FACADE REPETITION EXTRACTION USING BLOCK MATRIX BASED MODEL

Hongfei Xiao, Gaofeng Meng, Lingfeng Wang, Shiming Xiang and Chunhong Pan

Institute of Automation, CAS, {hfxiao,gfmeng,lfwang,smxiang,chpan}@nlpr.ia.ac.cn

ABSTRACT

Repetition extraction plays an important role in facade image analysis. In this paper, this task is handled within the graph cut based image segmentation framework. To model the repetitions, generalized translation symmetry (GTS) is introduced to enable aperiodic repetition layouts. More importantly, GTS is explicitly formulated in terms of matrix multiplication. That is, GTS is viewed as the product of a repetitive pattern and two block matrices. These two block matrices are employed to represent the vertical and horizontal symmetry respectively. On this basis, repetition extraction is formulated as a GTS constrained energy minimization problem. An alternatively optimization algorithm based on graph cut and dynamic programming is finally developed to solve the problem. Experimental results demonstrate the validity of our method.

***Index Terms*—** Repetition Extraction, Facade Labeling, Explicit Formulation, Image Segmentation

1. INTRODUCTION

One main task of facade image analysis [1, 2, 3, 4] is to extract the repetitive structures, such as windows, doors and other architectural elements. Actually, repetitive structures provide rich layout and texture information which is important in realistic building reconstruction [5, 6]. In addition, repetition extraction can assist facade image analysis to deal with the challenges coming from appearance variations, external occlusions and changing illuminations.

In the literature, there are two mainstreams of works modeling the repetitions: symmetry based detection [7, 8, 9] and grammar based parsing [10, 11, 12].

Symmetry based detection models the repetitions as translation symmetry. Previous work involved in translation symmetry detection include maximizing repetition quality [8], calculating similarity map [13], and generating translation map [14]. However, if varying distances exist among the repeated elements, they will be grouped into several dissociated lattices. Moreover, none of these approaches propose an explicit formulation for the repetitions. Recently, J. Liu et al. proposed an elegant method which formulates facades via Kronecker Product [9]. Yet, this approach intrinsically makes periodic assumption and thus cannot tackle facades with distance variations.

Grammar-based parsing applies shape grammar to interpret the facade regularity. Shape grammar recursively splits the facade images into basic shapes according to the parsing rules. Given an image, the splitting parameters are inferred via Markov Chain Monte Carlo [15], or random walk [10], or reinforcement learning [11]. However, when tested on an image several times, the results may be inconsistent. Rank-one approximation [16, 17] viewed grammar parsing as approximating an initial label with several rank-one matrices. Yet, the rank-one model does not constrain the matrix size and thus cannot faithfully describe the repetition property.

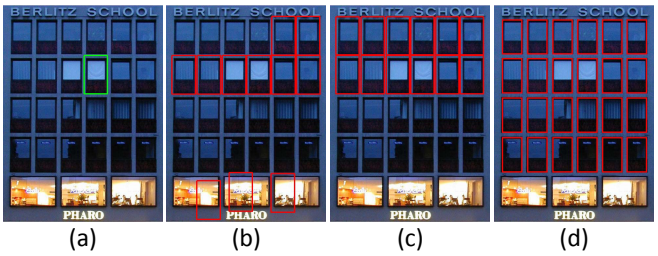

Fig. 1. Overview of our method. (a) The original image with a user-specified bounding box. (b) Template matching result. (c) Initial detection regularized from template matching. (d) The final result.

Our goal is to extract the repetitions given a user-specified bounding box. The core idea is to embed the regularity into the graph cut based image segmentation [18, 19]. The energy on the graph includes the likelihood term and the smoothness term. To model the regularity, we introduce a novel type of symmetry referred as generalized translation symmetry (GTS). GTS enables varying distances between the repetitive elements. Further, GTS is explicitly formulated via two block matrices which respectively express the symmetry along the vertical and horizontal directions. Constrained by GTS, the energy minimization problem is finally constructed. To solve the problem, dynamic programming (DP) and graph cut are alternatively utilized.

The characteristics of our method can be highlighted as follows: (i) we explicitly and intuitively describe the repetition property via block matrices; (ii) repetition detection and segmentation are jointly formulated as a symmetry constrained energy minimizing problem; (iii) the problem is alternatively optimized via DP and graph cut.

2. PROBLEM FORMULATION

In this section, GTS is introduced to model the repetitions, and is represented via matrix multiplication. Then, repetition extraction is formulated as a GTS constrained energy minimization problem.

This research work is supported by National Natural Science Foundation of China under grants 61203277, 61272331, 61370039 and 61305049.

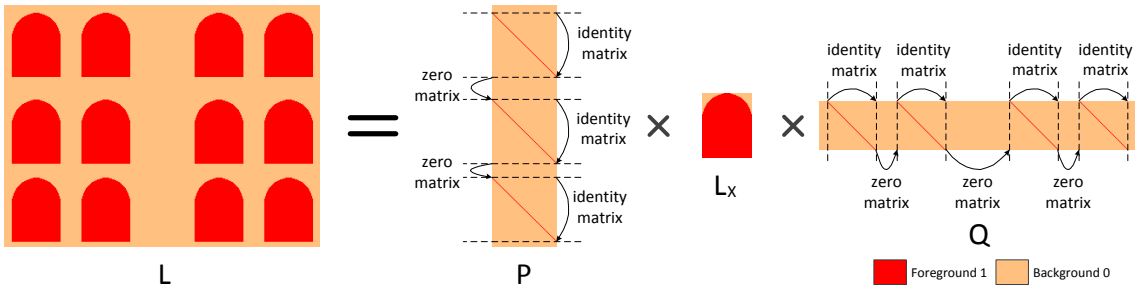

Fig. 2. A sketch for the explicit formulation of the facade label L. L can be formulated as the product of multiplying a repetitive sub-label L_X by two block matrices P and Q. P and Q are composed of alternating zero matrices and identity matrices. Specifically, for Q (P), the height (width) of its identity matrix equals the width (height) of L_X, and the identity matrices are located at the same columns (rows) of the repetitions. Hence, P and Q respectively represent the vertical and horizontal symmetries.

2.1. Formulating GTS via Matrix Multiplication

Real-world facades usually contain vertically and horizontally aligned repetitions. While many methods view the regularity as periodic translation symmetry, GTS is introduced to enable changing distances between the repetitions. For conciseness, we focus on the single symmetry case, namely, viewing one kind of repetitions as foreground and other parts of the facade as background.

For a rectified facade image $I \in \mathbb{R}^{h\times w}$, its 0-1 label L can be formulated in terms of matrix. That is, L can be modeled via two special block matrices whose elemental blocks are identity matrices. Let $L_X \in \mathbb{R}^{m\times n}$ denote the 0-1 label of a repeated element. As illustrated in Fig. 2, we have:

$$L = P \cdot L_X \cdot Q, \tag{1}$$

where P and Q encode the symmetry along vertical and horizontal directions respectively. To guarantee the repetition property, Q (P) must obey the *structured constraint*: being composed of alternating zero matrices and identity matrices along the horizontal (vertical) direction. Let $E \in \mathbb{R}^{n\times n}$ denote the identity matrix. Formally, the *structured constraint* for $Q \in \mathbb{R}^{n\times w}$ can be formulated as:

$$\begin{aligned} &1.\quad \forall i \in \{1,2,\dots,n\}, j \in \{1,2,\dots,w\}, Q_{ij} \in \{0,1\};\\ &2.\quad \forall i \in \{1,2,\dots,n\}, j \in \{1,2,\dots,w\},\\ &\qquad \text{if } Q_{ij} = 1, \text{then } Q(:, j-i+1 : j-i+n) = E. \end{aligned} \tag{2}$$

Furthermore, it should be mentioned that the *structured constraint* for P is similar to that for Q.

2.2. GTS Constrained Energy Minimization Problem

Based on the above facade regularity model, we formulate repetition extraction as a GTS constrained energy minimization problem:

$$\begin{aligned} \min_{L} \quad & U(I,L) + \lambda \cdot V(I,L)\\ \text{s.t.} \quad & L = P \cdot L_X \cdot Q,\\ \text{and} \quad & P,\ Q \text{ satisfy the structured constraint (2)}, \end{aligned} \tag{3}$$

where the term U and the term V will be discussed later, and λ is a positive trade-off between them. Note that optimizing P and Q signifies repetition detection, and calculating L_X stands for segmenting a repetitive pattern. Hence, repetition detection and segmentation are jointly formulated in a unified framework.

The term U expresses the preference of each pixel to be labeled as foreground or background. To evaluate this preference, the foreground and background RGB color distributions are respectively approximated with a full-covariance Gaussian Mixture Model (GMM) [19]. Let L_i be the label of pixel i and let I_i denote the RGB color of pixel i. The data term is defined as:

$$U = \sum_i -\log p(I_i|L_i;\theta), \tag{4}$$

where $p(\cdot)$ is the probability density calculated according to the GMM model, and θ is the parameters of GMMs. The parameters are initialized with k-means. In the optimization process, the parameters are iteratively updated.

The term V is used to enforce the smoothness of local regions and to align the boundary of label with image edges. Let $\mathcal{N}$ be the set of pairs of adjacent pixels. This term is defined as [18]:

$$V = \sum_{(i,j)\in\mathcal{N}} \frac{[L_i \neq L_j]}{\|i-j\|} \cdot \exp\left(-\frac{\|I_i - I_j\|^2}{2\sigma^2}\right), \tag{5}$$

where $[\cdot]$ equals 1 if $L_i \neq L_j$ and 0 otherwise, and $\|\cdot\|$ is the euclidean norm. The constant σ^2 is estimated as the average variances over the whole image. This term adaptively imposes strong or weak penalty for discontinuities between adjacent pixels.

3. OPTIMIZATION

The problem (3) contains four unknown variables P, Q, L_X and θ. It usually needs expensive computation to optimize over several variables simultaneously. Hence, we adopt a common alternating optimization strategy, i.e., minimizing w.r.t. the four variables one at a time. Given $L = P \cdot L_X \cdot Q$, updating θ is the classic problem which learns the parameters of a GMM via maximum likelihood estimation. This can be achieved by Expectation Maximization [20].

3.1. Optimizing L_X by Graph Cut

Given P and Q, optimizing L_X is the classic problem of labeling on a graph. Specifically, L_X can be solved via graph cut [18], imple-

mented as follows. Let $L_0 \in \mathbb{R}^{m\times n}$ be an all-ones matrix. Denote $R = P \cdot L_0 \cdot Q$ and let R_i be the i-th element of R. Given P and Q, the unknown region (to be segmented) is $T_U = \{i|R_i = 1\}$ and consist of several bounding boxes of the detected repetitive patterns. The known background region is $T_B = \{i|R_i = 0\}$. Let $\mathcal{E}$ denote the set of pairs of pixels which are located at the same rows and columns within different bounding boxes in T_U. The constraint $L = P \cdot L_X \cdot Q$ in problem (3) is implemented as a hard constraint on T_B and $\mathcal{E}$. That is, the labels on T_B are enforced to be 0 and each pair of pixels in $\mathcal{E}$ share the same label.

3.2. Optimizing P and Q by Dynamic Programming

Since P and Q have similar structures, the algorithm solving Q can be directly applied to P^T. Moreover, due to the special structure, Q can be optimized by DP. Its main idea is to decompose the original problem into a series of interrelated sub-problems and then solve the sub-problems recursively.

Optimizing Q can be viewed as placing a left-to-right sequence of identity matrices. Consequently, the placement of each identity matrix is "conditionally independent" of the others given the nearest identity matrix. Assuming that the rightmost column of the $(t\text{-}1)$-th identity matrix is at the S_t-th ($0 \le S_t \le w$) column of Q, the sub-problem, denoted as $\mathcal{P}(S_t)$, finds the optimal placement for the remaining $w - S_t$ columns. If the t-th identity matrix is placed and its n-th column is at the S_{t+1}-th column of Q, this placement results in a new sub-problem $\mathcal{P}(S_{t+1})$. Accordingly, we define U_t as the distance from S_t to S_{t+1}. If no identity matrix is placed, the placing process is finished and U_t is defined as 0. Thus, it follows:

$$S_{t+1} = \begin{cases} S_t + U_t, & n \le U_t \le w - S_t \\ w, & U_t = 0 \end{cases}. \tag{6}$$

Among these possible U_t, the optimal one can be found via enumeration, which is formulated as:

$$V_t(S_t) = \min_{U_t \in \{0, n:w-S_t\}} \{C_t(S_t, U_t) + V_{t+1}(S_{t+1})\}, \tag{7}$$

where $V_t(S_t)$ is the optimal cost of the sub-problem $\mathcal{P}(S_t)$, and $C_t(S_t, U_t)$ is the cost of U_t given S_t calculated according to (4) and (5). This enumerative process is recursively implemented, started from $S_t = w$ and finished when $S_t = 0$. Once this recursive process stops, the optimal Q^* will be reconstructed via back-tracking.

There are $O(w)$ sub-problems in the recursive process. For each sub-problem, $O(w)$ enumerations are evaluated, with each requiring $O(w \cdot h)$ additions. With pre-computation of the cost, the complexity optimizing Q is therefore $O(w^2)$.

3.3. Initial Detection and Convergency

A trivial initialization to the above alternating algorithm is to take only the user-specified bounding box as foreground and then learn GMMs. However, the foreground pixels may be too insufficient relative to the background pixels, leading to inaccurate GMMs and many false negatives. Hence, we adopt an initial detection strategy, implemented as follows. First, given the user-specified bounding box, we conduct normalized cross-correlation (NCC) based template matching on the grayscale image. Then, the detections whose HOG [21] has a NCC with that of the template lower than 0.8 are removed. Finally, some missing detections are remedied in accordance with the prior that the repetitions are horizontally and vertically aligned.

It can be shown that each step of the iterative algorithm minimizes the total energy. Hence, the energy decreases monotonically and the algorithm is guaranteed to converge at least to a local minima. It is straightforward to automatically terminate the iterations when the decreasing rate of the energy is smaller than a threshold (experimentally set as 10^{-2}). In our experiments, the algorithm typically terminates with two or three iterative steps.

Fig. 3. Four trials on a testing image with interactions on different positions. Given the green bounding box, the red curve shows the corresponding result obtained by our method.

4. EXPERIMENTS

We conducted a series of experiments to demonstrate the validity of our approach. The first experiment tests the stability of our method against the user interaction. In the second experiment, the proposed method is tested on 100 images from the Facade Database[1]. Finally, our method is compared with GraPes[2] [11]. Throughout the experiments, the results are measured via F-score, which is defined as:

$$F = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FN} + \text{FP}}, \tag{8}$$

where TP, FN and FP are the true positives, false negatives and false positives, respectively. The parameter λ is set as 10, since we found that $\lambda = 10$ is appropriate for most testing images. The number of Gaussian components of each GMM is set as 3 because generally the color distributions of facade images are relatively uncomplicated.

4.1. Stability against User Interaction

Since our approach needs a user-specified bounding box, the stability against user interaction is tested. For each window on the image with size 480×904 shown in Fig. 3, 100 trials are implemented. In each trial, the four borders of the bounding box deviate several pixels from

[1] http://www.kevinkaixu.net/k/projects/symbr.html
[2] http://vision.mas.ecp.fr/Personnel/teboul/grapes.php

Fig. 4. Representative results of our method. In each image, the green bounding box is user-specified and the red curve shows the segmentation result. Top left is a failure case of [13].

pre-defined baselines. The offsets are random integers in $[0, 15]$. Four trials and corresponding results are shown in Fig. 3. Among all the 800 trials on the image, the F-score is 0.973 ± 0.024 (mean$\pm$ standard deviation). Thus, our method is stable on the image.

4.2. Results on 100 images from Facade Database

We further test our approach on 100 images selected from Facade Database. For a kind of repetitions on each image, the bounding box of each repeated element has been manually given. Tested with these bounding boxes, the corresponding F-score is recorded. The final F-score of the image is calculated as the average of these trials. Among the images, 87% images achieve F-scores higher than 0.9, as shown in Fig. 5. It should be mentioned that an F-score higher than 0.9 means that all the repetitions are detected. In addition, our method produce higher F-score than template matching, and the iterative optimization significantly improves the results of initial detection. Some representative results of our method are shown in Fig. 4. As illustrated in Fig. 4, our method is robust to external occlusions, weakly changing illuminations and appearance variations.

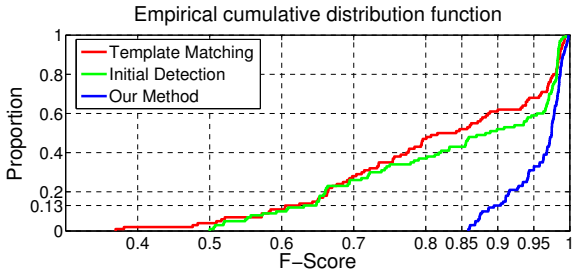

Fig. 5. The empirical cumulative distribution function of F-score on 100 images from Facade Database. Template matching and initial detection is discussed in section 3.3.

4.3. Comparative Results with GraPes

Our method is compared with GraPes [11], which implemented shape grammar parsing via reinforcement learning. Note that our method only needs the user-specified bounding box, while GraPes needs additional strokes on the background. Both approaches are ran 10 times on 3 testing images from Ecole Centrale Paris Facades Database[3]. The visual results are shown in Fig. 6 and the quantitative evaluations are illustrated in Fig. 7. As shown in the figures, our method produces more accurate results. Moreover, our labeling results in 10 trials are all the same. In contrast, there are large variances among the results of GraPes.

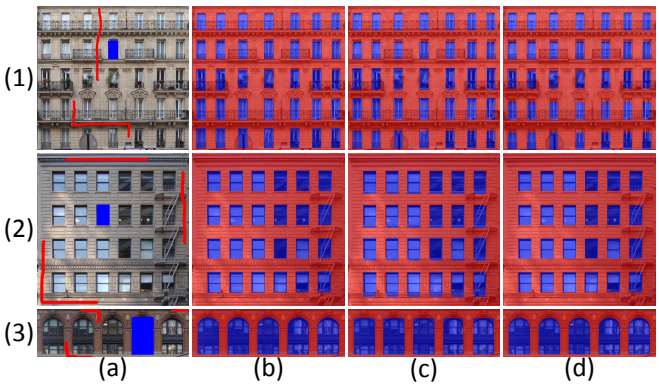

Fig. 6. The comparative results with GraPes. (a) The testing image with user interaction. Note that our method only needs the blue bounding box. (b) The manually labeled ground truth. (c) Among the 10 trails, the best result of GraPes. (d) Our segmentation result.

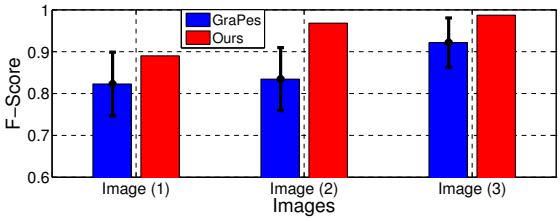

Fig. 7. The bar shows the mean F-score of both method in the 10 trials on the 3 images shown in Fig. 6. Our method produces the same results in the 10 trials, while the error bar shows the standard deviation of F-scores among the results of GraPes.

5. CONCLUSION

This paper implements a pixel-wise segmentation for repetition extraction on facade images. The repetitions are modeled by GTS and is explicitly formulated as the product of multiplying a repeated element by two block matrices interpreting the symmetries. This formulation is embedded into the graph cut based image segmentation framework. Therefore, repetition detection and segmentation are jointly formulated as a GTS constrained energy minimization problem. After the initial detection, the problem is alternatively solved by dynamic programming and graph cut. The experimental results demonstrate the effectiveness of our method.

[3] http://vision.mas.ecp.fr/Personnel/teboul/data.php

6. REFERENCES

[1] A. Martinovic, M. Mathias, J. Weissenberg, and L. V. Gool, "A three-layered approach to facade parsing," in *European Conference on Computer Vision*, 2012, pp. 416–429.

[2] D. Dai, H. Riemenschneider, G. Schmitt, and L. V. Gool, "Example-based facade texture synthesis," in *IEEE International Conference on Computer Vision*, 2013, pp. 1065–1072.

[3] S. AlHalawani, Y.-L. Yang, H. Liu, and N. J. Mitra, "Interactive facades analysis and synthesis of semi-regular facades," *Computer Graphics Forum*, vol. 32, no. 2, pp. 215–224, 2013.

[4] H. Zhang, K. Xu, W. Jiang, J. Lin, D. Cohen-Or, and B. Chen, "Layered analysis of irregular facades via symmetry maximization," *ACM Transactions on Graphics*, vol. 32, no. 4, pp. 121:1–121:10, 2013.

[5] P. Müller, G. Zeng, P. Wonka, and L. V. Gool, "Image-based procedural modeling of facades," *ACM Transactions on Graphics*, vol. 26, no. 3, pp. 85:1–85:9, 2007.

[6] Y. Li, Q. Zheng, A. Sharf, D. Cohen-Or, B. Chen, and N. J. Mitra, "2d-3d fusion for layer decomposition of urban facades," in *IEEE International Conference on Computer Vision*, 2011, pp. 882–889.

[7] M. Park, K. Brocklehurst, R. T. Collins, and Y. Liu, "Deformed lattice detection in real-world images using mean-shift belief propagation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 10, pp. 1804–1816, 2009.

[8] C. Wu, J.-M. Frahm, and M. Pollefeys, "Detecting large repetitive structures with salient boundaries," in *European Conference on Computer Vision*, 2010, pp. 142–155.

[9] J. Liu, E. Psarakis, and I. Stamos, "Automatic kronecker product model based detection of repeated patterns in 2d urban images," in *IEEE International Conference on Computer Vision*, 2013, pp. 401–408.

[10] O. Teboul, I. Kokkinos, L. Simon, P. Koutsourakis, and N. Paragios, "Segmentation of building facades using procedural shape priors," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 3105–3112.

[11] O. Teboul, I. Kokkinos, L. Simon, P. Koutsourakis, and N. Paragios, "Shape grammar parsing via reinforcement learning," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 2273–2280.

[12] H. Riemenschneider, U. Krispel, W. Thaller, M. Donoser, S. Havemann, D. W. Fellner, and H. Bischof, "Irregular lattices for complex shape grammar facade parsing," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1640–1647.

[13] P. Zhao and L. Quan, "Translation symmetry detection in a fronto-parallel view," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 1009–1016.

[14] P. Zhao, L. Yang, H. Zhang, and L. Quan, "Per-pixel translational symmetry detection, optimization, and segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 526–533.

[15] 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 Centres*, 2004.

[16] C. Yang, T. Han, L. Quan, and C.-L. Tai, "Parsing façade with rank-one approximation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1720–1727.

[17] T. Han, C. Liu, C.-L. Tai, and L. Quan, "Quasi-regular facade structure extraction," in *Asian Conference on Computer Vision*, 2012, pp. 552–564.

[18] Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images," in *IEEE International Conference on Computer Vision*, 2001, pp. 105–112.

[19] C. Rother, V. Kolmogorov, and A. Blake, ""grabcut": interactive foreground extraction using iterated graph cuts," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 309–314, 2004.

[20] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[21] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 886–893.