Building Reconstruction using Manhattan-World Grammars

Carlos A. Vanegas Daniel G. Aliaga Bedřich Beneš

Purdue University

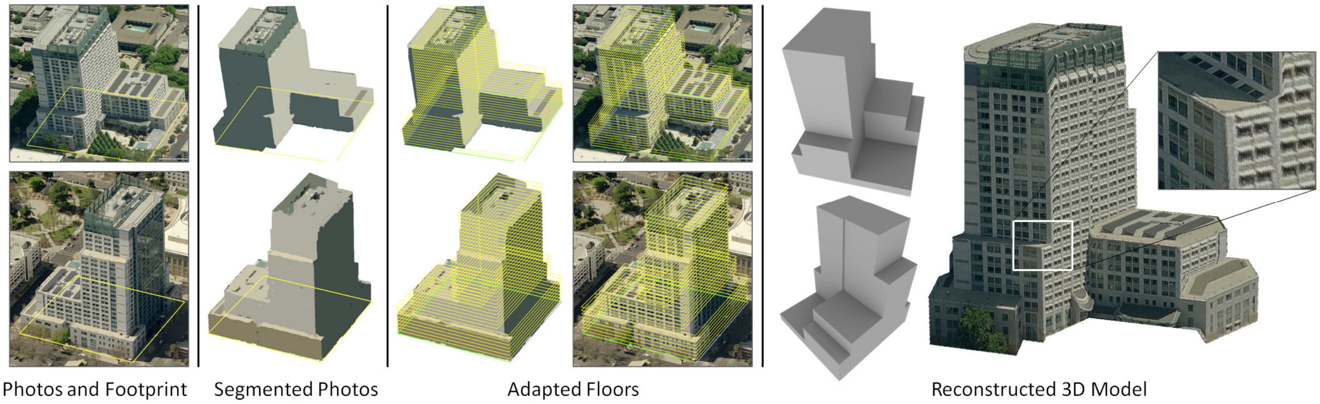

Figure 1. System Pipeline. The input to our system consists of one or more calibrated aerial images of a Manhattan-world building. After color segmentation and background/windows removal, our grammar-based algorithm adapts the geometry of the building that produces the façade orientation changes observed in the photos. The input photos are projected as textures onto the reconstructed model. The result is an automatically-generated complete, closed 3D model of the observed building.

Abstract

We present a passive computer vision method that exploits existing mapping and navigation databases in order to automatically create 3D building models. Our method defines a grammar for representing changes in building geometry that approximately follow the Manhattan-world assumption which states there is a predominance of three mutually orthogonal directions in the scene. By using multiple calibrated aerial images, we extend previous Manhattan-world methods to robustly produce a single, coherent, complete geometric model of a building with partial textures. Our method uses an optimization to discover a 3D building geometry that produces the same set of façade orientation changes observed in the captured images. We have applied our method to several real-world buildings and have analyzed our approach using synthetic buildings.

1. Introduction

Reconstruction of buildings and urban areas is crucial to a variety of applications including city planning, simulation, and training, and real-time uses such as gaming and virtual reality. Recently, aerial-view, ground-level, and oblique-angle images of urban areas have become available through Internet-based services such as Google Maps, Bing Maps, and Yahoo Maps that provide public access to geographic information system (GIS) style data. Providing an automatic mechanism to add the 3D geometry of the observed buildings would provide significant additional information for navigation, driving directions, and other related uses. We focus on providing a passive method that exploits the existing mapping and navigation databases to automatically create 3D building models.

To date models of building geometry can be generated by one of several mechanisms. Computer-assisted photogrammetric modeling methods require significant manual effort and time. Fully automatic approaches use laser-scans or LIDAR data, combined with aerial imagery or ground-level images (e.g., [4], [8], [21], [27]). However, most of the previous work suffers from one or all of low-resolution sampling, robustness, and missing surfaces. One way to improve quality or automation is to incorporate assumptions about the buildings. One such assumption is that buildings often contain planar faces. Recently, similar methods have focused on an important class of architectural structures obeying a so called *Manhattan-world* (MW) assumption [5]. It states that there is a predominance of a triple of mutually orthogonal directions in the scene. This assumption has been used to provide 3D reconstruction methods for building interiors and for more general architectural scenes observed with stereo pairs [9]. While these methods produce improved results, they still have missing surfaces and do not produce complete buildings.

Our key observation is that a Manhattan-world building can be represented by a parametric grammar describing a compact set of transitions between consecutive floors. The grammar encodes the transition types and the parameters the exact shape. Moreover, using a grammar facilitates the generation of a complete model for which hidden faces can be inferred. We seek to significantly extend previous work by using grammar-based techniques for modeling entire buildings observed by multiple oblique-angle aerial images. By using multiple views, we extend previous MW methods to yield coherent and complete building models.

Our approach reconstructs real-world buildings using a MW building grammar (Figure 1) and using a progressive building reconstruction method based on one or more calibrated oblique-angle aerial views. Our approach has two assumptions: i) the observed building can be represented by a sequence of building floors, each floor is composed of a set of connected faces, and each face is parallel to a MW direction, and ii) each Manhattan direction of a building floor is colored differently within each image. Our method uses a provided initial 3D building envelope (e.g., an extruded bounding box of the building footprint extracted from GIS data) that is further refined. The initial model is divided into a sequence of floors (i.e., expanded rules of the grammar) and each floor into an array of faces (i.e., terminal symbols of the grammar). Then, our approach defines and uses a rewriting rule which performs transitions to the floors of the initial 3D model in order to produce a new building model matching the one in the aerial views.

Our assumptions are based on the intuition that each of the three possible façade orientations of a MW building typically has a different (average) pixel value. Building walls are usually made of the same material and the input images are captured on days with few clouds such that the sun does not shine at the same angle to two or more façade orientations. Hence, by using color segmentation (e.g., mean shift segmentation [3]), each façade orientation has a different pixel value. As opposed to photometric stereo (e.g., [1]) or normal clustering (e.g., [12]), the absolute value of the pixel color is not relevant, as long as the pixel value within each image is different per façade orientation on the same floor. Further, pixel-level segmentation errors are overcome by the conditions imposed by our grammar – thus imprecision in albedo and segmentation is not critical which yields a significant advantage over a dense 3D reconstruction method. Moreover, the color for the same façade orientation can be different on two different images. This allows multiple views of a building to be taken under different illumination.

Our reconstruction method uses an optimization to discover a 3D building geometry that produces the same set of façade orientation changes and at the same location as seen by the oblique-angle aerial images. Since the possible façade orientations are known, their pixel value is not needed to reconstruct their geometrical orientation. To perform the transitions from the initial 3D model to the improved building model, we define a generalized rewriting rule that captures all plausible transitions from one floor to a next floor. Finding the parameters of the rule is expressed as an optimization that searches for the changes between two successive floors that reduces the value of an error metric between the observed changes and the changes produced by the geometric model. Altogether, our method simplifies the process of building shape detection to the sequential detection of floor-to-floor façade changes. Our approach has been applied to several real and synthetic buildings using only 2 to 4 images.

2. Previous Work

2.1. Building Acquisition Methods

Laser-scanning and/or LIDAR obtain dense 3D point clouds but suffer from robustness, noise, and incomplete models. Typical methods fit building envelopes to vertically extended footprints and use one of several possible roof geometries (e.g., [14], [25], [22]).

Other methods use ground-level video through a city with GPS (e.g., [18],[21]) and/or laser-scanning equipment (e.g., [8]). These approaches provide more visual detail but less overall coverage and do not capture tall buildings well.

Registering ground-level images to aerial images is another option. For example, Wang et al. [27] merges aerial and ground-based images to produce building models, but require user assistance and only produce buildings with vertical and planar walls. Other works attempt to register uncalibrated photographs to 3D laser scans (e.g., [16]).

Tools have been presented for manually reconstructing a building from photographs (e.g., [6], [11]). By enforcing epipolar, edge, and attachment constraints, the modeling process can be simplified but is still manual.

2.2. Manhattan World

The Manhattan-world assumption was first defined by Coughlan and Yuille [5]. Lee et al. [15] generate plausible views of the interior of rooms from a single view and Furukawa et al. [10] automatically construct models of building interiors. The same authors [9] recently presented a novel multi-view stereo algorithm exploiting the MW assumption. While their method has also been applied to outdoor buildings, it does not produce complete building models. In contrast, our method is able to infer a reasonable complete model even given partial occlusion.

2.3. Grammar-based Methods

Although grammars are traditionally used for generative modeling [17], in computer graphics they have enabled the design of complex architectural models [19] and in computer vision they have assisted in producing detailed models of façades semi-automatically (e.g., [1], [13], [20]). Aliaga et al. [1] also used grammars to create buildings from photographs, but their method provides an interactive tool with little automation. As opposed to reconstructions of façades and buildings from a small set of parameterized blocks [7], we use a grammar-based approach to automatically infer complete building models.

3. Manhattan Building Grammar

Our *Manhattan building grammar* exploits the coherency present amongst the floors of a building and provides a compact representation of the outer shape of a building. Rather than using arbitrary connected polygons, the structure imposed by our grammar is beneficial to ensure a plausible, coherent, and complete building is produced. Starting with an initial shape for the ground-level floor (e.g., a bounding box of the building footprint) and a constant floor height value, each successive floor up the building is constructed by applying a set of transitions to the previous floor. All floors use a constant and typical floor height value. However, knowing this value accurately is not necessary since floors are only an intermediate tool for reconstructing a building -- they need not match one-to-one with the actual floors. A floor is represented by a string of parameterized letters and a transition from one floor to the next is represented by an application of a rewriting rule of a linear grammar. The parameters of a rewriting rule encode the geometric sizes while its syntactic composition encodes the type of structural change.

After a careful observation and analysis of many different MW buildings, we concluded most floor-to-floor transitions can be encoded into a single parameterized rewriting rule. Transitions can be combined in multiple ways resulting in complex structures. Further, by the use of constraints we ensure that after the application of a rule the building remains plausible and coherent.

3.1. Building Representation

A MW building is represented as a sequence of floors $S = \{S_1, S_2, \ldots, S_N\}$ and each floor is formed by a sequence of planar quadrilateral faces $F_i = \{f_{i1}, f_{i2}, \ldots, f_{iM_i}\}$ where $i \in [1, N]$ and M_i is the number of quadrilateral faces for floor i. We assume the faces (i.e., building walls) are perpendicular to the ground plane and are aligned with a MW direction. For brevity, we treat a floor i as a polyline P_i with the angle between successive line segments being only ± 90 or 180 degrees and we use $f(l)$ to represent a generic segment of such a polyline of length l. We also use the notation $P_i(t)$ to refer to the point on the polyline at parametric position $t \in [0,1]$. Roof geometry is addressed separately and is not explicitly encoded in our grammar.

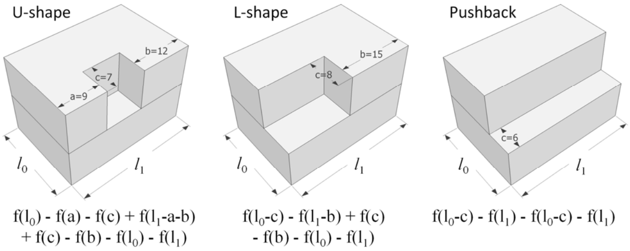
U-shape
b=12
c=7
a=9
l_0
l_1
f(l$_0$) - f(a) - f(c) + f(l$_1$-a-b)
+ f(c) - f(b) - f(l$_0$) - f(l$_1$)
L-shape
b=15
c=8
l_0
l_1
f(l$_0$-c) - f(l$_1$-b) + f(c)
- f(b) - f(l$_0$) - f(l$_1$)
Pushback
c=6
l_0
l_1
f(l$_0$-c) - f(l$_1$) - f(l$_0$-c) - f(l$_1$)

Figure 2. Generalized Rewriting Rule. The representative strings and geometries generated by GRR are shown for the U-shape (left), L-shape (middle) and pushback (right) cases.

Our string representation of a building uses the alphabet

$$A = \{f(l), +, -\} \quad (1),$$

where the letter f is an instance of the aforementioned parameterized floor segment and + and – are operations for changing the next segment's orientation by 90 degrees to the right or to the left, respectively. To convert a string into floor geometry, we use the turtle graphics formulation. This formulation sequentially reads the parameterized letters of a string and interprets them either as a geometric element or as a transformation. We assume the turtle is initially located in a corner of a floor and heading in the positive x axis as seen from above. For example, a rectangular floor (Figure 2, bottom) is represented by $f(l_0) - f(l_1) - f(l_0) - f(l_1)$.

3.2. Grammar and Rewriting System

Our rewriting system is defined by $< A, R, \omega >$ where A is the aforementioned alphabet, R is a single parameterized rewriting rule, and ω is the starting symbol (i.e., initial shape for ground-level floor). The string for a new floor shape is based on the previous floor's shape except for the substring that corresponds to the change. This change is efficiently captured by a rewrite rule that replaces a letter with a sequence of new letters. Multiple rule applications enable a variety of building styles and complexities to be represented.

We observed that the shape change from one floor to the next can be achieved by one or more transitions belonging to one of three types: i) L-shape, ii) U-shape, or iii) push-back. We can represent all of these transitions with one *generalized rewrite rule* (GRR) defined as

$$f(l) \rightarrow f(a) - f(c) + f(l - a - b) + f(c) - f(b) \quad (2),$$

where l is the length of the original segment, and the lengths a, b, and c are parameters of the rule such that $a + b < l$ and $c \geq 0$. For $a \neq 0$, $b \neq 0$, and $c \neq 0$ it corresponds to the U-shape transition. For $a = 0$ or $b = 0$ it describes a left or right L-shape transition, respectively. For $a = b = 0$ and $c > 0$ it represents a push-back transition. Figure 2 shows several example transitions from a rectangular floor to one of the aforementioned shapes.

An application of GRR on a segment f_{ij} generally affects the segment f_{ij} itself as well as the preceding segment $f_{i(j-1)}$ and succeeding segment $f_{i(j+1)}$ in the linear encoding of the floor. In particular, an L-shape transition will either affect the preceding segment ($a = 0$) or the succeeding segment ($b = 0$). A pushback transition will affect both the preceding and succeeding segment. A U-shape only affects the actual segment being rewritten. This could be represented by a context-sensitive rule; however, we would not be able to represent all possible cases as a single GRR and more rules would be necessary.

3.3. Building Constraints

To ensure a plausible structure, we enforce several intra- and inter-floor constraints during application of our GRR.

- *Closed-Floor Constraint.* We assume the polyline corresponding to each floor is closed. Assuming the initial floor shape is closed, this is implicitly accomplished by definition of the GRR.
- *Non-Intersecting Floor Constraint.* The polyline that describes a floor must not intersect with itself. This can be represented by ensuring $f_{ij} \cap f_{ik} = \emptyset$ for all $i \in [1, N]$, $j \in [1, M_i]$, $k \in [1, M_i]$, and $j \neq k$.
- *Containment Constraint.* We assume buildings usually "converge" from bottom to top (i.e., the top cross section is of the same size or smaller than the bottom cross section). Thus, we enforce $P_{\text{i}+1} \subseteq P_i$.
- *Intra-Floor Change Constraint.* We limit transitions to those that significantly alter the shape of a floor. Hence, we desire $|f_{ij}| \geq \epsilon_1$, where ϵ_1 is a small number, and only apply a GRR if $\{a, b, c\} \geq \epsilon_1$.
- *Inter-Floor Change Constraint.* We further limit transitions to those that generate a significant change between consecutive floors. This restriction reduces the sensitivity to noise in the input data. To enforce this constraint, we use a set of equally-spaced parametric positions $t_u \in [0,1]$, where $u \in [1, U]$. We only allow changes to floor $i + 1$ that satisfy

  $$\sum_{u \in [1,\text{U}]} \| P_{i+1}(t_u) - P_i(t_u) \| > \epsilon_2 \qquad (3)$$

  where threshold ϵ_2 is used to consider floors different.

4. Building Reconstruction

Our building reconstruction method consists of modifying each floor string so as to improve a measure of consistency between the building model and the captured images. Starting at the ground floor, our method determines the consistency between the currently estimated floor geometry and the building in the aerial images. When an inconsistency occurs, our method computes the parameter values for one or more applications of our GRR.

4.1. Geometry and Image Signal Functions

For each floor, our method computes two impulse signal functions, $G_s(t)$ and $I_s(t)$, parameterized by $t \in [0,1]$ on the floor contour and where an impulse represents the appearance of an event resulting from the building's shape (Figure 3). The signal G_s contains geometric events defined at the "turns" in the floor contour of the building model. The signal I_s contains photometric events defined at significant pixel value changes along the projection of the currently estimated contour in the captured images. A low correlation between these two signal functions implies an inconsistency between the floor geometry and the input images which triggers an application of our GRR.

4.1.1 Input Images

The signal functions exploit the observation that faces with different MW aligned normal vectors will typically have different observed pixel values in the input images. For this to be true in general, we assume i) a captured image is taken from a camera location such that two adjacent façades on the same floor are colored differently – this can be due to illumination or to a change of albedo between the façades, and ii) the appearance of windows and shadows is not dominant or, in the case of windows, recognizable as small dark patches. Small dark patches are easily identifiable and removed mostly automatically during image preprocessing. Further, our later described event weighting scheme will reduce the importance given to spurious pixel intensity value changes caused by windows and shadows. In addition, the method of [23] could be used to find/remove windows and the method of [24] could be used to give a façade a single overall intensity and mitigate the negative effect of shadows.

Given the above assumptions, for each floor we compute a signal function directly either from the geometric model or from color-segmented captured images. Given the currently estimated floor geometry, it is straightforward to obtain the positions t of contour changes for defining G_s. To define the photometric events for I_s, we perform a mean-shift segmentation [3] which results in faces pointing in the same MW direction mapping to the same color. A change in segmented pixel color along the contour implies a direction change. We do not need to define whether it was a "left turn" or "right turn" – this will be determined implicitly by our grammar and constraints.

4.1.2 Signal Functions

A signal function $G_s(t)$ or $I_s(t)$ is defined similarly to a floor's polyline but using V uniformly-separated point samples $p_1, p_2, \dots, p_V$ on the polyline for floor i and at the middle height of the floor (for brevity, we omit the i index from the p's). We set $G_s(t_v) = 1$ when there is a change in the geometric normal between p_v and p_{v+1}, where $t_v \in [0,1]$ corresponds to the parametric position of p_v along the contour and $v \in [1, V]$. Since the transition might occur at the end of the floor contour we let $p_{V+1} = p_1$. In a similar fashion, we set $I_s(t_v) = 1$ when the segmented pixel values corresponding to p_v and to p_{v+1} are different. For all other sampled points, the signal value is 0.

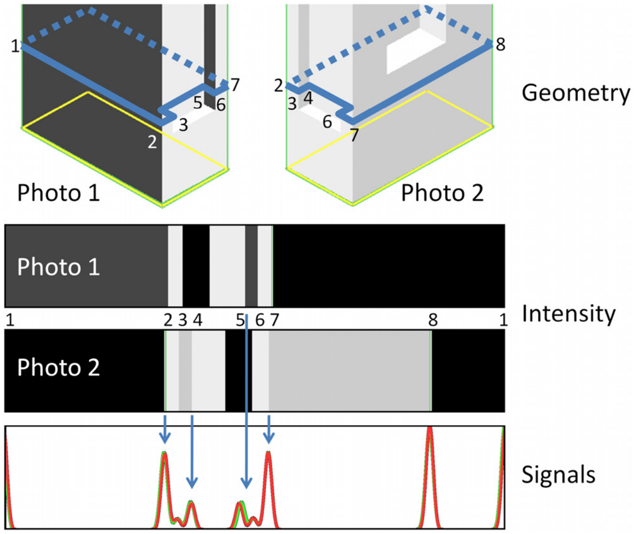

Figure 3. Geometric and Photometric Signals. These two signals register the events observed and in the geometry (turns) and in the photos (changes in intensity).

Photometric events from multiple captured images are simultaneously registered in the same signal I_s. This is possible because the floor contour along which the sample points are generated is the same for all images. Although a large number of images improves the reconstruction, our experience is that 2 to 4 images is sufficient.

As a next step, a per-event importance metric is used to scale the signal values. Our importance metric gives more weight to events that are distant from other events. Our intuition is that ignoring one of such events might yield a large error in overall floor shape. In contrast, ignoring an event in other parts will alter the accuracy of the detailed reconstruction but not the overall floor shape. In all cases, ensuring the events yield a closed, coherent, and plausible structure is enforced by the constraints. In the following, we describe the procedure for G_s. For when $G_s(t_v) = 1$, let $d^P{}_v$ and $d^N{}_v$ be the parametric distance from t_v to the previous and next sample points (events) whose associated signal value is also non-zero. Then, we perform

$$G_s(t_v) = G_s(t_v) \cdot (d^P{}_v + d^N{}_v)^2 \qquad (4)$$

for all $v \in [1, V]$ and normalize the signal. Afterwards, a similar procedure is applied to signal I_s.

Further, we smooth the generated signals in order to reduce noise that could lead to misinterpretation. Similarity between G_s and I_s impulse signals is computed using Pearson correlation because it is unaffected by the relevant maximum values. However, it is strongly affected by a small lateral shift in the impulses -- such a shift can be a common result of calibration error or image noise. To overcome this, we replace each sharp impulse by a zero-mean Gaussian distribution. Since the signals are parameterized by normalized values, a suitable width for the Gaussians is independent of the actual building sizes and is mostly determined by the typical frequency of and relative distance between events. For all buildings, we use a constant variance σ^2 that was experimentally determined.

Hence, the final form of the geometric signal is

$$G_s(t_v) = \frac{G_S(t_v)\cdot\left(d^P{}_v+d^N{}_v\right)^2}{\max_{v\in[1,V]}\left(d^P{}_v+d^N{}_v\right)^2} \sum\nolimits_{w\in[1,V]} \mathcal{N}(t_v; t_w, \sigma^2), \quad (5)$$

where $\mathcal{N}(t_v; t_w, \sigma^2)$ is a Gaussian function with mean t_w, variance σ^2, and evaluated at t_v. The function I_s is smoothed in a similar fashion.

4.2. Alteration of Floor Contours

Our floor alteration procedure determines if the next floor up the building requires alteration, within each floor which segments need to be rewritten, and what are the parameter values for each application of our GRR. To prevent applying our GRR to all floors, we only inspect floors up the building with a correlation value beneath a threshold. However, because the images are typically taken from a bird's eye perspective, the correlation values gradually change for several floors before stabilizing to a new lower value. This is due to the presence of small roof top structures that appear on the sides of the building because of the contour being push inwards from one floor to the next. Hence, the actual floor for a contour change is not immediately known but rather a range of potential floors is estimated. Our method then applies a contour change starting at each one of the floors in this range and ultimately chooses the floor that upon its alteration, and to all the ones above it, produces the best improvement.

Our GRR is applied only to the segments of a floor whose corresponding G_s and I_s signals mismatch. This selective application further reduces the number of times

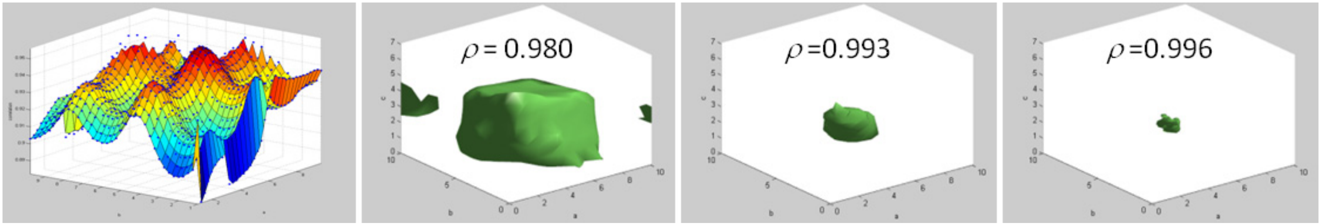

Figure 4. Smoothness of correlation between signals. (Left) The correlation between $\mathrm{G_s}$ and $\mathrm{I_S}$ for an example contour is shown as a function (height) of varying GRR parameters a and b (for a constant value of c). (Middle-left, middle-right, right) We vary all three parameter values (X, Y, Z axes).The isosurfaces at three correlation values are shown and indicate a well localized optimum.

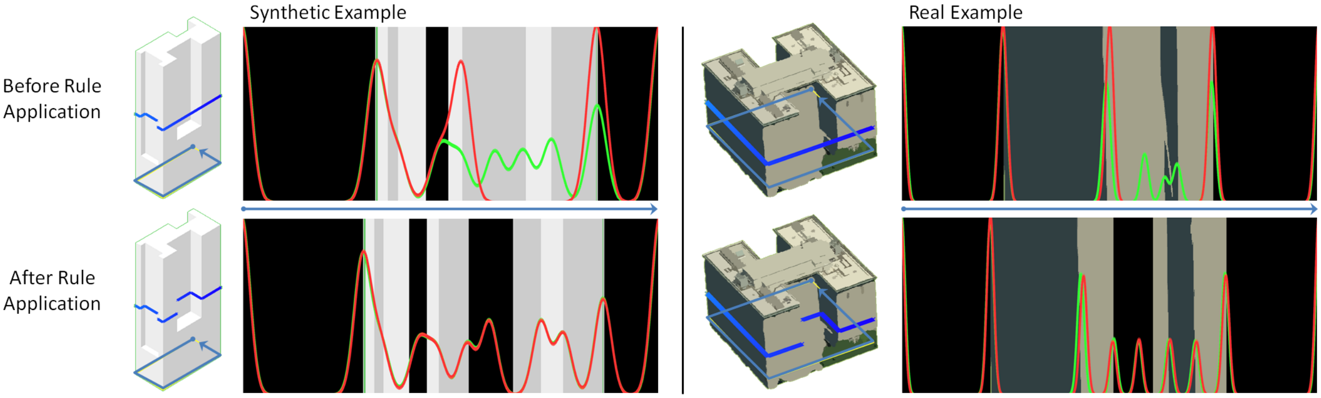

Figure 5. Optimization of Geometric and Photometric matching. The geometric (red) and photometric (green) signals are shown for a synthetic (left) and a real (right) example along the contour in blue. The signal match is low for the unmodified geometry (top) and high after GRR optimization. The intensity signal corresponds to the lower row of pixels of the box (other pixels shown for context).

the GRR optimization is performed. All segments to which GRR was tentatively applied and results in a correlation improvement are sorted in increasing order. The GRR is actually applied to the segment with the lowest correlation, and continues with all segments below a threshold correlation value. Segments affected by an application of GRR (e.g., the adjacent segments of a "L-shape" or "pushback" transition) are removed from the sorted list, have their correlation benefit recomputed, and are re-inserted into the list.

Applying the GRR to a segment consists in determining the values of a, b, and c that maximize the correlation $\rho_{G,I}$ of the corresponding signals G_s and I_s, given by

$$\rho_{G,I} = \frac{\sum_{\text{v}\in[1,\text{V}]}(G_S(t_v)-\overline{G_S})(I_S(t_v)-\overline{I_S})}{(V-1)\sigma_S\sigma_I} \quad (6)$$

where $\overline{G_s}, \overline{I_s}, \sigma_S, \sigma_I$, are the means and standard deviations of G_s and I_s. Equation (6) is applied to the portion of the signals corresponding to the segment plus an additional fraction of the adjacent segments. This helps to find adequate parameters when simultaneous and adjacent transitions occur. In preliminary experiments, we found the correlation values to vary smoothly as a function of a, b, and c and showing a clear localized optimum (Figure 4). Thus, we first perform a coarse sampling of the parameter space $\{a, b, c\}$. Then, we apply a nonlinear least-squares optimization starting with values that returned the largest correlation during coarse sampling.

5. Implementation Details

For the captured images, we used oblique-angle aerial imagery from Bing maps. Since in our prototype system we do not have easy access to pre-computed geo-referenced data, we use standard camera calibration to obtain camera focal length and pose parameters. Plane-based calibration can be performed using street vector data. The building footprint, or its bounding box, can be obtained from cadastral maps (or easily drawn by hand).

To remove segmented background pixels surrounding a building in a captured image, we first remove all pixels outside of a vertical extrusion of the building footprint's bounding box. Then, we remove segments of background pixels that intersect the aforementioned extrusion – this method works so long as the segments of background do not have the same color as the building. Segmented windows are mostly removed by selecting all small and dark patches. Segmentation imprecision is ameliorated by our signal weighting scheme and optimization. In practice, these preprocessing operations are nearly automated and require only a few mouse clicks.

6. Results and Discussion

We have used our approach to automatically reconstruct several real-world and synthetic buildings from one or more aerial views. Since the GRR is a key component of this method, we performed experiments on several test edges and floors to verify its behavior. Figure 5 shows the visualization of the geometric and photometric signals for two such test cases. Before the GRR is applied on a contour of the building (top), there is a clear mismatch between both signals, which results from the fact that some of the changes in intensity are not paired by changes in geometry. The GRR is applied with parameter values that maximize the correlation between geometric and photometric events.

Some of the real-world buildings are shown in Figures 1 and 6. For each building we show one or two of the input images, the automatically adapted floor contours superimposed on the images, and renders of the 3D model without textures and with textures computed by projecting the input images onto the model.

Each of the first three buildings in Figure 6 depicts a case of the GRR. The geometry of the first building is obtained by applying several pushbacks in two of the building façades. For this case, our optimization determines that the best matching between the geometry and the images is obtained by setting parameters a and b to zero. The second building is obtained by applying two U-shape instances of the GRR, each with different non-zero parameters a, b, c.

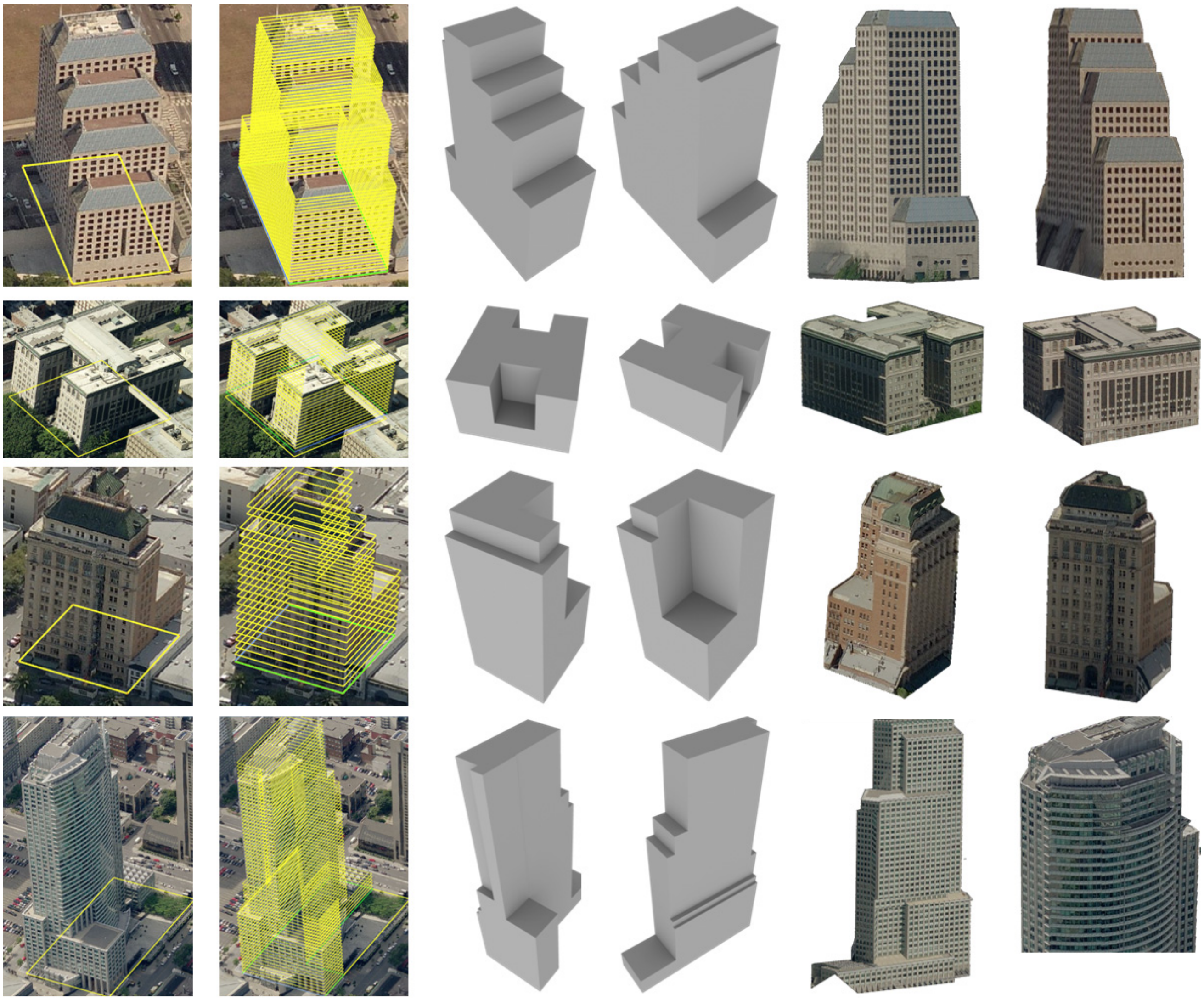

Figure 6. Results. Four of the buildings reconstructed by our method are shown. The first column shows one of the calibrated photos and the footprint used for each example. The second column shows the contours of each floor that have been automatically adapted to match the images. The last four columns show the reconstructed 3D models with and without projective texture mapping.

The third building is obtained by modifying a mid-height contour of the building with an instance of an L-shape GRR, followed by two pushbacks applied in the upper floors.

Notice that in all of these cases, the lower part of at least one façade of each building is occluded in the images either by smaller neighboring buildings or by trees. These occlusions are overcome by the logic we use to apply the GRR. The GRR is only applied if the correlation is significantly improved after its application. The correlation process is robust to noise (i.e., occlusions) because only structural changes that are determined to improve the matching of the signals, and obey the constraints of a plausible floor/building, are applied. Thus, a building can be reconstructed despite partial occlusions.

Figures 1 and 6 (bottom row) show two more complex buildings reconstructed using several applications of the GRR. In Figure 1, the reconstruction starts from the bounding box and automatically detects the L-shape floor after reaching the second floor. Due to shadows, our method fails to detect the shallow U-shape in the first floor of this building (top row, right façade) since no changes in intensity associated to this geometric event are apparent in the segmented photos. The bottom row of Figure 6 shows our method applied to a building that does not strictly follow the MW assumption. The curved façade of the building still exhibits a difference of intensity with respect to its neighboring façades which allows for a reasonable reconstruction. The angled façade in the first row of Figure 6 is modeled with a flat face and appears to be slanted only because of the applied projective texture mapping. Notice that all the reconstructed models consist of mutually orthogonal flat faces.

7. Conclusions and Future Work

We have presented an automatic method to reconstruct 3D building models from calibrated aerial imagery. Our method develops a grammar-based representation able to represent buildings with façade orientations that approximately follow a Manhattan-world assumption. The grammar converts the reconstruction of a building into a sequential process of refining a coarse initial building model (e.g., a box) using one generalized rewriting rule. The parameters values for each application of this rule can be robustly computed using color segmented aerial images where the actual pixel intensity values are not critical as long as façades with different Manhattan directions are colored differently. Our results show the capability of our approach using various real-world and synthetic models.

With regards to limitations and future work, there are several items we wish to pursue. First, we plan to support simultaneous applications of our GRR by explicitly optimizing for them -- currently such is not explicitly handled. Second, windows and shadows can be problematic. One option we will explore is detecting window symmetries and using normal clustering [12] as ways to improve grouping of façade pixels. Third, obtaining occlusion free images in dense building areas is challenging; thus, we look to symmetry-based methods to improve reconstructions in such cases. Fourth, our grammar could be extended to arbitrarily shaped buildings and be integrated with procedural modeling of façades [20]. Fifth, we will investigate using information in upper floors to self-correct erroneous rule applications in lower floors.

Acknowledgements. Work partially supported by NSF OCI 0753116 and Purdue Computing Research Institute.

References

[1] Alegre F., Dellaert F., A probabilistic approach to the semantic interpretation of building facades. Int'l Workshop on Vision Techniques Applied to the Rehabilitation of City Centres, 1-12, 2004.

[2] Aliaga D., Rosen P., Bekins D., Style Grammars for Interactive Visualization of Architecture. IEEE TVCG, 13(4), 786-797, 2007.

[3] Comaniciu D., Meer P., Mean Shift: A Robust Approach toward Feature Space Analysis. IEEE PAMI,34(5), 2002.

[4] Cornelis N., Leibe B., Cornelis K., and Van Gool L., 3D Urban Scene Modeling Integrating Recognition and Reconstruction. IJCV, 78(2), 121-141, 2008.

[5] Coughlan J.M., Yuille A.L., Manhattan world: Compass direction from a single image by bayesian inference. IEEE ICCV, 941–947, 1999.

[6] Debevec P. E., Taylor C. J., Malik, J., Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach. SIGGRAPH '96. 11-20, 1996.

[7] Dick A.R., Torr P.H.S., Ruffle S.J., Cipolla R., Combining single view recognition and multiple view stereo for architectural scenes. IEEE ICCV, 268-274, 2001.

[8] Früh C., Zakhor A., An Automated Method for Large-Scale, Ground-Based City Model Acquisition, IJCV, 60(1), 2004.

[9] Furukawa Y., Curless B., Seitz S.M., Szeliski R., Manhattan-world stereo, IEEE CVPR, 1422-1429, 2009.

[10] Furukawa, Y., Curless, B., Seitz, S.M., Szeliski, R. Reconstructing Building Interiors from Images. ICCV, 2009.

[11] Jiang N., Tan Ping, Cheong L.F., Symmetric Architecture Modeling with a Single Image. ACM TOG 28(5), 2009.

[12] Koppal S.J., Narasimham S., Clustering Appearance for Scene Analysis. IEEE CVPR, 1323-1330, 2006.

[13] Koutsourakis P., Teboul O., Simon L., Tziritas G., Paragios N., Single View Reconstruction Using Shape Grammars for Urban Environments. IEEE ICCV, 2009.

[14] Lafarge F., Descombes X., Zerubia J., Pierrot M., Building reconstruction from a single DEM. IEEE CVPR, 1-8, 2008.

[15] Lee D.C., Hebert M., Kanade T., Geometric reasoning for single image structure recovery. IEEE CVPR, 2009.

[16] Lingyun Liu., Stamos I., A systematic approach for 2D-image to 3D-range registration in urban environments. IEEE ICCV, 1-8, 2007.

[17] Mech R, Prusinkiewicz P., Visual Models of Plants Interacting with Their Environ. SIGGRAPH, 397-410, 1996.

[18] Micusik B., Kosecka J., Piecewise planar city 3D modeling from street view panoramic sequences, IEEE CVPR, 2009.

[19] Müller P., Wonka P., Haegler S., Ulmer A., Van Gool L., Procedural Modeling of Buildings. Proc. ACM SIGGRAPH, 614-623, 2006.

[20] Müller P., Zeng G., Wonka P., Van Gool L., Image-based Procedural Modeling of Facades. Proc. ACM SIGGRAPH, 26(3), 2007.

[21] Pollefeys M., Nistér D., Frahm J., Akbarzadeh A., Mordohai P., Clipp B., Engels C., Gallup D., Kim S., Merrell P., Salmi C., Sinha S., Talton B., Wang L., Yang Q., Stewénius H., Yang R., Welch G., Towles H., Detailed Real-Time Urban 3D Reconstruction from Video. IJCV, 78(2), 2008.

[22] Poullis C., You S., Automatic reconstruction of cities from remote sensor data, IEEE CVPR, 2775-2782, 2009.

[23] Sung C.L., Nevatia R., Extraction and integration of window in a 3D building model from ground view images. IEEE CVPR, 113-120, 2004.

[24] Troccoli A., Allen P., Building Illumination Coherent 3D Models of Large-Scale Outdoor Scenes. IJCV, 78(2), 2008.

[25] Verma V., Kumar R., Hsu S. 3D Building Detection and Modeling from Aerial LIDAR Data. IEEE CVPR, 2006.

[26] Vestri C., Devernay F., Using Robust Methods for Automatic Extraction of Buildings. IEEE CVPR, 133, 2001.

[27] Wang L., You S., Neumann U., Semiautomatic registration between ground-level panoramas and an orthorectified aerial image for building modeling. IEEE ICCV, 1-8, 2007.