Variation-Factored Encoding of Facade Images

Suhib Alsisan
KAUST

Niloy J. Mitra
KAUST/UCL

Abstract

Urban facades contain large-scale repetitions in the form of windows, doors, etc. Such elements often are in different configurations (e.g., open or closed) obscuring their regular arrangements to any direct low-level pixel matching based repetition detection. We propose a variation-factored representation for facade images by progressively favoring larger repeated structures while allowing relabeling using candidate element types. We formulate the problem as a Markov Random Field (MRF) based optimization, and evaluate the algorithm on a large number of benchmark facade images. Such a facade encoding is very compact and can be used for rapid generation of realistic 3D models with variations suitable for online map viewers or mobile navigation aids.

1. Introduction

Advances in acquisition techniques and computer vision algorithms have resulted in rapid growth of large collections of facade images [SSS06] and 3D scans [BFP*04]. Such data, however, come in low-level representations (e.g., pixels or points) providing little insight into the actual complexity of the corresponding facades (e.g., in Google StreetView, Microsoft Bing Maps).

Researchers have focused on extracting high level information from facade images using a range of priors like procedural grammar [TSKP10], manual annotations [MZWG07], reinforcement learning [TKS*11], repetition based scan consolidation [LZS*11], etc. They, however, do not characterize the inherent information content of the input facades. We propose to jointly analyze the basic facade elements (e.g., windows, doors) and use their cross-correlation to reveal the dominant modes of geometric variations. This not only leads to compact storage, but also provides a *variation-factored encoding* of the facades allowing rapid, realistic, and efficient creation of procedural facades in the space of extracted element variations.

Facade images display rich pixel-level variations depending on time and location of capture. Such variations are mainly due to: (i) *intrinsic* effects like valid part-level geometric alterations of the physical building like opening or closing of windows, rising or closing of blinds, etc., (ii) *extrinsic* effects like presence of outlier elements, varying illumination, etc. As humans, we routinely ignore these variations when observing facades, motivating a variation-factored facade representation. The primary goal of this work is to factor out intrinsic variations in facade images. For example, when looking at typical urban facades we remember the distribution of states of the window elements, rather than their individual spatial states. Hence, we extract and encode only the key configurations of the individual elements, while storing the statistics of their spatial distributions hierarchically. This leads to compact representations ($4-10$x in our tests). Further, the encoding naturally allows easy generation of multiple plausible facade variations starting from a single facade image (see Figure 1).

Performing such a factorization from a single image without further information is ill-posed. However, we observe that most facades contain large scale repetitions in certain (hidden) canonical configurations (see [TSKP10] and references therein). For example, in Figure 2 although the building does not exhibit any obvious image-level repetition, we as humans detect a 9×2 repetition pattern implicitly changing the windows to a common state, say open or closed. This

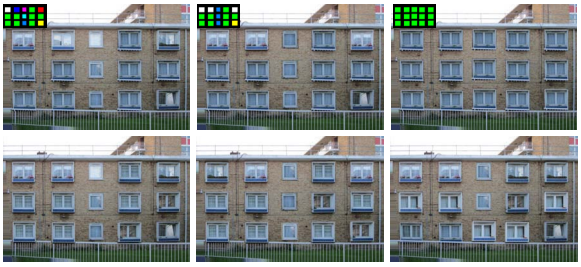

Figure 1: *Given a facade image (top-left) we use a novel MRF formulation to create different hierarchy levels of variation-factored encodings (top-middle/right) – insets show types of window elements in the respective levels. (Bottom row) The representation can then be used to generate plausible variations of the input facade.*

DOI: 10.2312/conf/EG2012/short/037-040

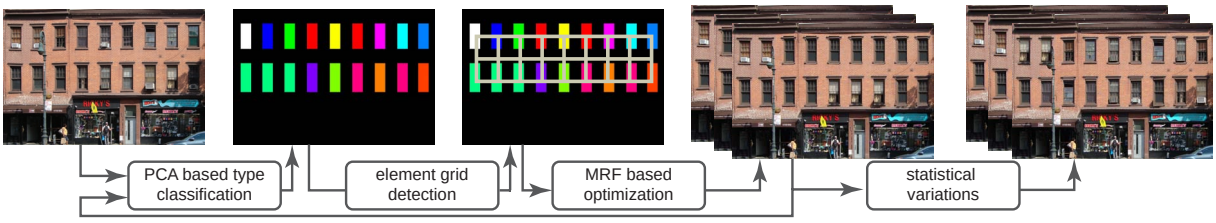

Figure 2: *Algorithm pipeline: Starting from a single facade image (left), we classify the detected elements into types, identify an initial repetition grid, and then use a MRF-based formulation to extract a variation-factored hierarchy. The hierarchical representation is then used to create novel facade variations (right).*

is not surprising since repetitions often relate to both aesthetic appeal and manufacturing convenience. As buildings are inhabited, local geometry gets modified non-uniformly, obscuring the original repetitions and patterns. Our goal is to simultaneously segment a facade into basic elements, extract key states of the elements, and bring them to a canonical configuration to restore the original regularity of the facades.

We introduce a novel Markov Random Field (MRF) formulation to automatically extract a hierarchy of cost-optimized representations progressively favoring more regular configurations, while suppressing variations (see Figure 1:top-right). We evaluate our algorithm on a large number of publicly available benchmark facade images with different extent of variations (see supplementary material). We use the extracted factored representation to enable novel possibilities: Starting even from a single facade image, we can (i) obtain a simple statistical model encoding possible variation states of the facade, and (ii) procedurally create plausible facade variations while breaking boring and unnatural regular repetition patterns (see Figure 4).

2. Related Work

Image-based modeling. Debevec et al. [DTM96] in their seminal paper propose an interactive image-based modeling method that exploits characteristics of architectural objects coupling an image-based stereo algorithm with manually specified 3D model constraints. Subsequently, automatic reconstruction of urban facades has been developed from unorganized photo collections (see [SSS06, XFT*08] and references therein) using photogrammetric reconstruction and image-based modeling techniques. Such systems produce massive collections of low-level unorganized textured points, which are not suited for low-memory footprint navigation or mobile interactions (e.g., Google Streetview).

Procedural modeling. Wonka et al. [WWSR03] use split grammars and an attribute matching system to synthesize buildings with varying styles. Later, Müller et al. [MZWG07] explore auto-correlation based analysis of rectified images combined with shape grammars towards urban reconstruction. They propose an interesting mix of user interaction and image analysis for rule-based procedural modeling. These methods, being interactive, are difficult to use for large scale modeling. Further, the methods do not analyze the allowable variations across similar elements, and hence the outputs often contain noticeable repetition patterns due to procedural generation.

Consolidated model synthesis. Multiple data sources (e.g., photographs, LiDAR scans, aerial images, GIS data) have been combined to improve the quality of 3D models [FJZ05, LZS*11]. Directly working with incomplete LiDAR scans, Zheng et al. [ZSW*10] use model scale repetitions to create a consolidated point cloud. Although the resultant point clouds have high resolution, the algorithms rely on multiple sources, have moderate to high memory foot-prints, and are not directly suited for creating realistic model variations.

Facade annotations. Our work is inspired by recent efforts of Teboul et al. [TSKP10] who perform supervised learning using shape grammar priors to create perceptual interpretation of building facades using random walks on the learned models. The method has been extended using recursive binary split grammar and reinforcement learning [TKS*11] to parse facades into element level masks (e.g., windows, doors) using training data. Neither of these efforts, however, models intrinsic variations across elements. We make use of such a classifier to initially parse the facade images and use our framework to extract the variations.

3. Algorithm

We now describe the proposed algorithm (see Figure 1).

Pre-processing. We first automatically rectify the input using extracted vanishing lines [WFP10] and then use publicly available Grapes package [TKS*11] to obtain a rough mask indicating possible element positions and sizes. We then group the elements into classes based on their classification types and their sizes. For example, all the detected windows of comparable size (based on respective mask boundary) are grouped together (see also Section 4).

Part clustering and grouping. We now look at the intensity variations of the extracted elements based on the masks to identify key element types. For noise robustness, we only retain the dominant signals using principal component analysis (PCA). Specifically, we map each of the (colored) element patches $\{P_1, P_2, \ldots\}$ of size say $m \times n$ to vectors in

$\mathbb{R}^{3mn}$ and extract their PCA modes. We retain only the top d-principal components to obtain projected patches. Next, we classify these patches into types by using mean-shift clustering (kernel width set to 0.1 of the size of element) on the d-coordinates in the PCA space ($d = 7$ in our tests). For each such cluster, as its *representative*, we take the projected patch closest (in the d-dimensional PCA space) to the cluster centroid. At the end of the patch grouping step, each part P_i has an associated type $T(P_i) \in \mathcal{T}$, where $\mathcal{T}$ denotes the set of possible types (e.g., see Figure 2). In the future, we plan to use normalization and high-level edge features to better handle noise from outliers and illumination variations.

Initial regularity grid detection. Next, we generate a *structure matrix* $\mathbf{S}^0$ to describe the facade, with each entry indicating the element type ($\in \mathcal{T}$) for patch P_i in the corresponding image location. First, the center of each patch P_i is computed. The centers are then used to initialize a 2-parameter repetition grid. If desired, we can directly operate on the rectified image to refine the estimates for the generators of the repetition grid [PMW*08]. Each patch P_i is then mapped to the nearest element in the refined structure matrix $\mathbf{S}^0$.

Optimization. Finally, we create the hierarchical encoding of the input facade. Specifically, for the i-th hierarchy level we search for a matrix $\mathbf{S}^i$ with higher repetition structure by performing relabeling of element types to the previous structure matrix $\mathbf{S}^{i-1}$. We formulate the relabeling as a Markov Random Field (MRF) minimization as follows:

$$\min_{\mathbf{x} \in |\mathcal{T}|^{mn}} E(\mathbf{x}) := E_{data}(\mathbf{x}^{i-1}, \mathbf{x}) + \alpha E_{reg}(\mathbf{x}) \qquad (1)$$

where, $\mathbf{x}^{i-1}, \mathbf{x}$ denote the corresponding structure matrices $\mathbf{S}^{i-1}, \mathbf{S}$ concatenated into vectors. The weight factor α determines the relative contributions between the data and the regularity terms, as described next.

Data term (E_{data}): This term penalizes relabeling of distinct element types. We construct a cost matrix $\mathbf{C}$ where $\mathbf{C}(a,b)$ measures the cost of relabeling a part of a-type by a part of b-type. Given two element types t_a and t_b, we first rescale the bigger-sized element representative to the smaller element representative, and then take their sum of squared differences (SSD) as $\mathbf{C}(a,b)$. Note that $\mathbf{C}$ is symmetric with diagonal entries as zeros. In order to favor replacements between element types with small proximity scores, we use a non-linear function f to adjust the data term, specifically, $f : x \rightarrow x^3$. We normalize the entries in $\mathbf{C}$ by its maximum element. Finally, the data term is computed as the accumulation cost as: $E_{data}(\mathbf{x}^{i-1}, \mathbf{x}) := \sum_k \mathbf{C}(\mathbf{x}_k^{i-1}, \mathbf{x}_k)$. In order to replace a part of a-type with b-type, we use the representative element for b-type (as computed earlier).

Regularity term (E_{reg}): This term measures the regularity of any given structure matrix and is critical for the success of

$$E_{reg}(\mathbf{x}) := |unique(\mathbf{x})| + repeat(\mathbf{x}) + \sum_{r \in subrect(\mathbf{x})} 1/size(r).$$

The first term represents the number of different types in the structure matrix $\mathbf{x}$, while the second term measures the extent of repetitions and is computed as the summation of the reciprocals of the number of times each type is repeated in $\mathbf{x}$. The last term is the summation of the reciprocals of the sizes of all sub-rectangles in the $\mathbf{x}$ and gives preference to large 2-parameter repetition patterns (as commonly found on facades). We use dynamic programming to identify the biggest such sub-rectangle consisting of similar type elements in $\mathbf{x}$, and then the next biggest sub-rectangle that does not overlap with any previous sub-rectangles, and so on (see Figure 3).

Figure 3: *(Left) Input facade image, (middle) third level of variation-factored hierarchy with corresponding non-overlapping regularity sub-rectangles, (right) corresponding structure and cost matrices. In this level, unique($\mathbf{x}$)=5, repeat($\mathbf{x}$) = 1/6+1/5+1+1/3+1/3.*

In order to solve for hierarchy levels with increasing preference for regular structures, we minimize energy given in Equation 1 with increasing values of α using the iterated conditional modes (ICM) MRF optimization [Bes86] stopping when all the facade elements have same labeling (typically $5-10$ hierarchy levels in our test images).

Statistical variations. We now use the hierarchy of abstraction levels to generate statistical variations of the input facade. First, we estimate the probability that a part type t_i in the input image is relabeled as type t_j in a given abstraction level. We then sample from these estimates to generate statistical variations of the input facade starting from the corresponding abstraction level. Essentially, this amounts to permuting elements among those elements of the same type in any particular hierarchy level (see Figure 4:bottom). In our experiments, we observed that random flipping of part types in the original input image leads to unsatisfactory results.

4. Evaluation

We tested our algorithm on a collection of publicly available facade images (see supplementary materials). Our test images produced $5-10$ hierarchy levels, taking 1-2 seconds to compute using our unoptimized Matlab implementation. Typically our proposed encodings required $4-10$x less space to store all the hierarchy levels. Specifically, instead of the original image, we only store the d-dominant PCA axes and PCA coordinates for the different element types along with their distribution for each level. For example, the number of hierarchy levels and the total compression ratios for the different examples are as follows: Figure 1: 6x and 5.8x; Figure 2: 9x and 8.1x; Figure 3: 7x and 10.4x; Figure 4: 8x

Figure 4: *Starting from a single facade sample (top-left), we use our variation-factored representation to create procedural facade faces with variations (bottom-row).*

and 4.0x; Figure 5: 5x and 10.0x, respectively. We generated a range of variations using the encoded representations.

Applications. In a small user study (with 5 users in our laboratory), we observed that the users rarely spotted the statistically created variations from the original facade images in their first glance. This demonstrates the benefit of having a variation-factored representation (with marginal memory overhead) in quickly creating plausible facade variations as commonly required in games, mobile browsing, online mapping applications (see Figures 4 and 5), etc. Note that our encoding allows easy generation of novel facades with different number of repetitions while retaining observed element-level variations in the original facade.

Limitations. Our regularity term is designed to measure regularity in rectilinear structure matrices. Although most facades indeed have such structures, the algorithm might produce suboptimal abstraction levels for those facades that do not conform to a rectilinear structure. In the future we want to consider alternate regularity terms to capture patterns in facades of general structure. Further, since our method makes use of [TKS*11] to acquire rough masks of the facade components in the preprocessing stage, poor masks could lead to unsatisfactory results. Another direction for future work is investigating the possibility of using the generated abstraction levels to enhance the input masks.

5. Conclusion

We presented variation-factored facade images, a novel representation that specifically stores distributions of variations across elements of the input facade. We presented a MRF optimization to automatically generate such encodings, and used them to create novel facade scenes both as images and also as low-memory footprint 3D street-view models. In the future, we plan similarly analyze 3D LiDAR scans leading to high-level variation-factored facade models.

Figure 5: *Starting from facade images, we compute their variation-factored representations, which can then be used for fast and efficient creation of 3D models with variations.*

References

[Bes86] BESAG J.: On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society 48* (1986), 259–302. 3

[BFP*04] BOSTROM G., FIOCCO M., PUIG D., ROSSINI A., GONCALVES J., SEQUEIRA V.: Acquisition, modelling and rendering of very large urban environments. In *3DPVT04* (2004), pp. 191–198. 1

[DTM96] DEBEVEC P. E., TAYLOR C. J., MALIK J.: Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach. In *SIGGRAPH* (1996), pp. 11–20. 2

[FJZ05] FRUEH C., JAIN S., ZAKHOR A.: Data processing algorithms for generating texture 3D building facade meshes from laser scans and camera images. *IJCV 61* (2005), 159–184. 2

[LZS*11] LI Y., ZHENG Q., SHARF A., COHEN-OR D., CHEN B., MITRA N. J.: 2d-3d fusion for layer decomposition of urban facades. In *ICCV* (Barcelona, Spain, November 2011). 1, 2

[MZWG07] MÜLLER P., ZENG G., WONKA P., GOOL. L. V.: Image-based procedural modeling of facades. In *SIGGRAPH* (2007). 1, 2

[PMW*08] PAULY M., MITRA N. J., WALLNER J., POTTMANN H., GUIBAS L.: Discovering structural regularity in 3D geometry. *ACM TOG 27*, 3 (2008), 43:1–43:11. 3

[SSS06] SNAVELY N., SEITZ S. M., SZELISKI R.: Photo tourism: Exploring photo collections in 3d. In *SIGGRAPH Conference Proceedings* (2006), ACM Press, pp. 835–846. 1, 2

[TKS*11] TEBOUL O., KOKKINOS I., SIMON L., KOUTSOURAKIS P., PARAGIOS N.: Shape grammar parsing via reinforcement learning. In *CVPR* (2011). 1, 2, 4

[TSKP10] TEBOUL O., SIMON L., KOUTSOURAKIS P., PARAGIOS N.: Segmentation of building facades using procedural shape prior. In *CVPR* (2010). 1, 2

[WFP10] WU C., FRAHM J.-M., POLLEFEYS M.: Detecting large repetitive structures with salient boundaries. In *ECCV* (2010). 2

[WWSR03] WONKA P., WIMMER M., SILLION F., RIBARSKY W.: Instant architecture. *ACM TOG 22* (2003), 669–677. 2

[XFT*08] XIAO J., FANG T., TAN P., ZHAO P., OFEK E., QUAN L.: Image-based facade modeling. *ACM TOG (SIGGRAPH Asia) 27* (2008), 161:1–161:10. 2

[ZSW*10] ZHENG Q., SHARF A., WAN G., LI Y., MITRA N. J., COHEN-OR D., CHEN B.: Non-local scan consolidation for 3D urban scenes. *ACM TOG 29*, 4 (2010), 94:1–94:9. 2