2D LATTICE EXTRACTION FROM STRUCTURED ENVIRONMENTS

Thommen Korah

Dept. Computer & Information Sciences
University of Delaware
korah@cis.udel.edu

Christopher Rasmussen

Dept. Computer & Information Sciences
University of Delaware
cer@cis.udel.edu

ABSTRACT

In this paper we investigate the problem of automatically detecting 2D grid structures such as windows on building facades from images taken in urban settings. The key assumption that the background is strongly structured allows searching for near-regular textures in the image. We describe a probabilistic framework using Markov Random Field modeling and Markov Chain Monte Carlo (MCMC) optimization to explicitly recognize and group rectangular structures that appear in a grid-like pattern. Results on a variety of images of building facades are shown.

***Index Terms*—** Image analysis, Segmentation, Pattern Recognition

1. INTRODUCTION

The motivation of this work is to parse images of man-made environments. It is part of a larger goal of robot-based urban modeling by incorporating semantic information into structure recovery and texture map construction. Purely geometric methods such as structure from motion rely on low-level features like points or lines. However, by restricting our application domain to structured environments like cities and campuses, we can exploit other high-level cues for modeling. One particular feature of urban scenes that we wish to detect are 2D lattices such as grids of windows or bricks.

Several researchers have attempted to use characteristics of man-made scenes for tasks such as camera calibration[1] and structure from motion [2]. Grammar-based interpretation of building facades in a Markov Chain Monte Carlo (MCMC) framework has been described by Alegre and Dellaert [3]. Mayer and Reznik [4] also use MCMC sampling for locating windows. However, both methods use highly specialized models and show examples on a very restricted set of images. We propose a more general grouping algorithm that is able to locate grid structures in a variety of building images. We also demonstrate how this discovered lattice can be used to infer occluded windows.

We frame the window extraction problem as one of texture discovery. Several papers have examined the problem of finding regular, planar patterns, with notable approaches based on RANSAC [5] and the cascaded Hough transform [6]. An algorithm for near-regular texture (NRT) discovery by lattice growth was described in [7]. Their method, which iteratively grows a lattice from a seed location, worked well on many images we tested but did not find the entire textured region when blocked by significant occlusions. Besides being computationally expensive (averaging over half an hour per image on an Intel Core2 Extreme X6800), their method finds tiles which do not correspond exactly to semantically meaningful units. Our approach benefits from several strong assumptions about the kind of texture we are looking for because of the urban scene domain.

In this paper, we describe an algorithm for grouping rectangular structures that appear on building facades. We first describe the bottom-up discriminative process that involves generating several hundred rectangle hypotheses from an image. Then a top-down grouping algorithm combines appearance, shape and topology cues in a Markov Chain Monte Carlo framework to group rectangles that form part of a 2D lattice on the facade. We show results on various images and demonstrate how this lattice discovery could be used for modeling applications.

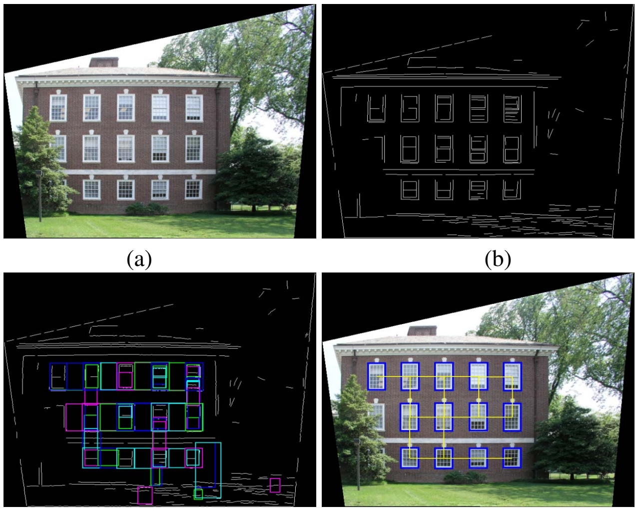
2. RECTANGLE HYPOTHESES

The first stage of our algorithm consists of finding as many straight lines as possible. Canny edge detection is applied to locate high gradient pixels in the image. Similar to the technique of Zhang and Kosecka [1], the orientation of each edge pixel is quantized into 8 bins. However, instead of computing the gradient direction from the Sobel operator, we convolve the image with filters from the RFS filterbank [8] at 8 orientations and a single scale. Each pixel is assigned to the bin of the filter that had the maximal response. This technique was found to be more robust to noise, especially in the presence of occlusions or weak edges. Adjacent edge pixels in the same orientation bin are grouped together using connected components to obtain a number of straight lines from the image. For efficiency, we retain only lines that had more than 15 pixels (approximately 2% of image dimensions) grouped together. Each connected component is parameterized as a line by its centroid, slope and magnitude. This is computed using least squares fitting.

We assume that our image has been rectified or the scene is mostly fronto-parallel where perspective effects are negligible. This allows us to represent rectangular structures with just 4 parameters – upper-left corner p_k, width w_k and height h_k for rectangle k. A discriminative bottom-up approach for rectangle finding has been mentioned in [1]. In that work, rectangles under perspective are hypothesized from two pairs of lines, each of which is picked from two different vanishing directions. The validity of the hypotheses is verified by looking for corner features in the image. Our rectangle detection is a top-down approach similar to [9], and can therefore afford to be more conservative about which rectangles pass through to the next stage for grouping.

To generate possible rectangles in the image, pairs of line segments (l_i, l_j) are exhaustively drawn and tested for parallelism. Due to our assumptions, these lines are also restricted to be within 10 degrees of the horizontal and vertical directions. For $l_i$and $l_j$to be parallel, they must satisfy the following conditions:

- Magnitude difference should be within 10% of the minimum length of the two.
- At least 90% of overlap between the lines in its orthogonal direction.



(a) (b) (c) (d)

Fig. 1. Extracting 2D window grids (a) Original rectified image; (b) straight lines fitted to canny edges; (c) Generating rectangle hypotheses from pairs of parallel line segments;(d) discovered lattice with inferred neighborhood relationships using our MCMC method. Image (a) was automatically rectified.

- The resulting rectangle should have an aspect ratio $a = \frac{width}{height}$ in the range $0.3 \leq a \leq 2$. This is typical of windows found on most buildings.

Every pair of lines that satisfy these constraints is used to propose a rectangle. Compared to [1] and [9] which needs all 4 lines, the sufficiency of line pairs makes it less sensitive to broken edges in the edge detection. Moreover, the strength of the higher-level grouping stage ought to overcome any ambiguities caused by having too many rectangles. Figures 1(b) and 1(c) show fitted straight lines and hypothesized rectangles for an input image. These rectangles are then passed to our lattice discovery algorithm described below.

3. DISCOVERING BUILDING TEXTURE

3.1. MRF Grid Extraction

We define a lattice-based Markov Random Field (MRF) [10] to model the grid topology. Similar models have been used for microarray analysis [11] and tracking deformable NRT lattices [12]. The rectangles generated by the technique described above constitute the nodes of the MRF graph. Other image discretization methods such as interest points and correlation map peaks [7], or superpixels of an over-segmented image [13] are also possible.

Similar to the texel proposal method of [7], an initial undirected grid graph $G_0 = (V, E)$ is constructed by a greedy scheme whereby each region connects with its lowest cost neighbor. The cost $E_c(v_i, v_j)$ for each pair of nodes v_i, v_j is measured as the total number of other nodes within a threshold distance of the line parameterized by the two nodes, scaled by rough shape similarity

$$B(v_i, v_j) = e^{-\alpha(|w_i - w_j| + |h_i - h_j|)}$$

. G_0 is then refined based on topology, shape and appearance following similar approaches to graph representation of adjacent regions in an over-segmented image [13, 14]. In our case, the nodes that are part of the lattice can be distributed over the entire image, and every pair of assignments between two nodes should be consistent in that they can be derived from a higher-order topological constraint T such as a 2-D rectangular grid. Thus grouping involves creating links along vectors $\mathbf{t}_i^o : o \in \{r, l, u, d\}$ to the most likely right, left, up and down neighbors of v_i without violating grid constraints. Given image I, we wish to obtain the MAP estimate

$$p(G|I, T, S) \propto p(I|T, S, G)p(T|G)p(S|G)p(G). \quad (1)$$

All the results in the paper have been produced with the image likelihood $p(I|T, S, G)$ set to unity. Color histograms, proximity of rectangle boundaries to image edges, or learned appearance models are all possible likelihood models. The shape prior $p(S|G)$ can be used to favor known shape models, though here we set it to unity since we are only dealing with rectangles. The graph prior $p(G)$ is set to be linear in the number of connected edges and unity if at least 50% of the nodes are connected. The topology prior $P(T|G)$ is represented as a pairwise MRF whose joint can be factored into a product of local node potentials Φ and clique potentials Ψ:

$$P(T|G) \propto \prod_i \Phi(v_i) \prod_{i,j \in E} \Psi(v_i, v_j).$$

To model a grid, we measure the symmetry of direction vectors from a node to its neighbors. Let $\delta(\mathbf{t}_1, \mathbf{t}_2) = e^{-\beta||\mathbf{t}_1 - \mathbf{t}_2||}$ be a similarity measure between two neighbor vectors assuming both edges are in G. The potentials are now defined as:

$$\Phi(v_i) = e^{-\gamma(4-n_i)} * \delta(\mathbf{t}_i^r, -\mathbf{t}_i^l) * \delta(\mathbf{t}_i^u, -\mathbf{t}_i^d),$$

$$\Psi(v_i, v_j) = \delta(\mathbf{t}_i^u, \mathbf{t}_j^u) * \delta(\mathbf{t}_i^d, \mathbf{t}_j^d) * B(v_i, v_j).$$

where n_i denotes the degree of node v_i. Thus we encourage left/right and up/down edge pairs to be 180 degrees apart with similar magnitudes. The interaction potential between horizontal neighbors forces their vertical edges to be approximately parallel. For missing edges, a small fixed value is assigned to δ. Even with these simple potentials, we are able to model many rigid and non-rigid grid configurations from the CMU NRT database [15].

3.1.1. Optimization

We use a Markov Chain Monte Carlo (MCMC) framework to iteratively maximize (1), probabilistically adding and removing edges from G_0 in a fashion similar to the multi-target tracking method of [16]. A Markov chain is defined over the space of configurations $\{G\}$ and the posterior is sampled using the Metropolis-Hastings [17] algorithm. A new state G_t' is accepted from state G_t with probability $p = \min(1, \frac{p(G_t'|I,T,S)q(G_t|G_t')}{p(G_t|I,T,S)q(G_t'|G_t)})$.

Proposal updates $q(G_t'|G_t)$ consist of edge additions or removals applied to a uniformly selected node v_k. Modifying one component at a time leads to a better success rate for transitions, although large state changes require more jumps. The edge transitions are made only in the up and right directions in order to keep the reverse transition probability simple. Let $e_{kl} : l \in \{1, \ldots, n_k\}$ be the edges from v_k to its neighbors. The edge in each direction is turned off with fixed probability p_{off} or assigned a neighbor by sampling from $E_c(k, \cdot)$. The simulation is run for 10000 iterations and the maximum a posteriori state estimate is returned.

4. LATTICE COMPLETION AND INFERENCE

The best state configuration returned by the MCMC simulation consists of separate connected components, each of which form a 2D grid configuration. On all the images in this paper, the window grid could be identified as the largest such component. On more complex

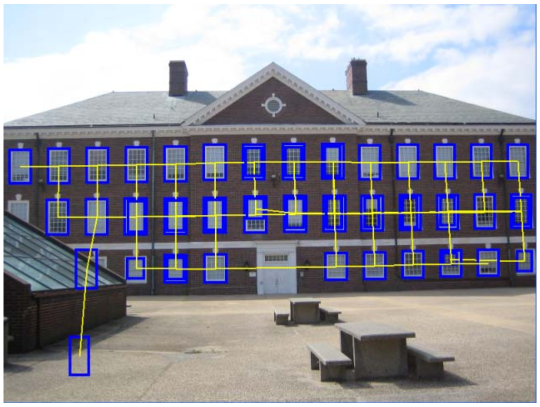


Fig. 2. The dominant 2D grid after grouping rectangles together. Since the grid is modeled locally using only pairwise relationships, it might deviate slightly from a single global model. This can be corrected in a post-processing stage.

images, alternative image-based metrics can be used to extract the most likely window grid.

Our method also facilitates inferring parameters of the regular grid from the grouped elements. The rectangle's median height and width as well as magnitudes of the horizontal and vertical $\mathbf{t}$ vectors are computed first. We then pick the node with the highest likelihood according to (1) as an origin. This completely specifies a regular grid that can be overlaid on the image to hypothesize missing or occluded lattice elements.

This approach is not as accurate if the actual windows deviate from the perfect grid assumption. For instance, windows on the ground floor have slightly different dimensions. However, we observed that windows on the same floor are similar and are centered horizontally and vertically with its neighbors. As explained in the previous paragraph, the parameters describing a regular grid (and which best approximates the discovered lattice) are first computed. Then for every window location, we test whether an actual window was detected in the grouping stage or not. If so, that window is drawn according to its parameters. Otherwise, a missing grid element is inferred at the location aligned horizontally and vertically with its neighbors. The size is set to be the same as that of its horizontal neighbor.

5. RESULTS

Our algorithm was tested on a variety of images consisting of highly regular and near-regular window lattice textures. Figure 1 shows results from each stage; bottom-up line fitting, rectangle hypotheses generation and finally the grouping algorithm described in section 3. The yellow lines in figure 1(d) denote the neighborhood relationships inferred between the grid elements. Zooming into the image reveals that rectangles have been localized accurately around all visible windows forming the grid lattice.

Figure 2 shows another example of a discovered grid. There were 571 lines fitted to the Canny edges from which 271 rectangles were proposed. 53 elements were discovered as part of the grid. As can be seen, the grouping algorithm does not require images to be rectified. Since we don't use the image likelihood function, it is hard to distinguish between an actual window and a spurious one as long as they are topologically consistent with the rest of the grid.

Figure 3 show more examples where the discovered grid (drawn in blue) is used to identify missing or occluded grid elements (drawn in red). The images in the top row are highly regular and can be fully described by a single set of global parameters. Predicting the location of occluded windows in figures 3 (c) and (d) however required the more adaptive approach detailed in section 4.

Figure 4 illustrates how this kind of semantic knowledge can be used to obtain clean texture maps for modeling. We can collect per-pixel statistics about the appearance of windows to detect and remove outlier elements. The reflected dome in fig. 3(b) has been scrubbed clean.

6. CONCLUSION

We have demonstrated a promising approach for accurate discovery and localization of 2D lattice structures from a single static image. Motivation for this work is to exploit higher-level domain cues found in man-made environments for a robot-based architectural modeling project. An MRF model for 2D lattices is defined where each node is a rectangle proposed by a bottom-up method. These rectangles are grouped together by topological constraints using an MCMC optimization framework.

Much work still remains to be done. We are currently experimenting with various image likelihood functions. Occlusions and errors in the edge detection might cause missing nodes in the lattice or generate a rectangle in a uniform part of the image. A drawback of the MRF model with local pairwise relationships is that there are no global constraints that force all nodes to adhere to a grid structure. Nevertheless, we demonstrate how topological constraints alone can assist in scene understanding.

7. REFERENCES

[1] Wei Zhang and Jana Kosecka, "Extraction, matching and pose recovery based on dominant rectangular structures," *High Level Knowledge in Vision Workshop*, 2003.

[2] Grant Schindler, Panchapagesan Krishnamurthy, and Frank Dellaert, "Line-based structure from motion for urban environments," in *3D Data Processing Visualization and Transmission (3DPVT)*, 2006.

[3] Fernando Alegre and Frank Dellaert, "A probabilistic approach to the semantic interpretation of building facades," in *International Workshop on Vision Techniques Applied to the Rehabilitation of City Centres*, 2004.

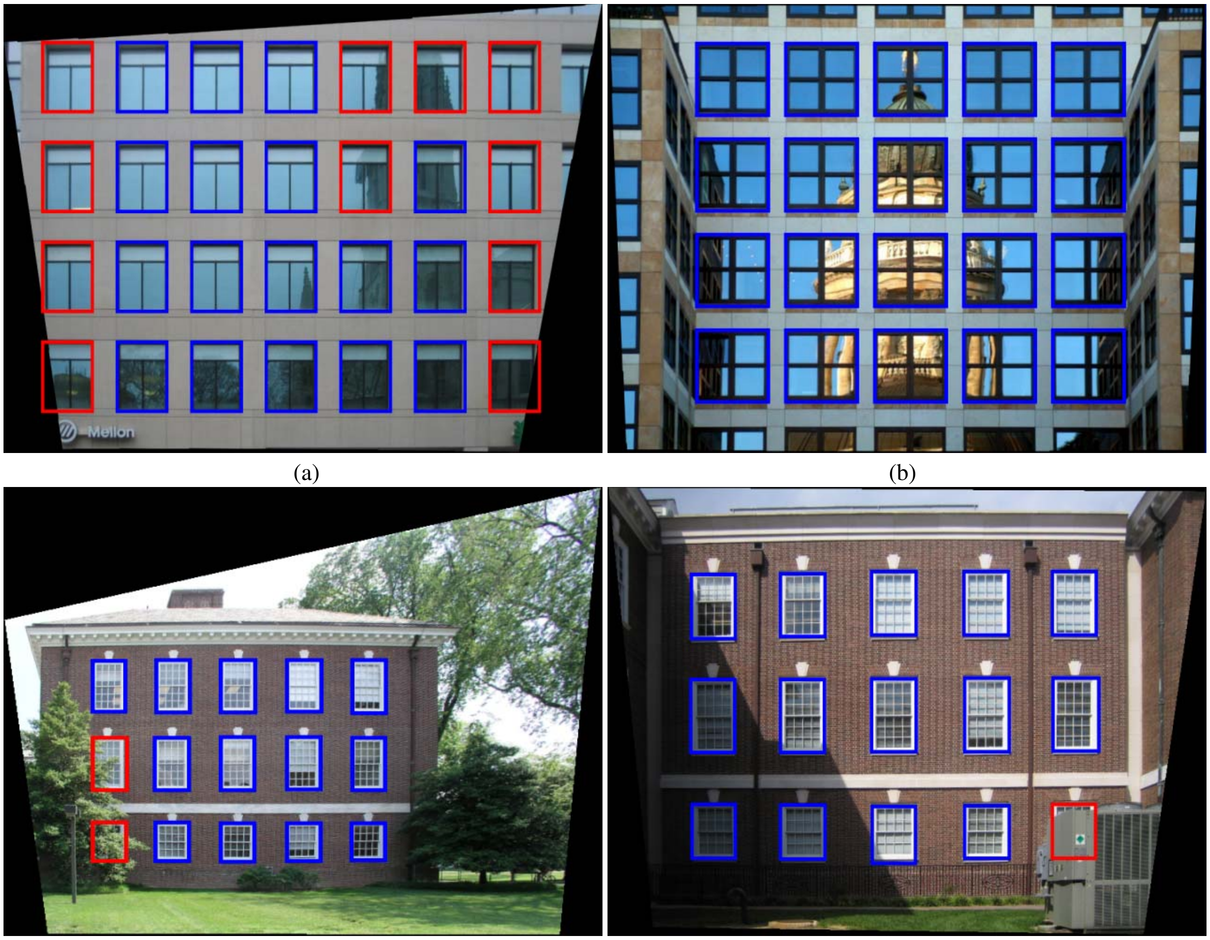
[4] H. Mayer and S. Reznik, "Building faade interpretation from image sequences," in *Proc. of the ISPRS Workshop CMRT 2005 - Object Extraction for 3D City Models, Road Databases and Traffic Monitoring - Concepts, Algorithms and Evaluation*, 2005.

[5] F. Schaffalitzky and A. Zisserman, "Geometric grouping of repeated elements within images," in *Proceedings of the 9th British Machine Vision Conference, Southampton*, 1998.

[6] A. Turina, T. Tuytelaars, and L. Van Gool, "Efficient grouping under perspective skew," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2001.

[7] James H. Hays, Marius Leordeanu, Alexei A. Efros, and Yanxi Liu, "Discovering texture regularity as a higher-order correspondence problem," in *9th European Conference on Computer Vision*, May 2006.

[8] M. Varma and A. Zisserman, "A statistical approach to texture classification from single images," *International Journal of Computer Vision: Special Issue on Texture Analysis and Synthesis*, vol. 62, no. 1–2, pp. 61–81, April 2005.



(a) (b) (c) (d)

Fig. 3. Discovered grid shown as blue rectangles on a variety of images. Images were automatically rectified as a pre-processing step. Rectangles plotted in red are occluded or missing windows inferred from the result of grouping.

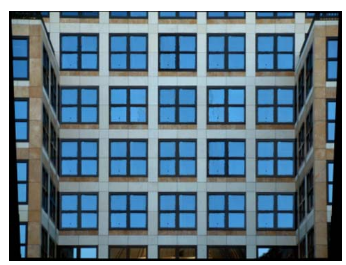


Fig. 4. Recovering clean texture maps by collecting per-pixel statistics over all windows.

[9] Feng Han and Song-Chun Zhu, "Bottom-up/top-down image parsing by attribute graph grammar," in *Proc. of the IEEE International Conference on Computer Vision (ICCV05)*, 2005.

[10] S. Z. Li, *Markov random field modeling in computer vision*, Springer-Verlag, 1995.

[11] Karsten Hartelius and Jens Michael Carstensen, "Bayesian grid matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 2, pp. 162–173, 2003.

[12] Wen-Chieh Lin and Yanxi Liu, "Tracking dynamic near-regular textures under occlusion and rapid movements," in *9th European Conference on Computer Vision*, 2006.

[13] Adrian Barbu and Song-Chun Zhu, "Generalizing swendsen-wang to sampling arbitrary posterior probabilities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, pp. 1239–1253, 2005.

[14] Jingbin Wang, Erdan Gu, and Margrit Betke, "Mosaicshape: Stochastic region grouping with shape prior," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005.

[15] James Hays and Yanxi Liu, "Carnegie mellon near regular texture database," http://graphics.cs.cmu.edu/data/texturedb/gallery/albums.php.

[16] Zia Khan, Tucker Balch, and Frank Dellaert, "Mcmc-based particle filtering for tracking a variable number of interacting targets," *Pattern Analysis and Machine Intelligence*, vol. 27, no. 11, pp. 1805–1918, November 2005.

[17] W.K. Hastings, "Monte carlo sampling methods using markov chains and their applications," *Biometrika*, vol. 57(1), pp. 97–109, 1970.