CENTER FOR
MACHINE PERCEPTION

CZECH TECHNICAL
UNIVERSITY

RESEARCH REPORT

ISSN 1213-2365

Windowpane Detection based on Maximum Aposteriori Probability Labeling

Jan Čech and Radim Šára

cechj@cmp.fek.cvut.cz

CTU–CMP–2007–10

May 10, 2007

The work was supported by the EC project FP6-IST-027113 eTRIMS.

Research Reports of CMP, Czech Technical University in Prague, No. 10, 2007

Published by

Center for Machine Perception, Department of Cybernetics
Faculty of Electrical Engineering, Czech Technical University
Technická 2, 166 27 Prague 6, Czech Republic
fax +420 2 2435 7385, phone +420 2 2435 7637, www: http://cmp.felk.cvut.cz

Windowpane Detection based on Maximum Aposteriori Probability Labeling

Jan Čech and Radim Šára

May 10, 2007

Abstract

Segmentation of windowpanes is formulated as a task of maximum aposteriori labeling. Assuming orthographic rectification of the building facade, the windowpanes are always axis-parallel rectangles of relatively low variability in appearance. Every image pixel has one of 10 possible labels, and the adjacent pixels are interconnected via links which defines allowed label configuration, such that the labels are forced to form 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 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 does 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.

1 Introduction

Markov Random Fields (MRFs) have been used in image analysis for a long time [15]. There are many papers on image segmentation using MRFs, e.g. [14, 7]. 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 introducing a more complicated structure prior model, where pairwise transition probabilities between labels are asymmetric. 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 [13] to demonstrate the functionality of the labeling solver.

The obvious drawback of the proposed approach is that we need more labels compared to similar segmentation with Potts model. We are using 10 labels and the problem is NP-complete as a result. 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 facade images. For instance, in [10, 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 facade. These approaches are very different from our simple formulation based on segmentation.

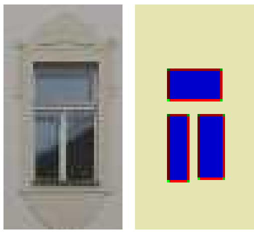
The paper is structured as follows: The problem is formulated in Sec. 2. Experimental validation on both synthetic data and images of real facades is given in Sec. 3. The Sec. 4 concludes the paper.

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, $X = \{$E,I,L,R,T,B,TL,TR,BL,BR$\}$, see Fig. 1.

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 task is to find the most probable labeling given the image data, i.e.



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

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

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, and the assumption of independence, this leads to

$$\mathbf{x}^* = \arg\max_{\mathbf{x}\in X^{|T|}} \frac{p(\mathbf{d}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{d})} = \arg\max_{\mathbf{x}\in X^{|T|}} \prod_t p(d|x_t) \prod_{t,t'} p(x_t, x_{t'}) \prod_t p(x_t), \tag{2}$$

where the notation x_t means the label $x \in X$ at pixel $t \in T$. Please note that $p(x_t)$ is *not* a marginal probability of $p(x_t, x_t')$. Their meaning is explained below.

The term $p(d|x_t)$ reflects data agreement for label x_t. The probability distribution $p(d|x)$ we call the *image model*. It is a probability distribution function of appearance for each individual labels $x \in X$, e.g. the windowpane label (I-label) is typically of dark or glossy color. This is learnt from examples.

The term $p(x_t, x_{t'})$ is a probability of coocurence of neighbouring labels. The probability distribution $p(x, x')$ we call the *structure model*, it represents the adjacent label compatibility, i.e. the rules that generate a 2D language of non-overlapping rectangles.

The last term $p(x_t)$ is a prior probability of the label.

Applying logarithm to (2), we obtain the max-sum labeling problem

$$\mathbf{x}^* = \arg\max_{\mathbf{x}\in X^{|T|}} \sum_t g_t(x_t) + \sum_{t,t'} g_{tt'}(x_t, x_{t'}), \tag{3}$$

where

$$\begin{aligned} g_t(x_t) &= \log p(d|x_t) + \log p(x_t), \\ g_{tt'}(x_t, x_{t'}) &= \log p(x_t, x_{t'}). \end{aligned} \tag{4}$$

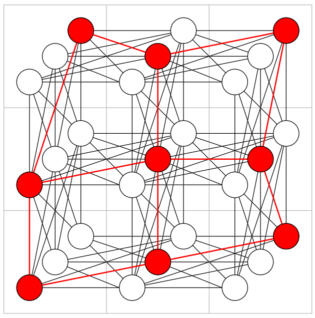


Figure 2: Illustration of the labeling problem.

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. The labels between adjacent pixels are interconnected via edges. Each node has assigned a node quality $g_t(x_t)$, each edge has assigned an edge quality $g_{tt'}(x_t, x_{t'})$, see (4). The task of the max-sum labeling (3) 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. 2in the sense of number of pixels and number of labels (we have 10 labels).

The problem (3) is NP-complete in general. There are solvable sub-classes [6], e.g. the number of labels is 2, or the underlying graph does not contain cycles and others. But, our task does not belong to any of the solvable sub-classes and remains NP-complete. However, there exist approximate algorithms which finds a sub-optimal solution, e.g. (loopy) belief propagation [11, 4], Kolmogorov's TRW-S algorithm [8], max-sum diffusion by Kovalevsky and Flach [9, 5], Schlesinger's linear programming relaxation [12, 13].

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 $c = (r, g, b)$ as red, green, blue intensity channels

$$p(d|x) = p(c|x) = p(r, g, b|x), \tag{5}$$

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 coocurence 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.

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 used Werner's implementation [13] of linear programming relaxation based max-sum solver by Schlesinger [12], since it seems to give good results for our problem, unlike the belief propagation [11, 4] which often oscillates and max-sum diffusion [5, 9] which is very slow.

3 Experiments

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

3.1 Synthetic experiment

We use a synthetic gray-scale test image simulating a facade, Fig. 4. It is 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) independently. 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), the local Bayes classifier - the Bayes decision incorporating prior probability of labels

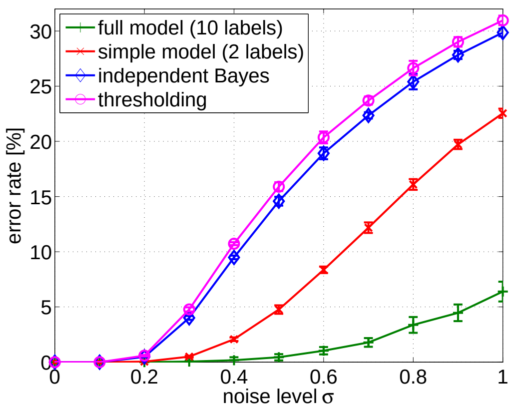


Figure 3: Labeling error rate as a function of noise level.

in each pixel independently (Independent Bayes) and thresholding with the optimal threshold $\tau = 0.5$.

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 are in Fig. 3. The results are averaged from 10 random trials and the plots have errorbars. The local decisions (thresholding and independent Bayes) have worse performance than methods modeling pixel neighbourhood relations (simple and full model). The independent Bayes decision is slightly better than thresholding, since the prior probabilities of segment does not differ much. The simple 2-label Potts model MRF segmentation outperforms both local methods, which is a known fact in segmentation literature. The full 10-label model (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 axis parallel non-overlapping rectangles has a large impact here.

The segmentation results for $\sigma = 0.3$ and $\sigma = 0.8$ are in Fig. 4. 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 methods does a few errors for such a noise level, it performs much better than the simple model.

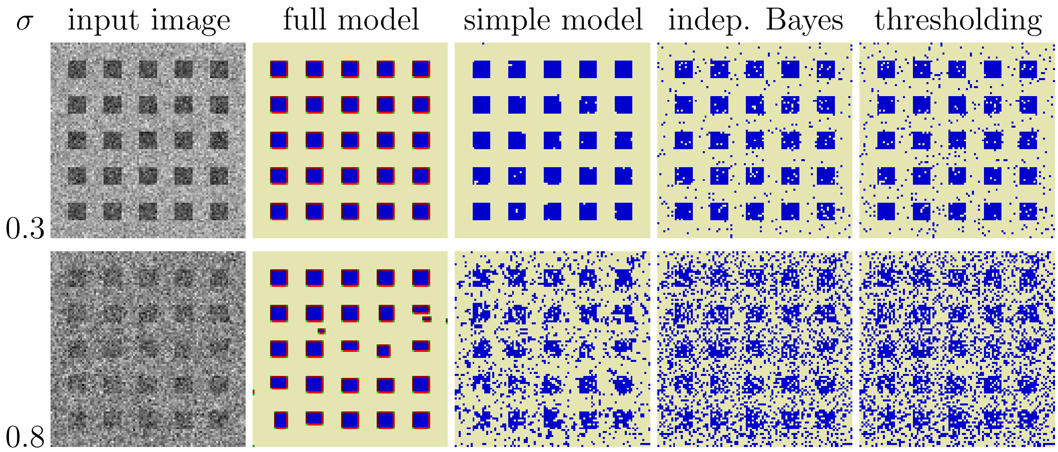


Figure 4: Results of the labeling. Labeling maps are color coded.

3.2 Real data

We run the proposed methods on several images of real facade. The images were from 0.1 to 0.7 Mpx. Both image model and structure model were learnt from a single facade image, Fig. 5-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 [13] 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. 5 and 6, that most of the windowpanes were correctly identified, despite the large variation in facade and windowpane color. The are few missing windowpanes or false positive detection. The reason for that is two-fold: (1) the actual image and structure model is wrong or imprecise, or (2) the solution we obtained from approximate algorithm is not a global optimum.

Comparison of results of full model and simple model is in Fig. 7. There are labeling map for image 6-top right. 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 truth windowpanes, but have not their structure, e.g. railing, tree branches, shadows, which persists at simple model results.

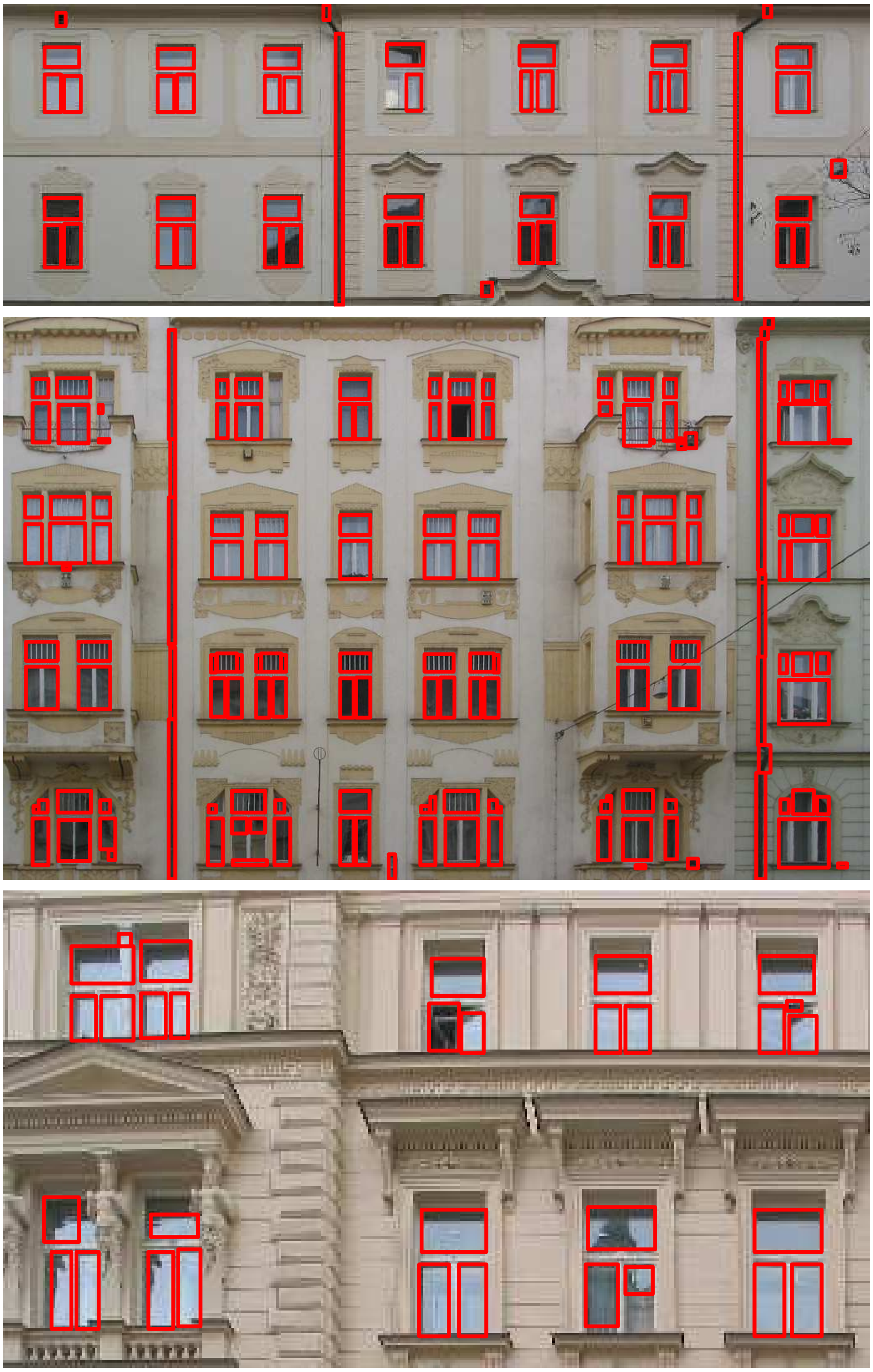


Figure 5: Results on real facade images

Figure 6: Results on real facade images

full model | simple model

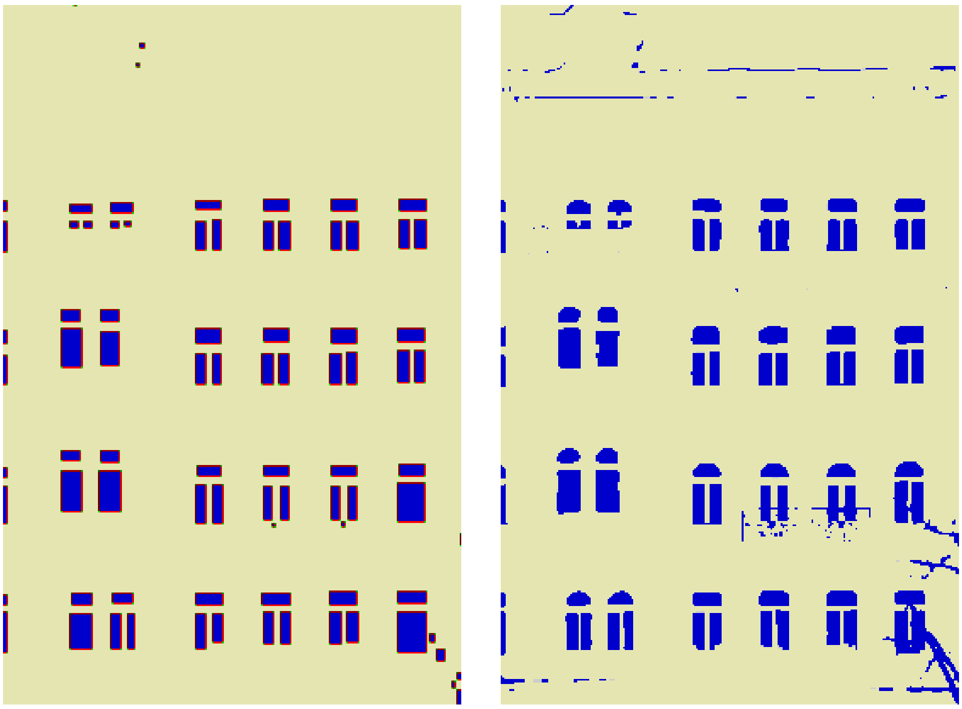


Figure 7: Labeling maps.

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 solution obtained from approximate algorithm [13] is acceptable.

The paper does not aim to bring 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 take more care to image model and its learning. Some pre/post-processing could also help to increase the performance.

References

[1] F. Alegre and F. Dallaert. A probabilistic approach to the semantic interpretation of building facades. In *Internatuional 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] Zoltan Kato and T.-C. Pong. A Markov random field image segmentation model for color textured images. *Image and Vision Computing*, 24:1103–1114, 2006.

[8] V. Kolmogorov. Convergent treee-reweighted message passing for energy minimization. *IEEE Trans. PAMI*, 28(10):1568–1583, 2006.

[9] 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.

[10] H. Mayer and Reznik S. Bulding facade interpretation from image sequences. In *Proc. of ISPRS Workshop CMRT*, pages 55–60, 2005.

[11] Judea Pearl. *Probabilistic reasoning in intelligent systems : networks of plausible inference*. The Morgan Kaufmann series in representation and reasoning. Morgan Kaufmann, San Francisco, 1988.

[12] M. I. Schlesinger. *Mathematical Tools of Image Processing*. Naukova Dumka, Kiev, 1989. In Russian.

[13] 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/.

[14] 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.

[15] J. W. Woods. Two-dimensional discrete Markovian fileds. *IEEE Trans. Information Theory*, 18:232–240, 1972.