, around 2001 the group at INRIA around Josaine Zerubia and Xavier Descombes, start to introduce the topics of stochastic geometry and point process theory to the field of image processing in terms of structure extraction and object detection. They aim at the detection of buildings and road networks in digital aerial images (Garcin et al., 2001; Descombes et al., 2001) or rectangular road markings (Tournaire et al., 2007). Ortner et al. (2003) continue the work on detecting parametrized objects and introduce a proposition kernel that allows to sample objects in the neighbourhood of existing objects. Ortner et al. (2007, 2008) use this technique for the extraction of building footprints from altimetric data in dense urban areas and the extraction of road networks and buildings from digital elevation models, which is done again by Tournaire et al. (2010) in a more efficient way to speed up the process. Lafarge et al. (2010a) extend this work to 3D and propose to reconstruct building from DSM based on a library of 3D models. These model suffer from a lack of generality, they aim at specific applications and the complexity of interactions between the objects does not generalize to another application. Most of them rely on many tunable parameters. Lafarge et al. (2010b) propose a more generalized MPP called multi-marked point process, which can be applied to a large range of applications without changing the underlying model. Verdie and Lafarge (2013) build on former approaches, but aim at their large-scale applications by introducing an efficient parallelization scheme. Chai et al. (2012) propose an hybrid representation of MRFs and MPPs to represent both, low-level information and high-level knowledge to provides a structure-driven approach for detecting buildings in aerial images. Lafarge and Gimel'farb (2008) and Lafarge et al. (2010b) provide a general model for extracting different types of geometric features from images, as line, rectangles and disks and propose jump-diffusion dynamic as alternative to the commonly used rjMCMC sampling with simulated annealing, cf. Sec. 2.5. Jump-diffusion adds to the reversible jumps of the MCMC process a stochastic diffusion dynamic within each continuous subspace. At high temperature of simulated annealing the diffusion performs large random steps to avoid trapping into local optima while at low temperature it acts as gradient descent. Jump-diffusion is an interesting development to speed up convergence tremendously and to increase the accuracy of the final solution. But, it assumes the energy landscape to be smooth near the optimum and an energy, for which we can evaluate the gradient efficiently. Both is not given in our application: the energy landscape is rough and the energies gradient can not be evaluated analytically. We may obtain the gradient from numerical differentiation, but this is slow and contradicts the desired speed-up. Therefrom we do not use jump-diffusion.

We are aware of two works, dealing with MPPs in the context of facade image interpretation. Burochin et al. (2014) formulate a stochastic process to sample rectangles, in order to detect openings in facades as indicator for the classification of blind facades, As the detection results are dedicated as indicator for the classification of blind facades, beside other features, they obviously do not require complete and precise detections.
Wang et al. (2015) aim at the detection of window grids in images using a Marked Point Processes. They propose structure-driven sampling in order to yield the assumed grid structure of windows and use an energy formulation whose data term depends on a probability map given by a pixelwise classification and whose prior term consist on a repulsive term scoring interacting rectangles in terms of their horizontal and vertical distance. Their results are not convincing compared to state of the art results, but to the best of our knowledge, this is the first approach dealing with marked point processes in the context of facade image interpretation.

1.2 Marked Point Processes

MPPs are statistical models for the analysis of observed patterns of points represented by the location of objects. They are of special importance in stochastic geometry for handling random sets of objects.
We call a set of objects $\mathcal{X} = \{\boldsymbol{x}_1 \dots \boldsymbol{x}_n\}$ a configuration. The point process $\underline{\mathcal{X}}$ is a random variable, whose realization $\mathcal{X}$ is a configuration of objects. The theory of MPPs allows to define a probability density $f(\mathcal{X})$ on configurations of objects. Therefrom, we may identify the most probable configuration by maximizing this density. We model our objects as points in $\mathcal{S} \subset \mathbb{R}^2$, attached with additional parameters from an arbitrary space $\mathcal{M}$, called marks. Then, $f(\mathcal{X})$ is the density of a point process, with respect to an underlying Poisson process with intensity μ that refers to the product space $\mathcal{S} \times \mathcal{M}$. To get a more detailed introduction into point processes, we strongly recommend Baddeley (2007).

Markov point processes (Baddeley and Lieshout, 1993) are defined in terms of local interactions between points of the configuration. We use the symbol $\sim$ to denote neighbouring objects. It is a symmetric and reflexive relation, thus, for $\boldsymbol{x}_i \sim \boldsymbol{x}_j \Leftrightarrow \boldsymbol{x}_j \sim \boldsymbol{x}_i \; \forall \boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{S}$, $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are said to be neighbours. The neighbourhood of an object $\boldsymbol{y}$ contains all its neighbours $\mathrm{Ne}(\boldsymbol{y}) = \{\boldsymbol{x} \in \mathcal{S} : \boldsymbol{x} \sim \boldsymbol{y}\}$. Given the Markov property and using the Hammersley-Clifford theorem, we model the density of the Markov marked point process $\underline{\mathcal{X}}$ as finite Gibbs process. Thus, we express $f(\cdot)$ in terms of a Gibbs energy $U(\mathcal{X})$, such that $f(\mathcal{X}) = \frac{1}{Z} e^{-U(\mathcal{X})}$. Instead of maximising $f(\mathcal{X})$ we minimize the energy $U(\mathcal{X})$. The advantages are obvious. We get rid of the normalizing constant Z. Further, the energy needs not be probabilistic. It is a matter of design to express the fitness of the configuration according the given image and expected configurations, and we will introduce our energy formulation in the next section.

2. ENERGY MODEL FOR FACADE INTERPRETATION

2.1 The Object Model

We model an image as continuous bounded set $\mathcal{S} = [0, I_R] \times [0, I_C] \subset \mathbb{R}^2$, using the number of rows I_R and columns I_C of the image. Objects in the image, we search for, are modelled as axis-parallel rectangles $\boldsymbol{x} = [x, y, w, h, c]$, with centre point $[x, y] \in \mathcal{S}$ attached with marks $[w, h, c] \in \mathcal{M} \subset \mathbb{R} \times \mathbb{R} \times \mathbb{N}$ for width, height, and class. As classes we consider $c \in \{1, 2, 3\}$, standing for window, entrance, and balcony.
We represent the image interpretation as unordered set of of objects $\mathcal{X} = \{\boldsymbol{x}_1 \dots \boldsymbol{x}_n\}$, using $n = N(\mathcal{X})$, the number of objects in the configuration. We consider the MPP $\underline{\mathcal{X}}$ to determine the configuration that best describe the image.

2.2 Energy Formulation

The energy should allow us to evaluate spatial interactions of objects and the consistency of single objects concerning the given image. We express the energy as sum of four terms

$$U(\mathcal{X}) = U_{\text{data}}(\mathcal{X}) + \lambda_1 U_{\text{geom}}(\mathcal{X}) + \lambda_2 U_{\text{conf}}(\mathcal{X}) + \lambda_3 U_{\text{num}}(\mathcal{X}) \tag{1}$$

where U_{data} is the unary energy or the data term, which expresses the local energy of all single objects in the current configuration $\mathcal{X}$. In our case, given the output of a classifier, it measures how well the objects fits the image content. This energy will be designed, such that attractive objects contribute negative energies, whereby we call an object attractive if its a posteriori probability is high. We denote U_{geom} the geometric energy and U_{conf} the configuration energy, which are given by prior knowledge from training data about typical geometry of objects or object pairs. The former expresses an additional unary energy, in terms of the objects size and location. The latter expresses the energy of neighbouring objects given by their spatial interaction. The structure of prior energies will be designed, such that configuration, not fitting the learned configuration statistics, get positive energies, thus, get punished. Finally, U_{num} represents an additional prior on the number of each classes objects. In contrast to the basic idea of a point process, which is guided by the reference density of the Poisson process, whose intensity reflects the accepted mean number of objects, this is an alternative model for the prior. Due to the different mean numbers of each classes objects, one underlying reference measure is not suited to reflect these numbers, wherefrom we decide to introduce this as a more adequate prior. Finally, the single terms are weighted relative to each other by positive constants λ_i. We detail each term in the next subsections.

2.3 Data Energy

We use the output of a classifier to capture evidence from bottom-up, from which we obtain an initial belief about its class. Therefor, we need a classifier, which yields a reliable estimate of the posterior probability $P(c \mid \boldsymbol{x})$ for the class, given the image data. We use an import vector machine (IVM) classifier (Zhu and Hastie, 2005; Roscher et al., 2012a,b) that was shown to get a state of the art classification performance. It is a discriminative classifier, therefore, usually ensures better discriminative power than generative models and it produces classwise probabilities for test samples. We train the IVM from given annotated training images. As features we use pairs of adjacent line segments (Wenzel and Förstner, 2012; Ferrari et al., 2008), from which we have learned representative shapelets from training images, again. The descriptor of an image patch, then, is given as histogram of shapelets, which serves as input for the classification.

The data term uses the output of the classifier to assign an energy to each object of a configuration. Thereby, we consider an object, $\boldsymbol{x} \in \mathcal{X}$, having class label $c \in \{1 \dots C\}$, attractive if its a posteriori probability $P(c \mid \boldsymbol{x})$ is large, preferably near 1. Vice versa, such objects should contribute a local unary energy $\Phi(\boldsymbol{x}) < 0$ to the overall energy. Therefrom, exploiting the Markov property, the overall data energy is given by the sum of each objects local energy

$$U_{\text{data}}(\mathcal{X}) = \sum_{\boldsymbol{x}_i \in \mathcal{X}} \Phi(\boldsymbol{x}_i) \;, \tag{2}$$

whereby, we use the reward function

$$\Phi(\boldsymbol{x}) = -\log_C P(c \mid \boldsymbol{x}) - 1, \quad \Phi \in [\infty, -1] \tag{3}$$

to express the object's local energy. Please note, the number of classes C as basis of the logarithm. This way, we ensure probable samples to get a negative local energy, while objects with rather uncertain class labels or objects even not belonging to the given class, thus, having a posteriori probability below $1/C$, get positive local energy. Nevertheless, the decrease of local energy is bounded, which is important to ensure stability of the Markov chain.
Obviously, the overall data energy might decrease infinitely by superimposing infinite copies of an attractive object $\boldsymbol{x}$. We avoid this by including a strong penalty term for overlapping objects in the pairwise prior energy, see below.

2.4 Prior Energy

The prior energy terms evaluate different aspects of the configuration, concerning the geometry of each single object or the geometric relation of neighbouring objects. In the following, we introduce our configuration model to describe these neighbourhood relations and derive the according terms of the prior energy.

2.4.1 The Configuration Model. We learn typical configurations of objects from training images. Thereby, we characterize a configuration by properties of single objects as width w, height h and location x and y and properties of neighbouring objects as intersection, distance, alignment, and size differences.

Interacting Objects. Usually, the neighbourhood relation $\boldsymbol{x}_i \sim \boldsymbol{x}_j$ is defined in terms of a distance, such that all objects within a ball of given radius are said to be neighboured. In our application, it is not possible to define such a fixed distance, it would differ from image to image and if the number of objects is low, objects even could be on opposite sides of the image. We define the neighbourhood relation in terms of a Voronoi diagram.

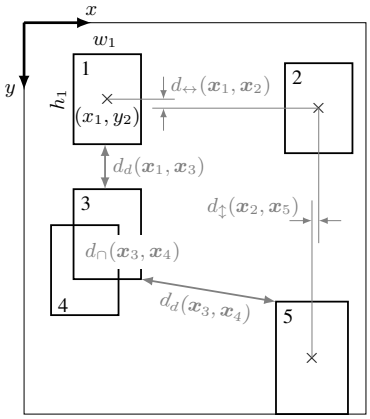

Figure 1: Properties of neighbouring objects.

Therefor, we reduce all rectangles to their center points and evaluate the according adjacency graph. Objects, neighboured within the adjacency graph, are said to be neighbours in image space. The reason, not to use the exoskeleton of the objects themselves is, that at high temperature of simulated annealing the objects might overlap, thus, get merged to one component for processing the exoskeleton.

We define different measures on interacting objects to describe the overall configuration. Not all of them are taken into account for each neighbouring object pair, e.g., we do not measure the vertical alignment for objects neighboured horizontally. In the following, we first describe the interacting measures and then detail, which measures are taken into account under which constraints.

Interacting Distances. Assuming the image scale s [px/m] to be known, we take into account the following properties of interacting objects $\boldsymbol{x}_i \sim \boldsymbol{x}_j$, cf. Fig. 1:

- intersection area
$$d_{\cap}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \frac{\text{area}\,(\boldsymbol{x}_i \cap \boldsymbol{x}_j)}{\min(w_i \cdot h_i, w_j \cdot h_j)} \qquad d_{\cap} \in [0, 1] \quad (4)$$
Objects not intersecting at all, get an intersection area of 0, while objects superimposed get an intersection area of 1.
- minimal distance $d_d(\boldsymbol{x}_i, \boldsymbol{x}_j)$, which is given by the minimal distance between any two points of the objects bounding box, normalized by the image scale s.
- alignment horizontal and vertical, referred to the centre of objects
$$d_{\leftrightarrow}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \frac{|y_i - y_j|}{s} \quad d_{\updownarrow}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \frac{|x_i - x_j|}{s} \quad (5)$$
- size difference in height and width
$$d_{\Delta h}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \frac{|h_i - h_j|}{s}\; d_{\Delta w}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \frac{|w_i - w_j|}{s}\;. \quad (6)$$

In order to shorten the notation, we denote $d_k^{(i,j)} := d_k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ the value of the k-th distance between two objects $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, i.e. $k \in \{\cap, d, \leftrightarrow, \updownarrow, \Delta h, \Delta w\}$.

As already mentioned, we do not take into account all distances for all interacting objects. To denote this special behaviour, we specialize the neighbourhood relation to $\boldsymbol{x}_i \sim^{d_k} \boldsymbol{x}_j$, to declare objects $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ neighbours with respect to distance d_k. Therefor, we define the following neighbourhood relations

$$\begin{aligned} &\boldsymbol{x}_i \sim^{d_{\leftrightarrow}} \boldsymbol{x}_j\;,\; \boldsymbol{x}_i \sim^{d_{\Delta h}} \boldsymbol{x}_j \quad \text{if} \quad d_{\leftrightarrow}^{(i,j)} \leq d_{\updownarrow}^{(i,j)} \\ &\boldsymbol{x}_i \sim^{d_{\updownarrow}} \boldsymbol{x}_j\;,\; \boldsymbol{x}_i \sim^{d_{\Delta w}} \boldsymbol{x}_j \quad \text{if} \quad d_{\leftrightarrow}^{(i,j)} > d_{\updownarrow}^{(i,j)}\;. \end{aligned} \quad (7)$$

This way, we take into account the size difference in height and the horizontal misalignment, only, if the objects are roughly beside each other. Vice versa, we take into account the size difference in width and the vertical misalignment, only, if the objects are roughly on top of each other. Further, the neighbourhood relations $(\sim^{d_{\Delta h}}, \sim^{d_{\Delta w}})$ and $(\sim^{d_{\leftrightarrow}}, \sim^{d_{\updownarrow}})$, respectively, are pairwise exclusive each, i.e. two neighbouring object can interact only in terms of one of them.

2.4.2 Learning the Configuration Statistics. Given annotated training data, we collect all bounding boxes of our target classes, and collect for all annotated objects their location (x, y), width w, and height h and for all interacting objects distances d_d, and $d_{\cap}$. Further, for each possible combination of object classes, we store distances $d_{\leftrightarrow}, d_{\updownarrow}, d_{\Delta h}$ and $d_{\Delta w}$ for all pairs of objects, interacting in terms of the according neighbourhood relation and according class combination. We represent these statistics by their histogram, which we denote $h_k(\cdot \mid \cdot)$, $k \in \{x, y, w, h, \leftrightarrow, \updownarrow, d_{\Delta h}, d_{\Delta w}\}$. We smooth the histograms, using a kernel density estimator, and to enable scoring of these properties, we normalize each histogram, such that $\max(h) = 1$. The minimal distance between neighbouring objects is an exception. We just store the minimal value $t_d\,(c_1, c_2)$, we have seen during training, according to each possible combination of object class c_1 and c_2.

2.4.3 Prior Energy Terms. The prior energy is represented by an unary term $U_{\text{geom}}(\mathcal{X})$ and a pairwise term $U_{\text{conf}}(\mathcal{X})$, which represent the confidence of proposed samples to the learned statistics over size and location of single objects and the local arrangement of neighboured objects. The former is given by the sum of local energies of single objects

$$U_{\text{geom}}(\mathcal{X}) = \sum_i \sum_{k \in \{x,y,w,h\}} \Psi\left(h_k\left(\boldsymbol{x}_i \mid c_i\right)\right)\;, \quad (8)$$

using the penalty function

$$\Psi(h) = -\log\left(2 \cdot h\right)\;, \quad \Psi \in [\infty, -\log(2)]\,. \quad (9)$$

Thus, for each object, we sum up the contributions of each part of the configuration model. Using Eq. (9), objects not fitting the learned function h_k get punished and contribute a positive energy, while objects near the global maximum of h_k, which is 1, even contribute a negative energy.
We denote $\mathcal{E} = \{e_{ij} \mid x_i \sim x_j\}$ the set of all neighbouring objects and define the prior energy regarding interacting object pairs by

$$U_{\text{conf}}(\mathcal{X}) = \textstyle\sum_{(i,j)\in\mathcal{E}} \left(\Psi_{\cap}\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) + \Psi_d\left(\boldsymbol{x}_i, \boldsymbol{x}_j \mid c_i, c_j\right) + \right.$$
$$\left. \tfrac{1}{N(\text{Ne}(\boldsymbol{x}_i))} \cdot \textstyle\sum_{k\in\{\leftrightarrow,\updownarrow,\Delta w,\Delta h\}} \Psi\left(h_k\left(\boldsymbol{x}_i, \boldsymbol{x}_j \mid c_i, c_j\right)\right)\right) \quad (10)$$

where we use the same penalty function $\Psi(\cdot)$ as before for alignment and size differences, while intersection and distance get punished more strongly. We use a hard core penalty function for overlapping rectangles

$$\Psi_{\cap}\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = -a \cdot \log(1 - d_{\cap}^{(i,j)})\;, \quad \Psi_{\cap} \in [\infty, 0]\,, \quad (11)$$

using a constant, e.g., $a = 100$. The function prevents to increase the overall energy by superimposing infinite many objects at medium and low temperature, nevertheless, it allows small overlaps at hight temperature, which is important to explore the state space. If $d_{\cap} = 0$ the objects do not overlap, then there is no penalty.
To penalize neighboured objects with distance below the class specific minimal distance $t_d(c_i, c_j)$, we use a function similar to the hat-function

$$\Psi_d\big(\boldsymbol{x}_i,\boldsymbol{x}_j|c_i,c_j\big)=-b\cdot\log\left(1-\max\left(0,\frac{t_d(c_i,c_j)-d_d^{(i,j)}}{t_d(c_i,c_j)}\right)\right)\ ,\ \Psi_d\in[\infty,0], \tag{12}$$

using a constant $b = 100$. Distances, slightly below $t_d(c_i, c_j)$, get a small positive energy, which strongly increases with decreasing distance. Distances, above $t_d(c_i, c_j)$, do not influence the overall energy.
Again, Eq. (10) sums up for each object the contribution of each interacting distance. To make the configuration energy independent on the number of interacting neighbours $\mathrm{Ne}(\boldsymbol{x}_i)$, we normalize by their number $N\left(\mathrm{Ne}(\boldsymbol{x}_i)\right)$.

Finally, in order to define a prior on the number of objects, we store the minimal and maximal numbers $c_{\min}$ and $c_{\max}$, respectively, of objects per class, we have seen in single images during training. Configurations fitting these numbers should not influence the overall energy, while deviations should get a punishment. Therefrom, we define the prior on the number of objects by

$$U_{\text{num}}(\mathcal{X}) = \sum_{y=1\ldots C} \Psi_n\left(\{\boldsymbol{x}\in\mathcal{X}\mid c=y\}\right)\ , \tag{13}$$

using

$$\Psi_n\left(\mathcal{X}\right) = \begin{cases} 0 & \text{if} \quad N\left(\mathcal{X}\right)\in[c_{\min},c_{\max}] \\ \log\varepsilon & \text{otherwise} \end{cases} . \tag{14}$$

This term is of special interest for low frequent classes, such as entrances. We observed that the regular structure of window grids, in most cases, overvotes the existence of a single entrance, which we avoid by using this prior.

2.5 Optimization

Our task is to find the configuration $\mathcal{X}$, which maximizes the unnormalized point process density or minimizes the energy $U(\mathcal{X})$, respectively. It is a complex function with rough landscape. Even its dimensionality is unknown due to the unknown number of objects. We optimize with rjMCMC coupled with simulated annealing to find the global optimum. Introducing the temperature parameter T, the optimizer is given by

$$\widehat{\mathcal{X}} = \underset{\mathcal{X}}{\operatorname{argmax}}\ f(\mathcal{X})^{\frac{1}{T_t}} = \underset{\mathcal{X}}{\operatorname{argmin}}\ \frac{U\left(\mathcal{X}\right)}{T_t},\ \lim_{t\to\infty} T_t = 0\ . \tag{15}$$

Geyer and Møller (1994) propose the so called Birth an Death algorithm to sample point processes, which turns out to be a special type of Green's rjMCMC sampler (Green, 1995). Given a point process $\underline{\mathcal{X}}$ with points $\boldsymbol{x}_i$ in $\mathcal{S}$, distribution $F_{\mathcal{X}}(\cdot)$, density $f(\cdot)$, and intensity $\mu(\cdot)$ of the reference Poisson process, Geyer and Møller (1994) built a Markov Chain that, provided $f(\cdot)$ fulfils the stability condition, given by what is called the Papangelou conditional intensity, was proven to build a Markov Chain that is $F_{\mathcal{X}}(\cdot)$ invariant, thus, simulates the point process $\underline{\mathcal{X}}$ (Geyer and Møller, 1994; Ortner et al., 2003). Furthermore, it was proven that it simulates a $F_{\mathcal{X}}(\cdot)$ irreducible Markov chain that is recurrent and ergodic, which guarantees its convergence to its target distribution $F_{\mathcal{X}}(\cdot)$.
To improve the mixing properties of the Markov chain, Ortner et al. (2003) extende this algorithm and introduce additional proposition kernels for the moves within the Markov chain. Beside the basic birth and death, they proposed to use birth and death in a neighbourhood and non jumping transformations, such as translation, rotation or dilation of existing objects of the configuration. In our work, we use two types of moves: (1) dimensional jumping transformation: birth and death, (2) non jumping transformations: translation, dilation, switching. The latter randomly selects an object from the current configuration and randomly perturbs its marks. The generic point process sampler algorithm is given in Alg. 1, which is equivalent to Hasting's and Green's Algorithm. Having M different moves, we take the proposition kernel of the Markov chain as a mixture of $m = 1\ldots M$ proposition distributions, each having a probability $j_m(\mathcal{X})$ to choose move type m being at $\mathcal{X}$. The kernel $\mathcal{Q}_m$ can be interpreted as instruction, how to throw a new sample $\boldsymbol{x}$ being at $\mathcal{X}$, using move type m, which we define in more detail, in the following.
We set $\mathcal{S} = \{[1, I_C] \times [1, I_R]\}$, using $\mathcal{M} = \{[W_{\min}, W_{\max}] \times [H_{\min}, H_{\max}] \times [1, C]\}$, using ranges $[W_{\min}, W_{\max}]$ and $[H_{\min}, H_{\max}]$ for width and height, respectively, taken from the training sample's sizes. In order to clarify the notion, we denote the move types not by numbers. We use $m \in \{\mathrm{b}, \mathrm{d}, \mathrm{nj}\}$ to denote birth, death, and non jumping moves, whereby the latter comprises different move types leading to the same Green ratio.

Birth and Death. In case of birth, we create a new object $\boldsymbol{x} \in \mathcal{S} \times \mathcal{M}$ and propose $\mathcal{Y} = \mathcal{X} \cup \boldsymbol{x}$. Noting $f\left(\mathcal{Y}\right)/f\left(\mathcal{X}\right) = \mathrm{e}^{-U(\mathcal{Y})}/\mathrm{e}^{-U(\mathcal{X})} = \mathrm{e}^{U(\mathcal{X})-U(\mathcal{Y})}$, and introducing the temperature parameter T_t, of simulated annealing, we obtain Green's ratio for birth and death kernels

$$R_{\mathrm{b}}(\mathcal{X},\mathcal{Y}) = \frac{\mu(\mathcal{S})}{N(\mathcal{Y})}\,\mathrm{e}^{\left(\frac{U(\mathcal{X})-U(\mathcal{Y})}{T_t}\right)} \qquad R_{\mathrm{d}}(\mathcal{X},\mathcal{Y}) = \frac{N\left(\mathcal{X}\right)}{\mu(\mathcal{S})}\,\mathrm{e}^{\left(\frac{U(\mathcal{X})-U(\mathcal{Y})}{T_t}\right)}\ . \tag{16}$$

Non Jumping Transformations. With this type of moves, we randomly perturb the marks of an existing object. Therfor, we uniformly select an object $\boldsymbol{x} \in \mathcal{X}$ and throw random numbers $u \sim Z_{(\mathcal{X},u)}$ according a suitable distribution, usually uniform in a certain range of values. Given a function $g : \boldsymbol{y} = g(\boldsymbol{x}, u)$, which transforms the object we propose $\mathcal{Y} = (\mathcal{X} \setminus \boldsymbol{x}) \cup \boldsymbol{y}$.

- **Translation:** manipulates the centre of an object $\boldsymbol{x}$. We throw $\boldsymbol{u} \in \mathbb{R}^2,\ \boldsymbol{u} \sim U\left([-\Delta x, \Delta x] \times [-\Delta y, \Delta y]\right)$. $g : \boldsymbol{y} = g(\boldsymbol{x}, \boldsymbol{u}) = [x + u_x,\ y + u_y,\ w,\ h,\ c]^{\mathsf{T}}$.
- **Dilation:** manipulates width and height of an object $\boldsymbol{x}$. We throw $\boldsymbol{u} \in \mathbb{R}^2,\ \boldsymbol{u} \sim U\left([-\Delta w, \Delta w] \times [-\Delta h, \Delta h]\right)$. $g : \boldsymbol{y} = g(\boldsymbol{x}, \boldsymbol{u}) = [x,\ y,\ w + u_w,\ h + u_h,\ c]^{\mathsf{T}}$.
- **Switching:** changes the label an object $\boldsymbol{x}$. We throw $u \in \mathbb{N},\ u \sim U\left(\{1, \ldots, C\} \setminus c\right)$, thus, we throw a random number out of C classes, which is not the current one c. $g : \boldsymbol{y} = g(\boldsymbol{x}, u) = [x,\ y,\ w,\ h,\ u]^{\mathsf{T}}$.

We fix control parameters Δx, Δy, Δw, and Δh to $1/8$ of the image's height and width, respectively. In each case, the inverse transformation $\boldsymbol{x} = g^{-1}(\boldsymbol{y}, \boldsymbol{u})$ exist and is as possible as g is, which ensures reversibility. Thus, if the translation and its inverse are equally probable and the according perturbation variables where thrown from the same distribution, Green's ratio simplifies to

$$R_{\mathrm{nj}}\left(\mathcal{X},\mathcal{Y}\right) = \mathrm{e}^{\left(\frac{U(\mathcal{X})-U(\mathcal{Y})}{T_t}\right)}\ , \tag{17}$$

Algorithm 1: Generic point process sampler

Input: state $X^{(t)} = \mathcal{X}$ of the Markov chain at time t, probabilities j_m for choosing move type m
Output: X_{t+1}

with probability $j_m\left(\mathcal{X}\right)$ choose proposition kernel $\mathcal{Q}_m$
sample $\boldsymbol{x} \sim \mathcal{Q}_m\left(\cdot \mid \mathcal{X}\right)$
propose $\mathcal{Y} = \mathcal{X} \cup \boldsymbol{x}$
compute Green's ratio $R_m\left(\mathcal{X},\mathcal{Y}\right)$ c.f. (16), (17)
sample $\alpha \sim \mathcal{U}(0, 1)$
if $\alpha < min(1, R_m)$ **then**
    $X^{(t+1)} = \mathcal{Y}$ `// accept the move`
else
    $X^{(t+1)} = X^{(t)}$ `// reject`
end

Figure 2: Sample images from used datasets. From top to bottom: Basel old town, Basel row houses, City houses with balconies.

which was shown by Ortner et al. (2003) and completes Alg. 1.

Exceptions. Impossible proposals for moves, i.e. $\boldsymbol{y} \notin \mathcal{S} \times \mathcal{M}$, are not used. In that case, we reject the according proposition, without any impact on the invariant distribution as pointed out by Ortner et al. (2003).

3. EXPERIMENTS

This section shows result for the proposed MPP for facade interpretation. We learn the model on three different datasets, showing different characteristics. The goal is to show that we are able to learn a model from few images that we can use to evaluate new images with similar characteristics. To prove the learned prior model, we simulate configurations, ignoring bottom up evidence from image data. We visualize successful interpretation results and provide a qualitative pixelwise evaluation.

3.1 Datasets

From eTrims image database (Korc and Förstner, 2009), we assembled different image collections, which are characterized by similar facade structure and as similar as possible appearance of addressed facade objects. Fig. 2 shows samples of datasets we use for this work. The first dataset consist of six images from Basel old town, characterized by a medium number of highly structured facade objects, which are regularly arranged. There are just windows and entrances. Due to different decorations, the appearance of facade objects differ between the images. The second dataset consist of 11 images showing Basel row houses, characterized by a sparse configuration of few facade objects, which are windows and entrances. The configuration of facade objects is homogeneous within the dataset, while the appearance differ due to occlusions and decorations. The last dataset consist of 8 images showing apartment houses, characterized by a large number of facade objects, which are 135 windows and 52 balconies in total. Although the windows are regularly arranged, which seems to be easy to evaluate, the configuration of balconies differ between the images of the collection, which results in less distinctive configuration statistics.

3.2 Simulation

To asses our concept of learning the configuration statistics and the modelling of the Gibbs energy, we initially simulate configurations based on individually learned configuration statistics, ignoring the dataterm in Eq. (1). We fix $U_{\text{data}} = -1$ for all proposals and simulate the Markov chain, using a fast geometric temperature schedule with $T_{t+1} = T_0 \cdot \alpha^t$ with fixed parameters $T_0 = 2$ and $\alpha = 0.9999$ for all runs. Fig. 3 shows samples based on the configurations statistics learned from each given image collection. Actually, we simulate configurations similar to those of the given images, which reflect the characteristics of underlying training images. Samples of the balcony dataset are more variable due to the diversity of configurations of given training images. Nevertheless, homogeneity of window lattices is realized.

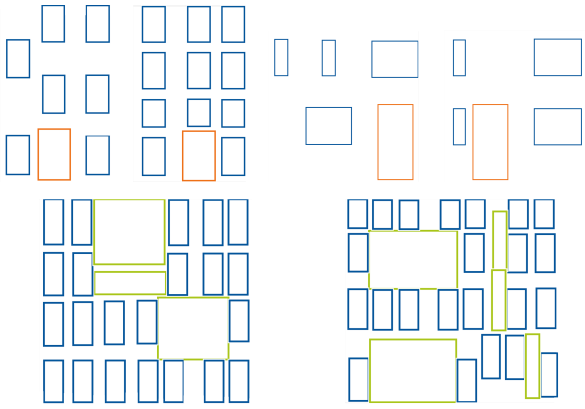

Figure 3: Simulations without using bottom up evidence by an image, just based on learned configuration statistics from datasets: **Top:** two images each: Basel old town, Basel row houses. **Bottom** city houses with balconies. **Colours:** Blue: windows. Orange: entrances. Green: balcony.

3.3 Evaluation

For evaluation on real images, we perform object detection in a leave-one-out cross-validation setting, i.e. we use one image of a dataset for testing and all other for learning the classifier and the configuration statistics. For all experiments, we use a slow geometric temperature schedule, using $\alpha = 0.999999$ and $T_0 = 2$. The latter was determined empirically, such that the average acceptance rate at beginning was around 70%. The weights λ_i, c.f. Eq. (1), were set empirically, too. Up to know, we did not learn them from data, which is planned using cross validation. The weight λ_3, for the prior on the number of objects, is set to 1 for all experiments, while the weights λ_1 and λ_2 were individually set for each dataset. The range of marks, thus, range of width and hight for sampling new or manipulating existing objects, is given by the training data. The optimization needed on average 7 million iterations.

Fig. 4 visualize exemplary results. For shown samples of datasets Basel and Basel row houses, the detection of facade objects is complete, thus, we do not miss any object, except few small windows. This is stressed by their confusion matrices, cf. Tab. 1, which provide a qualitative evaluation in terms of pixelwise comparison to ground truth for each pixel and all images of the according dataset. Shown numbers proof high detection rates up to an accuracy of 94%. Anyway, deviations in the objects outline, especially for class windows, result in a pixelwise accuracy, which is lower.

The dataset city houses with balconies is more challenging. We miss some windows, which, due to protrusion of balconies and perspective distortion, overlap with bounding boxes of neighbouring balconies. As our energy is designed to prevent overlapping objects, the process at equilibrium does not accept these object proposals. Further, we miss some windows or do not get their exact outline. Both might be explained by the loose structure of neighbouring objects. For example, the height of windows

Table 1: Confusion matrices for for pixelwise evaluation. Numbers are given as percentage of pixels that belong to the according class. **Left:** Basel old town. **Middle:** Basel row houses. **Right:** City houses with balconies.

		prediction		
		bg	win	entr
truth	bg	**90.29**	8.68	1.03
	win	16.39	**83.61**	0.00
	entr	10.75	0.00	**89.25**

		prediction		
		bg	win	entr
truth	bg	**94.12**	4.45	1.43
	win	13.05	**86.95**	0.00
	entr	5.10	0.00	**94.90**

		prediction		
		bg	win	balc
truth	bg	**91.55**	5.02	3.43
	win	16.61	**83.13**	0.26
	balc	10.94	3.43	**85.63**

Figure 4: Sample results. **1st row:** Basel old town dataset, using weights $\lambda_1 = \lambda_2 = 0.3$. **2nd row:** Basel row houses dataset, using weights $\lambda_1 = \lambda_2 = 0.4$. **3nd and 4th row:** city houses with balconies, using weights $\lambda_1 = \lambda_2 = 1/5$. Colours as given in Fig. 3.

is wrongly detected if they do not have horizontally neighboured windows, whose hight supports them. The same holds for few missing windows. Their appearance differs from common windows of this dataset, thus, their data energy is weak, and their existence is not supported by vertically neighboured windows.

These effects show that the energy should take into account global neighbourhood relations. The pairwise energy term we use, takes into account objects, which are directly adjacent. In this dataset we may argue that the regular window grid is disturbed by balconies in between. Thus, we should model the global layout of each individual class, which is not the case in our model.

In order to demonstrate the sensitivity of the approach w.r.t. weights λ_i, we show variations in Fig. 5. We observe that increasing weights of the prior energy lead to hallucinations of missing objects, e.g. additional windows, or even overvotes the prior on the number of objects per class, which leads to an additional window instead of an entrance.

4. SUMMARY AND CONCLUSION

In this paper, we proposed a novel method for facade image interpretation based on a marked point process, combining bottom up evidence, given by an object classifier, and prior knowledge, about typical configurations of facade objects, from top-down. We represent facade objects by a simplified rectangular object model and present an energy model, which evaluates the agreement of a proposed configuration with the given image and the learned statistics about typical configurations. Due to the learned prior energies, our model is almost free of tunable parameters, in contrast to other approaches, dealing with marked point processes. We show promising results on three datasets and provide a qualitative evaluation, which demonstrates the capability of complete and accurate detection of facade objects.

However, we are not competitive to state of the art results, e.g. Teboul et al. (2010) in terms of complexity. In contrast to them we deal with few classes per image. We are aware on the weakness of our evaluation: We proved our approach on few datasets with small sample size. Their appearance as well as structure of facade objects are homogeneous within each dataset. But, as soon as more training data are available and the procedures are parallelised for GPU processing, we may evaluate more data and may learn the remaining weights from data, e.g., by cross validation. Nevertheless, compared to grammar based approaches, we are more flexible in terms of underlying structure of facade objects and even able to deal with very sparse structure. We believe that

Figure 5: Results for varying parameters. **From left to right:** groundtruth | results, using the datasets setting $\lambda_1 = \lambda_2 = 0.3$ | $\lambda_1 = \lambda_2 = 1$ | $\lambda_1 = 0.5, \lambda_2 = 0.1$.

we prospectively may overcome the limitations of grammar based approaches, which have to be designed individually for each type of facade and are dedicated for large, regularly structured types of facades, such as Hausmanian, cf. Teboul et al. (2010).
In future work we will try to enhance the energy model to express the structure of more complex facades, including more classes, especially balconies. Nonetheless, we will implement methods to learn the remaining weights from data.

References

Alegre, F. and Dellaert, F., 2004. A probabilistic approach to the semantic interpretation of building facades. In: Int. Workshop on Vision Techniques Applied to the Rehabilitation of City Centres, pp. 1–12.

Ali, H., Seifert, C., Jindal, N., Paletta, L. and Paar, G., 2007. Window Detection in Facades. In: ICIAP, pp. 837–842.

Baddeley, A. J., 2007. Lecture Notes in Mathematics: Stochastic Geometry. Vol. 1892, Springer Verlag , Berlin Heidelberg, chapter Spatial Point Processes and their Applications, pp. 1–75.

Baddeley, A. J. and Lieshout, M. N. M. V., 1993. Stochastic geometry models in high-level vision. J. of Appl. Stat. 20(5-6), pp. 231–256.

Börcs, A. and Benedek, C., 2012. A Marked Point Process Model for Vehicle Detection in Aerial Lidar Point Clouds. In: Int. Ann. Photogramm. Remote Sens. (ISPR'12), Vol. I-3, pp. 93–98.

Bredif, M., Tournaire, O., Vallet, B. and Champion, N., 2013. Extracting polygonal building footprints from digital surface models: A fully-automatic global optimization framework . P&RS 77, pp. 57–65.

Burochin, J.-P., Vallet, B., Bredif, M., Mallet, C., Brosset, T. and Paparoditis, N., 2014. Detecting blind building facades from highly overlapping wide angle aerial imagery . P&RS 96, pp. 193–209.

Čech, J. and Šára, R., 2008. Windowpane detection based on maximum aposteriori labeling. In: IWCIA.

Chai, D., Förstner, W. and Ying Yang, M., 2012. Combine Markov Random Fields and Marked Point Processes to Extract Building from Remotely Sensed Images. In: Int. Ann. Photogramm. Remote Sens. (ISPR'12), Vol. I-3, pp. 365–370.

Cohen, A., Schwing, A. and Pollefeys, M., 2014. Efficient Structured Parsing of Facades Using Dynamic Programming. In: CVPR, pp. 3206–3213.

Descombes, X., Stoica, R., Garcin, L. and Zerubia, J., 2001. A RJMCMC Algorithm for Object Processes in Image Processing. Monte Carlo Methods and Applications 7(1-2), pp. 149–156.

Dick, A. R., Torr, P. H. S. and Cipolla, R., 2004. Modelling and Interpretation of Architecture from Several Images. IJCV 60, pp. 111–134.

Ferrari, V., Fevrier, L., Jurie, F. and Schmid, C., 2008. Groups of adjacent contour segments for object detection. TPAMI 30(1), pp. 36–51.

Fröhlich, B., Rodner, E. and Denzler, J., 2010. A Fast Approach for Pixelwise Labeling of Facade Images. In: ICPR, pp. 3029–3032.

Garcin, L., Descombes, X., Le Men, H. and Zerubia, J., 2001. Building detection by Markov object processes. In: ICIP, Vol. 2, pp. 565–568.

Geyer, C. J. and Møller, J., 1994. Simulation Procedures and Likelihood Inference for Spatial Point Processes. Scandinavian Journal of Statistics 21(4), pp. 359–373.

Green, P. J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82(4), pp. 711–732.

Korc, F. and Förstner, W., 2009. eTRIMS Image Database for Interpreting Images of Man-Made Scenes. Technical Report TR-IGG-P-2009-01, University of Bonn, Dept. of Photogrammetry.

Koziński, M., Gadde, R., Zagoruyko, S., Obozinski, G. and Marlet, R., 2015. A MRF shape prior for facade parsing with occlusions. In: CVPR, pp. 2820–2828.

Lafarge, F. and Gimel'farb, G. L., 2008. Texture Representation by Geometric Objects using a Jump-Diffusion Process. In: BMVC.

Lafarge, F., Descombes, X., Zerubia, J. and Pierrot-Deseilligny, M., 2010a. Structural Approach for Building Reconstruction from a Single DSM. TPAMI 32(1), pp. 135–147.

Lafarge, F., Gimel'farb, G. and Descombes, X., 2010b. Geometric Feature Extraction by a Multimarked Point Process. TPAMI 32(9), pp. 1597–1609.

Lee, S. C. and Nevatia, R., 2004. Extraction and integration of window in a 3D building model from ground view images. In: CVPR, Vol. II, pp. 113–120.

Martinović, A., Mathias, M., Weissenberg, J. and Gool, L. V., 2012. A Three-Layered Approach to Facade Parsing. In: ECCV, pp. 416–429.

Müller, P., Zeng, G., Wonka, P. and Van Gool, L., 2007. Image-based Procedural Modeling of Facades. ACM Trans. Graph.

Ortner, M., Descombes, X. and Josiane Zerubia, J., 2008. A Marked Point Process of Rectangles and Segments for Automatic Analysis of Digital Elevation Models. TPAMI 30(1), pp. 105–119.

Ortner, M., Descombes, X. and Zerubia, J., 2003. Improved RJMCMC point process sampler for object detection by simulated annealing. Technical Report 4900, INRIA.

Ortner, M., Descombes, X. and Zerubia, J., 2007. Building Outline Extraction from Digital Elevation Models Using Marked Point Processes. IJCV 72(2), pp. 107–132.

Park, M., Brocklehurst, K., Collins, R. and Liu, Y., 2010. Translation-Symmetry-based Perceptual Grouping with Applications to Urban Scenes. In: ACCV, Vol. 3, pp. 1631–1645.

Recky, M. and Leberl, F., 2010. Windows Detection Using K-means in CIE-Lab Color Space. In: ICPR, pp. 356–359.

Ripperda, N., 2008. Determination of Facade Attributes for Facade Reconstruction. In: In Int. Ann. Photogramm. Remote Sens. (ISPRS'08), Vol. B3a, pp. 285–290.

Roscher, R., Förstner, W. and Waske, B., 2012a. I2VM: Incremental Import Vector Machines. Image and Vision Computing 30, pp. 263–278.

Roscher, R., Waske, B. and Förstner, W., 2012b. Incremental Import Vector Machines for Classifying Hyperspectral Data. TGARS 50(9), pp. 3463–3473.

Teboul, O., Simon, L., Koutsourakis, P. and Paragios, N., 2010. Segmentation of building facades using procedural shape priors. In: CVPR, pp. 3105–3112.

Tournaire, O., Bredif, M., Boldo, D. and Durupt, M., 2010. An efficient stochastic approach for building footprint extraction from digital elevation models. P&RS 65, pp. 317–327.

Tournaire, O., Paparoditis, N. and Lafarge, F., 2007. Rectangular road marking detection with marked point processes. In: PIA, pp. 149–154.

Tylecek, R. and Sara, R., 2010. A Weak Structure Model for Regular Pattern Recognition Applied to Facade Images. In: ACCV, pp. 445–458.

Verdie, Y. and Lafarge, F., 2013. Detecting parametric objects in large scenes by Monte Carlo sampling. IJCV 106(1), pp. 57–75.

Wang, J., Fang, T., Su, Q., Zhu, S., Liu, J., Cai, S., Tai, C. and Quan, L., 2015. Structure-driven Facade Parsing With Irregular Patterns. In: ACPR.

Wendel, A., Donoser, M. and Bischof, H., 2010. Unsupervised Facade Segmentation Using Repetitive Patterns. In: DAGM, pp. 51–60.

Wenzel, S. and Förstner, W., 2012. Learning a Compositional Representation for Facade Objects. In: Int. Ann. Photogramm. Remote Sens. (ISPR'12), pp. 197–202.

Wenzel, S., Drauschke, M. and Förstner, W., 2007. Detection and Description of Repeated Structures in Rectified Facade Images. PFG 7, pp. 481–490.

Werner, T. and Zisserman, A., 2002. New Techniques for Automated Architecture Reconstruction from Photographs. In: ECCV, pp. 541–555.

Zhu, J. and Hastie, T., 2005. Kernel Logistic Regression and the Import Vector Machine. JCGS 14(1), pp. 185–205.