Windowpane Detection based on Maximum Aposteriori Probability Labeling

Jan Čech and Radim Šára

Center for Machine Perception
Czech Technical University
Prague, Czech Republic
{cechj,sara}@cmp.felk.cvut.cz

Abstract. Segmentation of windowpanes in images of building façades is formulated as a task of maximum aposteriori probability labeling. Assuming orthographic rectification of the image, the windowpanes are always axis-parallel rectangles of relatively low variability in appearance. Every image pixel has one of 10 possible labels, and the labels in adjacent pixels are constrained by allowed label configuration, such that the image labels represent a set of non-overlapping rectangles. The task of finding the most probable labeling of a given image leads to NP-hard discrete optimization problem. However, we find an approximate solution using a general solver suitable for such problems and we obtain promising results which we demonstrate on several experiments.
Substantial difference between the presented paper and the state-of-the-art papers on segmentation based on Markov Random Fields is that we have a strong structure model, forcing the labels to form rectangles, while other methods do not model the structure at all, they typically only have a penalty when adjacent labels are different, in order to make resulting patches more continuous to reduce influence of noise and prevent over-segmentation. The difference is assessed experimentally.

1 Introduction

Markov Random Fields (MRFs) have been used in image analysis for a long time [18, 7]. There are many papers on image segmentation using MRFs, e.g. [17, 8]. The spatial relationships of pixels in the image domain is often modeled by the *Potts model* [2]. It means there is a zero penalty for adjacent pixel having the same label and a constant penalty if the adjacent pixels have different labels. This prior model reflects a natural assumption on the segmentation to be locally homogeneous. The homogeneous patches have higher probability to become a part of the solution.

On the other hand, the Potts model cannot incorporate any stronger assumption on the *shape* of the patches, i.e. on the structure of the segmentation. In this paper, we make an explicit requirement for the shape of the patches to be segmented. This is done by introducing a prior model of more complicated structure, where pairwise transition probabilities between labels are asymmetric.

The resulting MRF is non-Gibbsian. With 10 labels and such structure prior, we force the solution to be a set of axis-parallel non-overlapping rectangles, which represent windowpanes. Similar structure prior models appeared in [15, 16] to demonstrate the functionality of their labeling solver.

The obvious drawback of the proposed approach compared to similar segmentation with Potts model is that it leads to NP-complete problem. However, we will show that an approximate algorithm will give an acceptable results and the segmentation quality of the proposed method is superior to the method using the Potts model.

Of course, there are several different methods to detect windows in the façade images. For instance, in [12, 3] they have a *parametric* model of windows. Changing the parameters (as width, aspect ratio, brightness, etc.) using Markov Chain Monte Carlo sampling they try to generate the image which is the most similar to the given image. In [1], they use stochastic context-free grammars to represent a hierarchical regular structure of a façade. These approaches are very different from our simple formulation based on segmentation.

The paper is structured as follows: The problem is formulated in Sec. 2. Experimental validation on both synthetic data and images of real façades is given in Sec. 3. The Sec. 4 concludes the paper.

2 Problem formulation

The image lattice is a finite set of pixels T, where we denote pixel $t' \in T$ an immediate 4-neighbour of pixel $t \in T$. Every pixel $t \in T$ is in one of states defined by the finite set of labels X. In our problem, $X = \{\text{E,I,L,R,T,B,TL,TR,BL,BR}\}$, see Fig. 1.

E	E	E	E	E	E	E	E	E
E	E	TL	T	T	T	TR	E	E
E	E	L	I	I	I	R	E	E
E	E	L	I	I	I	R	E	E
E	E	L	I	I	I	R	E	E
E	E	L	I	I	I	R	E	E
E	E	BL	B	B	B	BR	E	E
E	E	E	E	E	E	E	E	E

Fig. 1. Labeled image. Image and the labeled image encoded in color (left) and allowed configuration of labels (right), E - external label (represents a façade wall), I - internal windowpane label, L,R,T,B are left, right, top, bottom edge respectively and TL, TR, BL, BR are corners.

There are rules defining the configuration of adjacent labels, which are allowed, such that the labels always form a set of non-overlapping rectangles. For instance, right to T only another T or TR is allowed, down to B only E is allowed, etc. The full list is given in Sec. 2.1.

Let us have the labeling of the image $\mathbf{x} = (x_1, x_2, \ldots, x_{|T|}) \in X^{|T|}$, where x_t is a label $x \in X$ at pixel $t \in T$, and image data (observation) $\mathbf{d} = (d_1, d_2, \ldots, d_{|T|}) \in D^{|T|}$, where d_t is a data (feature) vector $d \in D$ at pixel $t \in T$. In our case, the space D is the RGB-color space and each d_t is a 3-vector of red, green and blue intensities in pixel t.

The task is to find the most probable labeling given the image data, i.e. to find the labeling maximizing the posterior probability

$$\mathbf{x}^* = \arg \max_{\mathbf{x} \in X^{|T|}} p(\mathbf{x}|\mathbf{d}), \tag{1}$$

where $\mathbf{x}$ is the image labeling and $\mathbf{d}$ is the image data. Using the Bayes law, we get

$$\mathbf{x}^* = \arg \max_{\mathbf{x} \in X^{|T|}} \frac{p(\mathbf{d}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{d})}, \tag{2}$$

where $p(\mathbf{d}|\mathbf{x})$ is a probability of the observation given the labeling. We assume independence, so the probability distribution can be written as

$$p(\mathbf{d}|\mathbf{x}) = \prod_t p(d_t|x_t). \tag{3}$$

We refer to $p(d|x)$ as the *image model*. It is a probability distribution function of appearance for each individual label $x \in X$, e.g. the windowpane label (I-label) is typically of dark or sky color. This is learnt from examples.

The term $p(\mathbf{x})$ in (2) is a prior probability of the labeling. We model it as

$$p(\mathbf{x}) = \prod_{t,t'} p(x_t, x_{t'}), \tag{4}$$

where $t' \in T$ is an immediate 4-neighbour of pixel $t \in T$ and the product is over all pairs of image neighbours. We refer to $p(x, x')$ as the *structure model*. It is a probability distribution of co-occurrence of neighbouring labels. It reflects compatibility of adjacent labels, i.e. the rules that generate a 2D language of axis-parallel non-overlapping rectangles. Besides allowing certain label configuration, the non-zero probabilities represent the shape of rectangles (e.g. their size, aspect ratio). This probability distribution is learnt from labeled examples.

The last term $p(\mathbf{d})$ in (2) is a prior probability of the observed image, which does not influence the location of the optimum $\mathbf{x}^*$ and can be omitted, so

$$\mathbf{x}^* = \arg \max_{\mathbf{x} \in X^{|T|}} \prod_t p(d_t|x_t) \prod_{t,t'} p(x_t, x_{t'}). \tag{5}$$

Applying logarithm to (5), we obtain the max-sum labeling problem

$$\mathbf{x}^* = \arg \max_{\mathbf{x} \in X^{|T|}} \sum_t g_t(x_t) + \sum_{t,t'} g_{tt'}(x_t, x_{t'}), \tag{6}$$

where

$$\begin{aligned} g_t(x_t) &= \log p(d_t|x_t), \\ g_{tt'}(x_t, x_{t'}) &= \log p(x_t, x_{t'}). \end{aligned} \tag{7}$$

The illustration of the labeling problem is in Fig. 2. There is an image. Pixels $t \in T$ are sketched as squares, each containing a finite set of labels $x \in X$ creating nodes of the underlying graph defining the problem topology. The labels between adjacent pixels are interconnected via edges. Each node has assigned a node quality $g_t(x_t)$, which reflects an agreement of the data with the label. Each edge has assigned an edge quality $g_{tt'}(x_t, x_{t'})$, which reflects label compatibilities, see (7). The task of the max-sum labeling (6) is to select one of labels at each pixel that maximize the sum of node qualities and corresponding edge qualities for the entire image. A typical maximizer nodes and corresponding edges are marked by color. Note, that the graph of our problem is much larger than in Fig. 2, in the sense of number of pixels and number of labels (we have 10 labels).

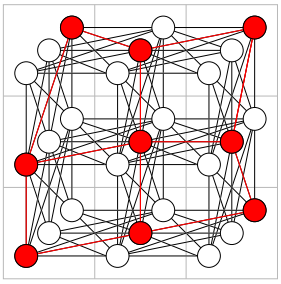

Fig. 2. Illustration of the labeling problem.

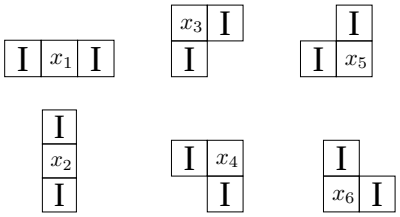

Fig. 3. Configurations of the I-label. Relates to a proof the labels form a set of rectangles.

The problem (6) is NP-complete in general. There are solvable sub-classes [6], e.g. when the number of labels is 2, or when the underlying graph does not contain cycles and others. But, our task does not belong to any of the known solvable sub-classes and probably remains NP-complete. The brute force approach to the problem is of complexity $\mathcal{O}(|X|^{|T|})$. However, there exist approximate algorithms which finds a sub-optimal solution, e.g. (loopy) belief propagation [13, 4], Kolmogorov's TRW-S algorithm [9], max-sum diffusion by Kovalevsky and Flach [11, 5], Schlesinger's linear programming relaxation [14–16] or recent method via dual decomposition by Komodakis et al. [10].

2.1 Implementation

We use a very simple image model $p(d|x)$ in order to be easily learnable from exemplar images. For each label separately, we learn a probability distribution of the pixel color $d = (r, g, b)$ as red, green, blue intensity channels

$$p(d|x) = p(r, g, b|x), \tag{8}$$

where the $p(r, g, b|x)$ is assumed to be of Gaussian distribution. The mean value vector and covariance matrix are estimated from annotated training images.

The structure model $p(x, x')$ was set in order to create the language of axis-parallel non-overlapping rectangles. There are two types of transition probabilities: horizontal $p_h(x, x')$ where x' is the neighbour of x to the right, and vertical $p_v(x, x')$ where x' is the neighbour of x to down. Note that $p(x, x') \neq p(x', x)$ and $p_h \neq p_v$. The asymmetry of the label co-occurrence is a necessary prerequisite for complex structure modeling.

The allowed horizontal left to right transitions are: (E,E), (E,TL), (E,L), (E,BL), (I,I), (I,R), (L,I), (R,E), (T,T), (T,TR), (B,B), (B,BR), (TL,T), (TR,E), (BR,E), (BL,B). The allowed vertical up to down transitions are: (E,E), (E,TL), (E,T), (E,TR), (I,I), (I,B), (L,L), (L,BL), (R,R), (R,BR), (T,I), (B,E), (TL,L), (TR,R), (BL,E), (BR,E), see Fig. 1 - right. The probabilities $p(x, x')$ of listed allowed transitions are non-zero, the forbidden transitions has zero probability.

We can easily prove, that the only labeling with non-zero probability is a set of axis-parallel non-overlapping rectangles. Assume the set is non-empty, i.e. there is at least one I-label, see Fig. 1. If the I-labels do not form an axis-parallel rectangle, there must be at least one of the configuration of adjacent labels in Fig. 3 where $x_i \neq \mathrm{I}$ is allowed. By the above rules, the only possibility is $x_i = \mathrm{I}$, which is a contradiction. Similarly, we can show, the set of rectangles can be empty or has one or more elements.

Probabilities $p(x, x')$ are estimated from labeled examples, as a relative frequency of individual transitions $(x, x') \in X \times X$ in all transitions appearing in the labeled examples.

We selected Werner's implementation [15, 16] of linear programming relaxation based max-sum solver by Schlesinger [14], since it seems to give good results for our problem, unlike the belief propagation [13, 4] which often oscillates and max-sum diffusion [5, 11] which is very slow.

3 Experiments

We performed experiments on both simulated data and on images of real façade.

3.1 Synthetic experiment

We use a synthetic gray-scale test image simulating a façade, see Fig. 5. It is of 100×100 pixels containing 25 dark rectangles of intensity $\mu_0 = 0$, in the light background of intensity $\mu_1 = 1$. We added independent Gaussian noise with increasing standard deviation σ to both segments (rectangles and background). So, the statistical image model of the rectangles and background segments is $N(\mu_0, \sigma^2)$, $N(\mu_1, \sigma^2)$ respectively.

In a repeated experiment, we measured an error rate, i.e. percentage of pixels which were labeled incorrectly, as a function of noise level σ. Beside the proposed

method (full model, 10 labels), we measure the performance of MRF segmentation with 2 labels and Potts model (simple model), and the local Bayesian classifier

$$\forall t \in T,\ x_t^* = \arg\max_{\mathbf{x}_t \in X} p(d_t|x_t)p(x_t), \tag{9}$$

were $p(x)$ is a prior probability of the label. So the Bayesian decision is performed in each pixel independently (we call the method the independent Bayes).

The prior probabilities of the Bayesian classifier and the transition probabilities of both simple and full prior model were estimated from ground-truth labeling as relative frequencies of transitions. The parameters of the image models were set $\sigma = 0.2$ and kept fixed throughout the experiment.

The results shown in Fig. 4 are averaged from 10 random trials and the plots have error-bars. The local decision (independent Bayes) has worse performance than methods modeling pixel neighbourhood relations (simple and full model). The simple 2-label Potts model MRF segmentation outperforms the local method, which is a well known fact in segmentation literature. The full 10-label model (the proposed method) is the best here. The reason is that it precisely models the structure of the segment. The force of the full model allowing only axis-parallel non-overlapping rectangles has a large impact here.

The segmentation results for $\sigma = 0.3$ and $\sigma = 0.8$ are in Fig. 5. The full model is free of errors for $\sigma = 0.3$, unlike other methods. All methods except for the proposed full model fail for noise level $\sigma = 0.8$. Notice the difference between the simple and full model. Although the proposed method makes a few errors for such a noise level, it performs much better than the simple model.

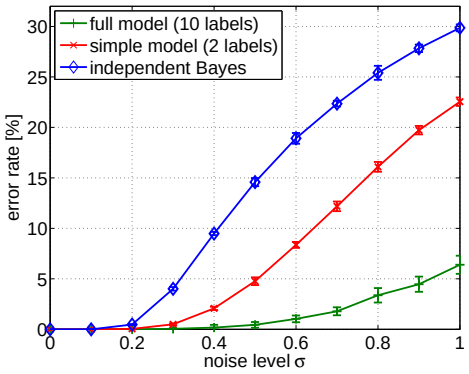

Fig. 4. Labeling error rate as a function of noise level.

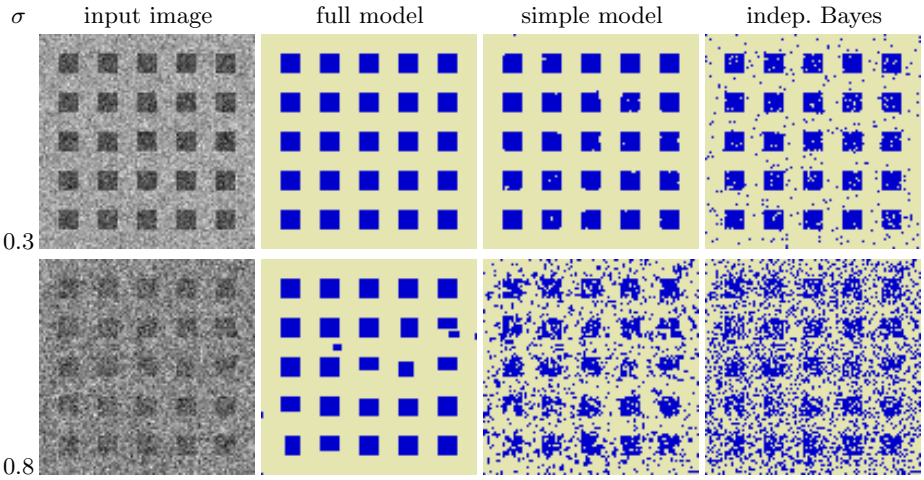

Fig. 5. Results of the labeling, color coded segmentation maps. For the full model, all labels except the background label (E) are displayed in blue.

3.2 Real data

We run the proposed method on several images of real façade. The images were from 0.1 to 0.7 Mpx. Both image model and structure model were learnt from a single façade image, Fig. 6-top image, with manually annotated windowpanes. The model parameters were kept fixed for all images.

The CPU time for the proposed full model is less than 10 seconds per image on C2 2.4 GHz, mostly depending on the scene complexity. The solution for the simple 2-label model is also obtained by [15] in much shorter time. Even faster solution of the 2-labeling problem can be obtained by Max-Flow algorithm, which is polynomial and optimal for this problem.

We can see, Fig. 6 and 7, that most of the windowpanes were correctly identified, despite the large variation in façade and windowpane color. There are few missing windowpanes or false positive detections. The reason for that is two-fold: (1) the actual image and structure model are locally violated, or (2) the solution we obtained from approximate algorithm is not the global optimum.

Comparison of results of the full model, simple model and the local method (independent Bayes) is shown in Fig. 8. We can see the full model forces the windowpanes to be rectangles and helps reject regions which have the appearance (the pixel color) same as true windowpanes, but do not have their structure, e.g. railings, tree branches, shadows. The rejection cannot occur in simple model or in the local method.

We also tested sensitivity of the method to structure model violation. Rotating an image plane breaks the orthographic rectification assumption: windowpanes are not axis-parallel rectangles any more.

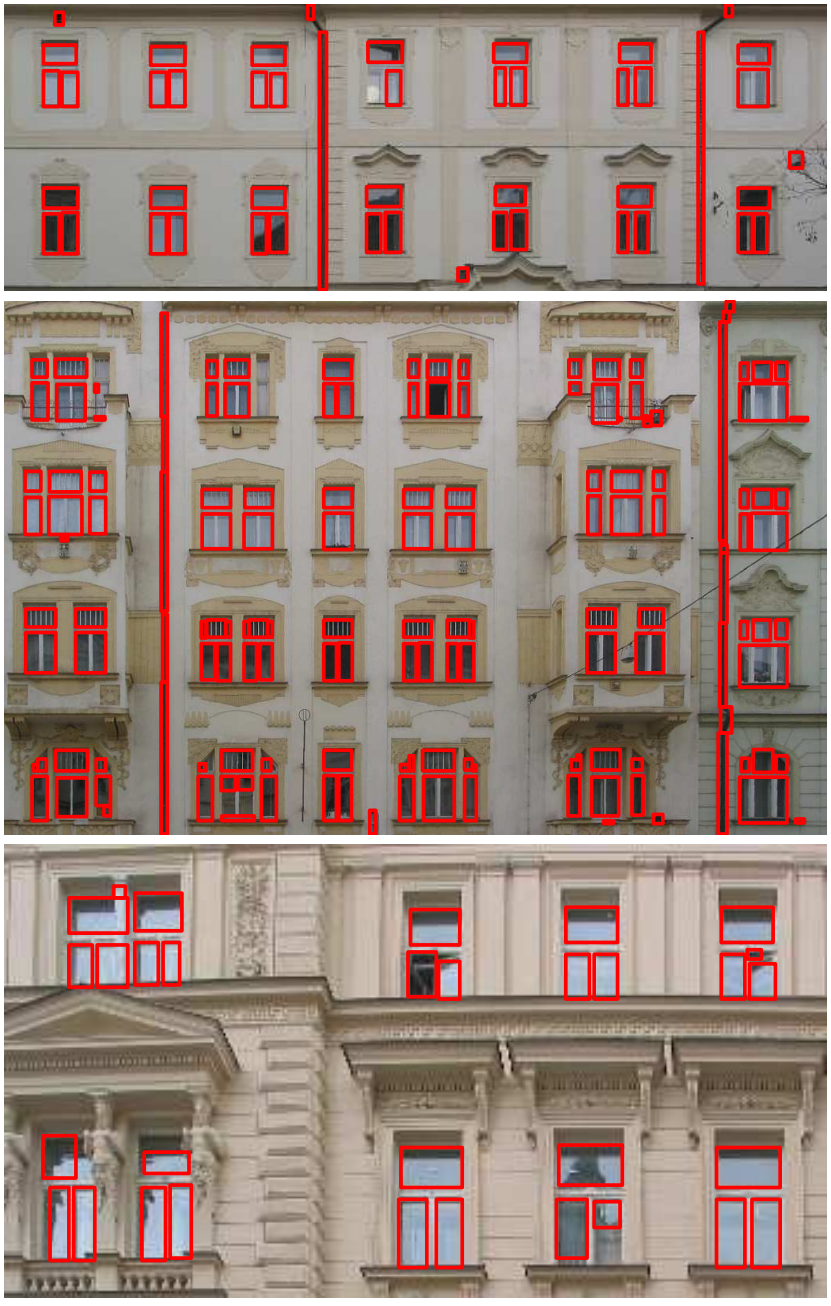

Fig. 6. Results on real façade images. This paper is best viewed in full-color electronic version.

Fig. 7. Results on real façade images.

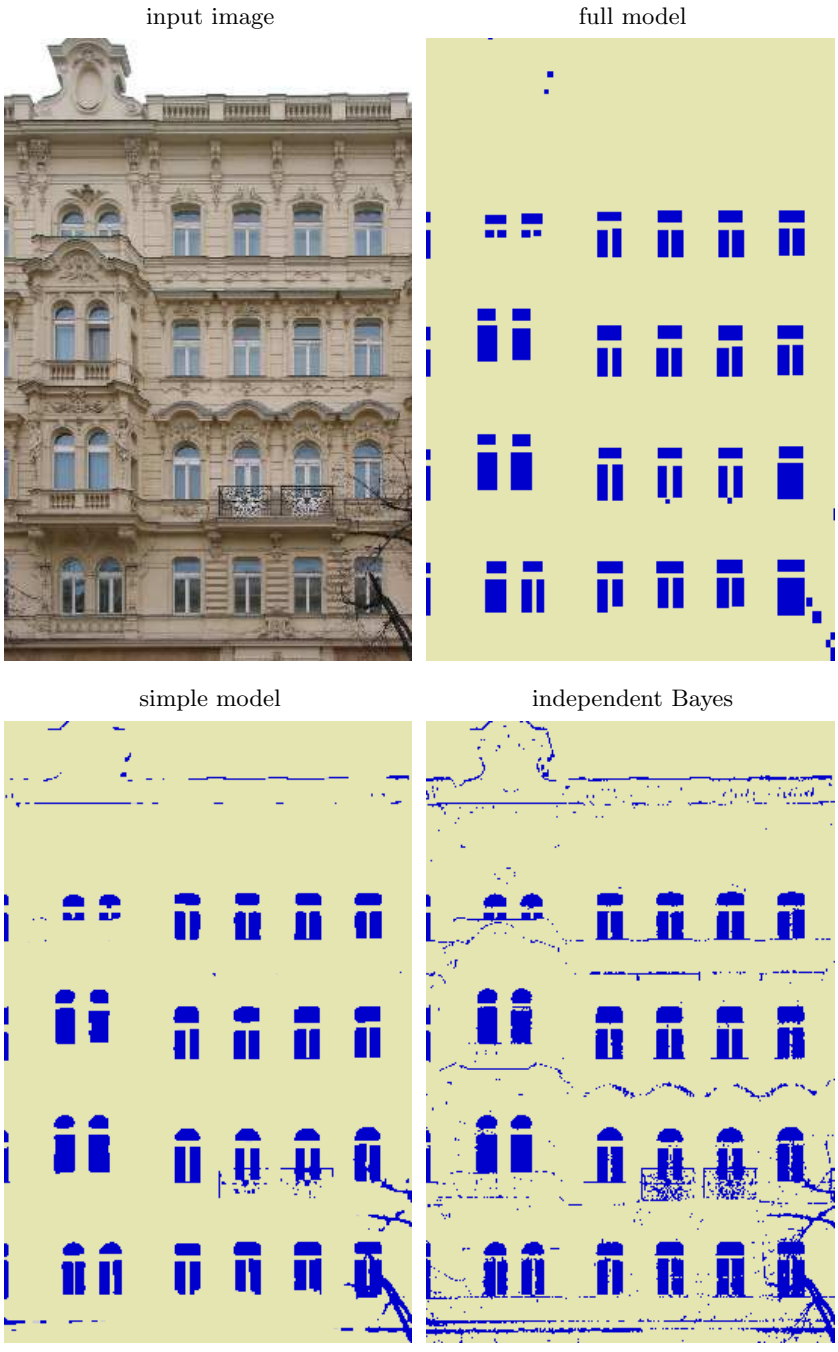

Fig. 8. Segmentation maps using models of decreasing complexity.

Having manually annotated windowpanes as the ground-truth segmentation, we measured the error rate as a function of image plane rotation, Fig. 9 (left). Obviously, the quality of the segmentation using the simple Potts model or independent Bayesian decision is invariant to image plane rotation. However, we can see the segmentation using the full model deteriorates quickly with increasing rotation angle. The full model still results in a set of axis-parallel non-overlapping rectangles, despite the data is far from the model assumptions, see Fig. 9 (right). This is a natural drawback of such a strong model.

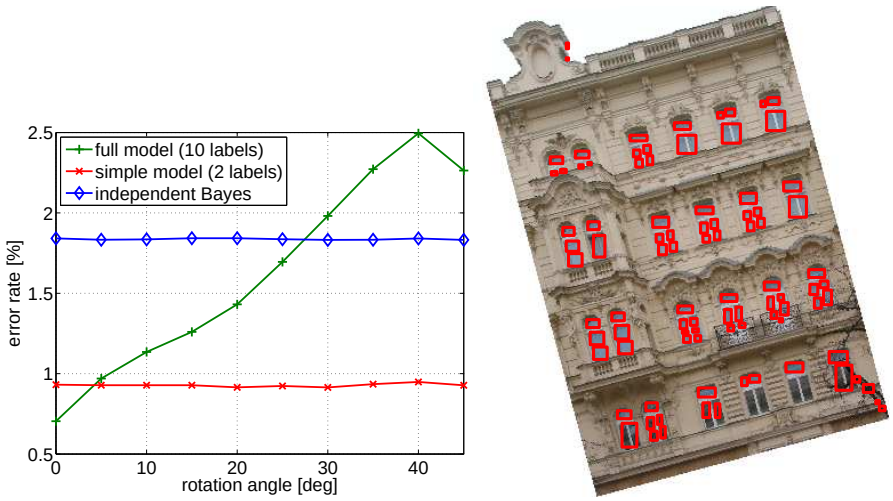

Fig. 9. Error rate as a function of image plane rotation (left). Detection results of the full model for rotation angle of 15 degrees (right).

4 Conclusion

We showed that using the structure model of the region to be segmented has a large impact on the quality of the segmentation results. We showed experimentally on both synthetic and real data that the proposed full structure model (forcing rectangles) outperforms a traditional 2-label Potts model which enforces continuity only without modeling the structure of the region at all. Although our problem leads to NP-complete task, we show that solutions obtained from an approximate algorithm [15, 16] are acceptable.

The paper does not aim at bringing a perfect windowpane detector. Instead, we wanted to present an interesting method which is based on a pure and simple global formulation and whose solution is found by a general solver. Of course, for a later practical usage of the windowpane detector there will be necessary to improve image model and its learning. Some pre/post-processing could also help increase the performance.

Acknowledgement. The research was supported by Czech Academy of Sciences project 1ET101210407 and by EC project FP6-IST-027113 eTRIMS.

References

1. F. Alegre and F. Dellaert. A probabilistic approach to the semantic interpretation of building facades. In *International Workshop on Vision Techniques applied to the rehabilitation of city centers*, pages 1–12, 2004.
2. R.J. Baxter. *Exactly Solved Models in Statistical Mechanics*. Academic Press, New York, 1990.
3. A. Dick, P. Torr, and Cipolla R. Modelling and interpretation of architecture from several images. *IJCV*, 60(2):111–134, 2004.
4. P. F. Felzenszwalb and D. P. Huttenlocher. Efficient belief propagation for early vision. *IJCV*, 70(1), 2006.
5. B. Flach. A diffusion algorithm for decreasing energy of max-sum labeling problem. Fakultät Informatik, Technische Universität Dresden, Germany, 1998. Unpublished manuscript.
6. B. Flach and M. I. Schlesinger. A class of solvable consistent labeling problems. In *Proc. of IAPR International Workshops on Advances in Pattern Recognition*, pages 462–471, 2000.
7. S. Geman and D. Geman. Stochastic relaxation, Gibbs distribution, and the bayesian restoration of images. *IEEE Trans. PAMI*, 6(6):721–741, 1984.
8. Z. Kato and T.-C. Pong. A Markov random field image segmentation model for color textured images. *Image and Vision Computing*, 24:1103–1114, 2006.
9. V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Trans. PAMI*, 28(10):1568–1583, 2006.
10. N. Komodakis, N. Paragios, and G. Tziritaz. MRF optimization via dual decomposition: Message passing revisited. In *Proc. of ICCV*, 2007.
11. V. A. Kovalevsky and V. K. Koval. A diffusion algorithm for decreasing energy of max-sum labeling problem. Glushkov Institute of Cybernetics, Kiev, USSR, 1975. Unpublished manuscript.
12. H. Mayer and S. Reznik. Bulding facade interpretation from image sequences. In *Proc. of ISPRS Workshop CMRT*, pages 55–60, 2005.
13. J. Pearl. *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. The Morgan Kaufmann series in representation and reasoning. Morgan Kaufmann, San Francisco, 1988.
14. M. I. Schlesinger. *Mathematical Tools of Image Processing*. Naukova Dumka, Kiev, 1989. In Russian.
15. T. Werner. A linear programming approach to max-sum problem: A review. Technical Report CTU–CMP–2005–25, Center for Machine Perception, K13133 FEE Czech Technical University, Prague, Czech Republic, December 2005. http://cmp.felk.cvut.cz/cmp/software/maxsum/.
16. T. Werner. A linear programming approach to max-sum problem: A review. *IEEE Trans. PAMI*, 29(7), July 2007.
17. R. Wilson and C.-T. Li. A class of discrete multiresolution random fields and its application to image segmentation. *IEEE Trans. PAMI*, 25(1):42–55, 2002.
18. J. W. Woods. Two-dimensional discrete Markovian fileds. *IEEE Trans. Information Theory*, 18:232–240, 1972.