2D-3D Fusion for Layer Decomposition of Urban Facades

Yangyan Li[†] Qian Zheng[†] Andrei Sharf[‡,†] Daniel Cohen-Or[◇] Baoquan Chen[†] Niloy J. Mitra[*]

[†]SIAT, China [‡]Ben-Gurion Univ. [◇]Tel Aviv Univ. [*]UCL

Abstract

We present a method for fusing two acquisition modes, 2D photographs and 3D LiDAR scans, for depth-layer decomposition of urban facades. The two modes have complementary characteristics: point cloud scans are coherent and inherently 3D, but are often sparse, noisy, and incomplete; photographs, on the other hand, are of high resolution, easy to acquire, and dense, but view-dependent and inherently 2D, lacking critical depth information. In this paper we use photographs to enhance the acquired LiDAR data. Our key observation is that with an initial registration of the 2D and 3D datasets we can decompose the input photographs into rectified depth layers*. We decompose the input photographs into rectangular planar fragments and diffuse depth information from the corresponding 3D scan onto the fragments by solving a multi-label assignment problem. Our layer decomposition enables accurate repetition detection in each planar layer, using which we propagate geometry, remove outliers and enhance the 3D scan. Finally, the algorithm produces an enhanced, layered, textured model. We evaluate our algorithm on complex multi-planar building facades, where direct autocorrelation methods for repetition detection fail. We demonstrate how 2D photographs help improve the 3D scans by exploiting data redundancy, and transferring high level structural information to (plausibly) complete large missing regions.*

1. Introduction

Fast and accurate digital acquisition of urban buildings remains a challenging task. While procedural modeling provides an attractive and effective option for generating high quality models of buildings, and has been employed for creating virtual cities, it remains unsuitable for digital archival of existing cities. A common solution for such digital acquisition is image-based modeling, which has been used to produce realistic 3D textured models using various degrees of manual assistance. However, even state-of-the-art image-based modeling methods require a large number of photographs to create models with sufficient 3D geometric details. An alternative is to directly use 3D scanners. In either case, the resultant 3D models are represented as low level primitives, and lack explicit encoding of high level structures, which are characteristics of urban facades.

High resolution laser scanners can be used for high quality model acquisition even with mm-resolution accuracy. Such scanners, however, are slow and have small working volumes, making them unsuitable for digitizing buildings and city blocks. Alternately, 3D LiDAR scanners are attractive as they are fast, easy to use, and capable of generating rough coherent scans of large structures like building facades. Unfortunately, such scans are noisy, sparse, and typically have large missing parts (see Figures 1 and 8). Although they provide a cursory impression of the scanned buildings, in the raw point cloud form they are unsuited for any practical application or digital inspection.

On the other hand, 2D photographs have important complementary characteristics to 3D scans. They are high resolution, noise-free, and, unlike LiDAR scans, they typically cover more of the building facade. Importantly, as cameras are portable and ubiquitous, it is easy to obtain photographs from various viewpoints, possibly from locations where it is challenging to scan from, e.g., rooftops.

The complementary traits of 3D scanners and photographs naturally suggest the use of a multi-modal acquisition approach. We present an algorithm for careful fusion of these two acquisitions modes, targeted specifically towards urban buildings with large-scale repetitions.

Given multi-modal inputs comprising of an incomplete noisy 3D point set $\mathcal{S}$ and a single or multiple photographs $\{I_i\}$, our goal is to create an enriched 3D model with information extracted and fused across the two modes. A major challenge is due to large missing regions common in LiDAR scans (see Figure 1). We rectify the photographs and register the two modalities together. However, projecting the 3D information from the scan over the registered 2D photograph only partially augments the photograph with depth values (Figure 1 (mid-left)).

Another challenge is detecting repetitive patterns, prevalent in urban facades. Typical building facades are not restricted to simple planar faces. As a result, even in

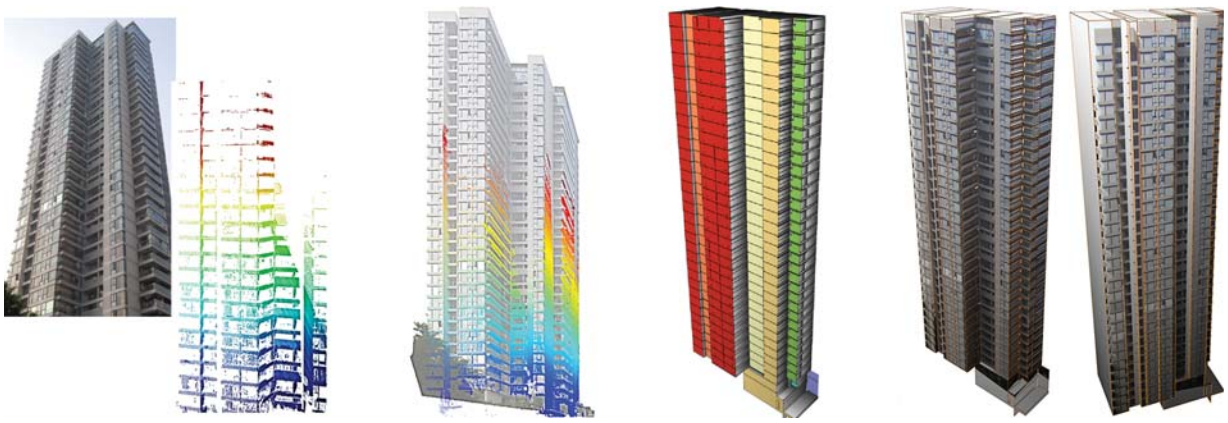

Figure 1. Given a 2D photograph and 3D LiDAR scan of a building (left), we overlay the scan over the rectified photograph (semi-transparent) (mid-left). Analyzing the fusion of the two modes allows decomposing the scene into depth-layers (distinctively colored, mid-right) followed by a per-layer symmetry detection that allows completing and augmenting the LiDAR scan with enhanced texture information (two views, right).

a perfectly rectified photograph, repeated protruding elements like balconies may not appear as regularly spaced elements in image space due to the perspective distortion (see Figure 4). Thus, any global autocorrelation-based symmetry detection approach fails. Hence, learning the self-symmetries requires decomposing the building facades into multiple planar depth-layers. We accomplish such a decomposition by augmenting parts of the photograph with depth values transferred from corresponding parts of the scan, and propagating depth labels across the whole facade using a multi-label assignment. Finally, we detect self-symmetries in the depth-augmented photos and propagate both geometry and texture across repetitions to complete missing parts in the scan (see Figure 2).

Specifically, fusing the two modes enables: i) augmenting the photos with depth, ii) decomposition of depth-augmented photos into consistent depth layers, iii) automatic image space repetition detection in each layer, and (iv) completion and integration of per-layer geometry and texture information.

2. Related Work

Given the large volume of work on urban modeling, we refer the reader to the recent survey by Vanegas et al. [2009] for a comprehensive coverage. Here we only focus on previous works most closely related to ours, in particular those addressing self symmetries and 2D/3D integration in the modeling process.

Image-based modeling. Works on automatic reconstruction of urban scenes have mostly been based on collections of photos [3, 6, 8, 18, 20, 22] or multi-view video [14], relying on photogrammetric reconstruction and image-based modeling techniques. Debevec et al. [2] 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. More recently, Sinha et al. [18] present an interactive modeling system using unordered sets of photographs, leveraging the piecewise-planarity of architectural models. Mueller et al. [12] perform analysis on 2D facade images in order to generate a 3D procedural model counterpart. Xiao et al. [22] efficiently model facades from images by decomposing facades into rectilinear elementary patches. Later they extend the semantic segmentation and analysis approach to more general scenes, to produce visually compelling results by imposing strong priors of building regularity [23]. Special shape symmetries can also be leveraged to model architectural objects from a single image [9].

Integrating 2D with 3D. Diebel and Thrun [4] combine low-resolution range images and high-resolution registered camera images to create high-resolution range images using Markov random fields. Stamos et al. [19] perform automatic registration of 2D images with 3D range scans by matching linear features between the range scans and the photographs. The alignment is then used to optimally texture map the photographs onto the dense model. Images can also be enhanced or augmented with information presented in 3D, as exemplified by the work of DeepPhoto [10]. Unlike previous attempts, we tightly couple two input modalities, images and 3D range scans, to produce 3D geometry.

Symmetry analysis. Symmetries, repetitions, and regularity have been extensively studied in the context of image analysis, and to a lesser extent for 3D geometry, with application towards procedural modeling, scan completion, and improvement. We review only a subset of works that are

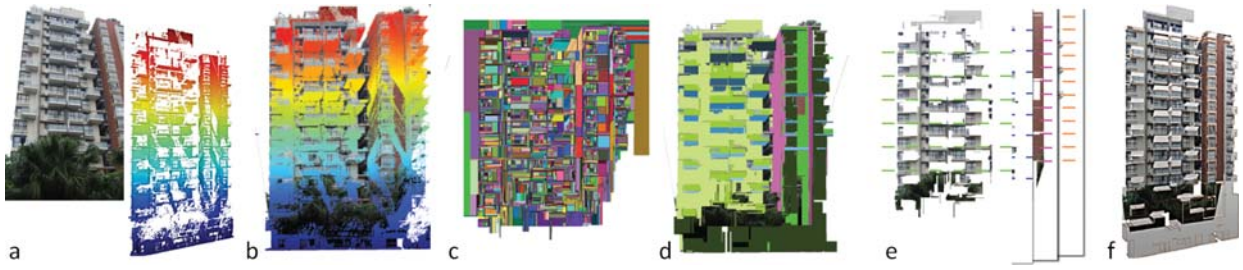

Figure 2. Starting from an input photo and a LiDAR scan (a), the two acquisition modes are registered (b). 3D primitives in the input scan are detected and used to over-segment the photograph into rectangular fragments (c). Solving a multi-label assignment problem yields a depth-layer decomposition of the fragmented image (d). In each depth layer, repetitions are detected (e) and used to consolidate the inputs to produce a textured polygonal 3D model (f).

most related to our setting.

Schaffalitzky and Zisserman [15] automatically detect repeated image elements in planes, which are typical in urban facades. They compute features and use RANSAC to detect repetitions under projective transformations. Korah and Rasmussen [11] address the problem of automatically detecting 2D grid structures such as windows on building facades from images taken in urban settings. Pauly et al. [13] introduce a transform domain analysis coupled with a non-linear optimization to detect regular k-parameter structures in 3D shapes. Recently, user-assisted repetition patterns have been used for effectively improving sparse building scans [21, 24]. Wu et al. [21] present a feature-based method that extracts repetition and symmetry hypotheses from the rectified image. They make simplifying assumptions such as constant repetition height, and no gaps between floors to find the repetitive pattern. Nevertheless, due to perspective distortion, such methods fail to correctly detect repetitions in image space in scenes with multiple depth layers (e.g., see Figure 4).

3. Overview

Given multi-modal inputs comprising of an incomplete noisy 3D point set $\mathcal{S}$ and a photograph I, our goal is to create an enriched 3D model with information extracted and fused across the modes. The method easily extends to incorporate multiple images, when available (see Figure 8). A key observation is that the two data sources carry complementary information. The two modes can be jointly explored to produce 3D models of superior quality, which are otherwise difficult to achieve using a single source of acquisition. Specifically, while operations like depth estimation, planar facade detection and model consolidation can be robustly executed on 3D scans, others involving edge detection and finding translational repetitions, which are common in urban facades, are better performed in the dense image plane.

A typical building facade does not consist of a single dominant plane, but of multiple planar sections separated across depth, introducing different foreshortening in different layers. This poses a challenge to direct image space symmetry detection even on a rectified facade image (see Figure 4). Our algorithm first uses the depth information from the scan $\mathcal{S}$ to partition the image I into polygonal regions with consistent depth, denoted as *depth-layers*. We register the two data sources to simplify tasks like segmentation and depth extraction. Next, we compute depth-layers together with reliable 2D edge information by solving a multi-label assignment problem. Finally, we consolidate the data using repetitive patterns extracted from each depth layer to produce a complete textured polygonal mesh.

4. Depth-layer Decomposition

We decompose an input image into depth layers using a sparse LiDAR scan in the following principal stages: (i) rectification of image I and registration with scan $\mathcal{S}$, (ii) segmentation of image I into polygonal regions using edges from I and $\mathcal{S}$, and (iii) assigning consistent-depth to the extracted segments using a multi-label assignment formulation. We now elaborate the steps (see Figure 2).

2D-3D registration. We first manually mark two horizontal and two vertical lines on the input image I, extract the respective vanishing points, and solve for the metric rectification that take the extracted vanishing points to infinity and restore orthogonality relations. The user then marks two pairs of rectangles, two in 2D and two in 3D to indicate rough correspondence, which is then used to extract the camera pose for I. Later we use this estimated camera matrix to project $\mathcal{S}$ onto I.

2D-3D segmentation. We use the registered 2D-3D data sources to partition image I into segments, which are later used for depth-layer extraction. First, using RANSAC, we extract planar regions from the input scan $\mathcal{S}$ similar to [16]. We prune out outlier planes arising out of sparse sampling and noise using the *Manhattan-world* prior [1,5], which biases the planes to lie along principal directions. We estimate the three major axes by clustering plane orientations

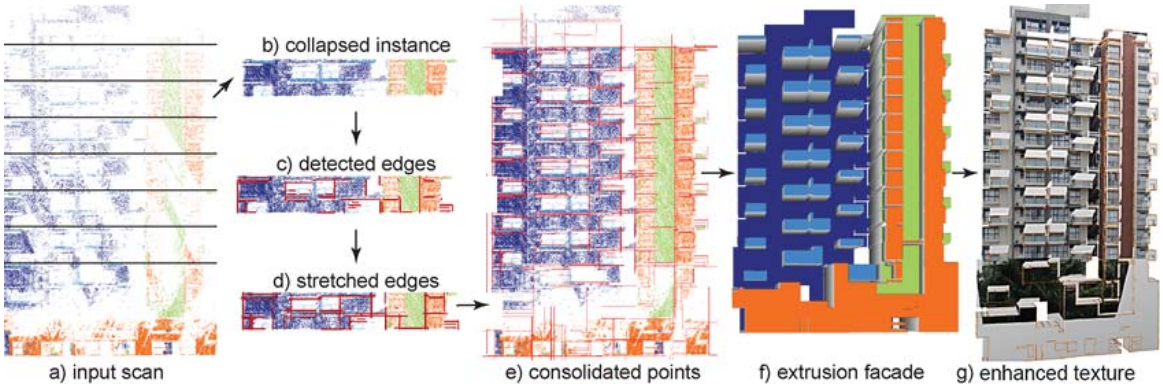

Figure 3. For each detected repetition pattern extracted from an image depth-layer, we collapse the corresponding points from scan to a single slab, project the slab to the frontal plane, perform edge detection biased to horizontal/vertical directions, propagate the edges to the consolidated point sets to create a rectangular fragmentation of the ortho-image. The fragment image is extruded, shown in gray, using the corresponding depth layer information to produce consolidated geometry (f) and texture, and thus a textured polygonal model (g).

and selecting the dominant three orthogonal clusters. Next, we compute two types of 3D edges: (a) edges extracted by identifying intersections of nearby planes, and (b) in-plane edges are extracted by applying a state-of-the-art edge detector [7] to the points in each plane. We overlay the visible extracted 3D edges and the principal planes on the image I. We greedily stretch each edge until it meets another (maximum of 5% of their respective lengths) and intersect them to form polygons, thus over-fragmenting the (rectified) image into polygonal segments.

Multi-label depth-layer generation. We use a multi-label assignment formulation to assign depths to the image fragments resulting in depth consistent layers using information from the set of planar primitives Φ extracted from $\mathcal{S}$. More formally, we look for a labeling that assigns for each segment $s \in I$ a label $p \in \Phi$, such that the labeling creates coherent unfragmented regions from the over-fragmented image. We create a graph with each segment $s \in I$ as a node, and connecting nodes s_i and s_j by an edge only if s_i and s_j share a boundary in I. Assigning consistent depths to the segments then amounts to solving a multi-label assignment problem on the constructed graph while balancing contradictory energy costs consisting of data and smoothness terms.

The data term E_{data} measures the cost of assigning a segment $s_i \in I$ to an extracted planar component $p_j \in \Phi$, and is broken up into terms measuring effects of neighbors and of occlusion. We utilize the 2D-3D registration to project points of p_j to pixels X_{p_j} on image I. Then, we measure the cost of assigning a segment s_i to an extracted planar component p_j as the average distance from all pixels of segment s_j to the closest pixels in X_{p_j}

$$ED_{neighbor}(s_i, p_j) = \sum_{x_k \in s_i} d(x_k, X_{p_j})/|s_i|,$$

where, $d(x_k, X_{p_j})$ measures the distance from pixel x_k to the closest pixel in X_{p_j}, and $|s_j|$ denotes number of pixels in segment s_j.

Further, due to sparsity and perspective projection of 3D scan points, foreground and background points can project to same planar segment resulting in ambiguities (see Figure 2). In other words, since points in $\mathcal{S}$ are sparse, points from different planes may be projected to the same area in I. In such cases, we favor assigning the area to the front plane rather than the back ones using the following occlusion term (see also [17]). For each scan point $q \in \mathcal{S}$, we define its occlusion influence region using a decay function:

$$occ_{eff}(x_k, q) = R^{1-d(x_k,q)^2/R^2},$$

where $d()$ is the distance from the overlayed scan point q to the pixel x_k and R is the maximal occlusion radius limiting occlusion effect to only pixels in the proximity of q. We set R to be the average distance between points in $\mathcal{S}$. For each assignment of a segment $s_i \in I$ to a certain planar component $p_j \in \Phi$, we compute for each pixel $x_k \in s_i$, the occlusion cost occ_{cost} as the sum of occlusion effects that are in front ("$\prec$") of the assigned planar component p_j, as

$$occ_{cost}(x_k, p_j) = \sum_{p_m \prec p_j, q \in p_j} occ_{eff}(x_k, q).$$

Finally, the occlusion data cost of a segment $s_i \in I$ under the assignment to a planar component $p_j \in \Phi$ is simply:

$$ED_{occlusion}(s_i, p_j) = \sum_{x_k \in s_i} occ_{cost}(x_k, p_j)/|s_i|.$$

By construction, our 2D image is over-segmented into polygons, while we have a primitive-guided conservative segmentation in 3D. Hence, the smoothness term E_{smooth} encourages adjacent segments $(s_i, s_j) \in I$, which are not

Figure 4. (Left) A global autocorrelation based method like that of Wu et al. [21] cannot correctly capture repetitions due to different repetitions depths. In contrast, our method operates on separate depth layers to produced desired result (right).

connected by any projected separating 3D edge, to be assigned to the same depth plane. More specifically,

$$E_{smooth}(s_i, s_j) = \begin{cases} T & f_{s_i} \neq f_{s_j} \text{ AND } \cup_k (\Phi_{e_k} \cap (s_i \cup s_j)) = \emptyset \\ 0 & \text{otherwise} \end{cases}$$

where, Φ_{e_k} denotes projected 3D edge e_k, T is a constant penalty for assigning neighbor segmentations with no 3D edge between them to different depth planes. In contrast, if there is no 3D edge between neighboring segments, the smoothness term encourages them to be assigned to the same plane. In all our experiments we used $T = 2$.

We use the combined energy $E = ED_{neighbor} + 4 \cdot ED_{occlusion} + E_{smooth}$ to minimize a multi-label assignment problem for the extraction of planar components from 3D scan and 2D segments. This yields a decomposition of the image into planar depth layers (see Figure 2).

5. 3D Geometry and Texture Fusion

An important source of data improvement lies in data redundancy, since urban facades often have large scale repetitions. Since photographs typically cover much more of the building than scans, we detect in-plane repetitions by analyzing the depth-layers of image I, and then transfer the recovered information to 3D in the consolidation stage.

In order to correctly detect available repetitive patterns, we apply the repetition detection method of Wu et al. [21] on each depth layer separately (see Figure 4). Subsequently, we use the extracted pattern from the image I to consolidate both the geometry of 3D point cloud $\mathcal{S}$ and the associated texture for plausible completion of missing data using the detected repetitions, as described next.

First, using the detected symmetry pattern, for each set of repetitions we bring the corresponding floors of the point set $\mathcal{S}$ to a single slab. We prune outlier points that have no nearby points in corresponding slabs. The points linked to all parallel depth planes are orthogonally projected onto a single frontal plane (see Figure 3b). We detect edges based on discontinuity of local density image [7]. Again using the Manhattan-world assumption, we bias the solution to horizontal and vertical edges. Finally, we transfer the extracted edges back to the constituent floors, and extend the edges to create local rectangular fragments.

The rectangular fragments inherit associated depth values from their respective depth planes. We use this information, to create an extrusion surface, thus reconstructing a scaffold for the facade faces (see Figure 3f). The associated textures are extracted from the input image I.

We render the model from the camera view to find corresponding texture patches in the input photograph. If the texture blocks do not suffer from missing parts or are not occluded (easily detected from camera view rendering of layers), we copy back the texture onto the scaffold reconstruction. When the corresponding texture fragment is corrupted, we use good texture blocks corresponding to repeated parts to produce plausible texture consolidation. Goodness of texture blocks is judged based on the consistency with their symmetric counterparts using SSD measure. Note that unlike for 3D geometry consolidation, we prefer to only touch texture parts in corrupted regions — this retains the subtle variations often exhibited by facade faces, rather than produce a sterile repetitive texture reconstruction (Figure 3g).

6. Results and Discussion

We tested our algorithm on a large number of datasets. Since many buildings have only a few distinctive styles, we present only representative examples to highlight different aspects of our technique (see also supplementary material). We experimented with medium to very tall buildings. Figures 1 and 5 show the results for tall and medium height buildings, respectively. Due to acquisition range limitations, the LiDAR scans are sparse and top floors are completely missed; some other parts go missing due to occlusions. With the assistance of 2D photographs, our method successfully detects repetitive patterns that are then used to complete missing geometry and consolidate its texture. Figure 5 demonstrates the advantage of fusing the two acquisition modes even when the input image has strong perspective distortion. In all our examples, only the described minimal user intervention was required to register the modes.

In Figure 7, we utilize a non-trivial repetition pattern composed of two repetition sets. Since separate repetitive patterns are handled independently by our fusion system, we can successfully handle complex repetition patterns.

In Figure 8 we demonstrate the effective power of incorporating multiple photos for accurate layer decomposition. Since 2D photographs are view-dependent, large parts of the building can be occluded in a single photograph (top row). Nevertheless, since photos are easy to acquire, we

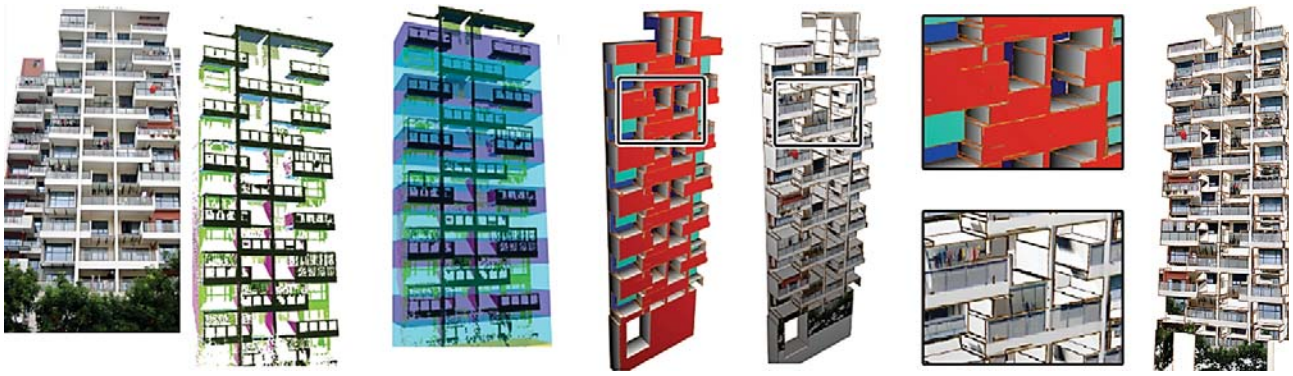

Figure 5. Top floors of buildings are barely visible to the LiDAR scanner resulting in large missing parts in the 3D scan. From left-to-right, figure shows input photo and scan, repetition pattern, and the resulting enhanced 3D geometry and texture (with zoom-ins).

can take multiple photos to get a larger coverage of the building. Our algorithm scales well to handle multiple photographs (bottom row). Through rectification and 2D-3D registration, photos are aligned together with the 3D LiDAR scan, thus occluded texture in one photo is completed from photos with different views. We decompose the photographs into depth-layers and accurately detect repetitive structures (see Figure 4), which allow completing and enhancing the 2D texture. Furthermore, we utilize repetitive patterns in 2D photos and their registration with 3D LiDAR scans, to consolidate the scan point by transferring geometry across constituent floors, completing missing regions and enhancing data.

In Figure 9 we evaluate our algorithm on a synthetic 3D model. Using a 3D building model, we virtually scan it by ray casting from only 3 views resulting in 190k points. The input point cloud simulates a real scan i.e. it is dense at lower floors and very sparse at top floors. Our algorithm accurately recovered all the depth-layers from the 2D photograph and 3D scan. The resulting textured polygonal facade (Figure 9-right) is identical to the input model although the topmost floors could not be recovered due to completely missing 3D samples and local repetitions.

Since we neither solve any non-linear systems nor does our algorithm contain any nested loops, our method scales roughly linearly with data size. We ran our algorithm on a 2.83 GHz Intel Core Q9550 with 4GB RAM and report the performance in Table 1. These are non optimized timings, and sum up to few minutes, which is negligible when compared to the actual acquisition times. Our method is controlled by a minimal set of parameters that stayed constant through all of our experiments: minimum support for RANSAC plane detection, 100 points; distance threshold, 0.05m. Please refer to additional material and video for further evaluation and examples of our method.

Limitations. Although symmetry detection and data fusion work in an unsupervised mode in our system, we still expect the user to bootstrap the system using initial markups. Spurious symmetry detection naturally produces imprecise final models (see Figure 6). Additionally, our method is ill-suited for enriching buildings with free-form facades that violate our piecewise planar facade assumption.

model	# pts.	image res.	lab.-assg.	rep.-det.	conso.
Fig. 1	338K	2592×3872	914s	357s	257s
Fig. 5	143K	2592×3872	981s	327s	213s
Fig. 7	109K	856×1624	239s	172s	47s
Fig. 8	463K	1296×1936	756s	680s	84s

Table 1. Performance statistics on various models.

Figure 6. In case of poor quality input and insufficient repetitions, we may fail to identify variations among repeated elements, e.g., different potted plants and railings across balconies. As a result, the final polygonal model incorrectly contains identical balconies.

Conclusions. We presented a method to enhance 3D urban models using depth-layer decomposition of photographs. Fusing 3D LiDAR and photographs exploits the advantage of both, thus enabling consistent decomposition. The information transfer is bidirectional: the 3D information is transferred to the photos, and then the depth-augmented photograph information is transferred back to 3D data. The final output comprise of consolidated geometry and texture information.

We have shown that a multitude photos further improves the quality of the results. However, as we showed in our examples, we use less than a handful of photos only, and not a dense set, as typically needed for SfM or stereo methods.

A research avenue is to further exploit different types of

Figure 7. Multi-modal data fusion produces textured polygonal model, with missing parts in LiDAR data being completed using extracted repetition patterns. In this example, two sets of repetitions, in dark blue (top floors) and light blue (low floors), are discovered.

Figure 8. Photographs can have significant parts occluded due to trees and other obstacles (top row). The occlusions, however, are usually different across views thus resulting in improved geometry and texture consolidation when we use more images. In this example we added another photograph from a different view to significantly improve the resulting 3D model (bottom row).

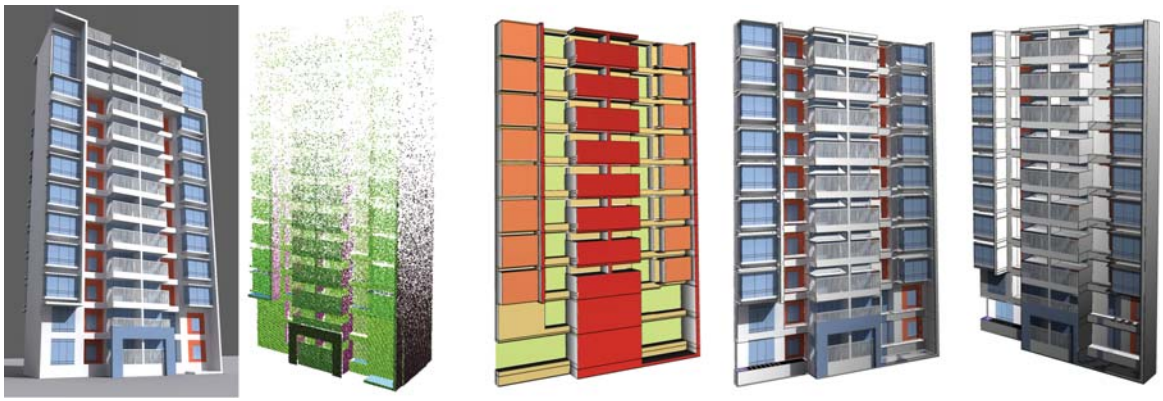

Figure 9. Evaluation of our multi-modal method on synthetic data. We virtually scan a 3D model using a ray-casting sampling technique (left). We accurately compute depth layers (middle) and utilize repetitions to compute a textured polygonal 3D facade (right).

photos as valuable sources of information in assisting the enhancement of the street-level acquisition, for example, satellite or aerial photography. We believe that the fusion of a multitude of modalities is a promising research direction, and that our work is only a step in that direction.

Acknowledgements. We would like to thank Guowei Wan for initial experiments on this project and Dror Aiger for inspiring discussions. This work was supported in part by NSFC (60902104, 61025012, 61003190), 863 Program (2011AA010500), CAS One Hundred Scholar Program, CAS Visiting Professorship for Senior Int'l Scientists, CAS Fellowship for Young Int'l Scientists, Shenzhen Science and Technology Foundation (JC201005270329A, JC201005270340A), Israel Science Foundation and European FP7 under grant agreement 276982.

References

[1] J. M. Coughlan and A. L. Yuille. Manhattan world: Compass direction from a single image by bayesian inference. In *Proc. ICCV*, pages 941–947, 1999.

[2] P. E. Debevec, C. J. Taylor, and J. Malik. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. *SIGGRAPH*, pages 11–20, 1996.

[3] A. R. Dick, P. H. S. Torr, and R. Cipolla. Modelling and interpretation of architecture from several images. *Int. J. Comput. Vision*, 60(2):111–134, 2004.

[4] J. Diebel and S. Thrun. An application of markov random fields to range sensing. In *Proc. NIPS*, 2005.

[5] Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski. Manhattan-world stereo. In *Proc. IEEE CVPR*, pages 1422–1429, 2009.

[6] Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski. Reconstructing building interiors from images. In *Proc. ICCV*, 2009.

[7] R. G. Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall. LSD: A fast line segment detector with a false detection control. In *IEEE PAMI*, volume 32, pages 722–732, 2010.

[8] M. Goesele, N. Snavely, B. Curless, H. Hoppe, and S. Seitz. Multi-view stereo for community photo collections. In *Proc. ICCV*, pages 1–8, 2007.

[9] N. Jiang, P. Tan, and L.-F. Cheong. Symmetric architecture modeling with a single image. *ACM SIGGRAPH Asia*, 2009.

[10] J. Kopf, B. Neubert, B. Chen, M. Cohen, D. Cohen-Or, O. Deussen, M. Uyttendaele, and D. Lischinski. Deep photo: Model-based photograph enhancement and viewing. *ACM SIGGRAPH Asia*, 27(5):116:1–116:10, 2008.

[11] T. Korah and C. Rasmussen. Analysis of building textures for reconstructing partially occluded facades. In *Proc. ECCV*, pages 359–372, 2008.

[12] P. Müller, G. Zeng, P. Wonka, and L. J. V. Gool. Image-based procedural modeling of facades. *ACM Trans. on Graphics*, 26(3):85, 2007.

[13] M. Pauly, N. J. Mitra, J. Wallner, H. Pottmann, and L. Guibas. Discovering structural regularity in 3D geometry. *ACM Trans. on Graphics*, 27(3), 2008.

[14] M. Pollefeys, D. Nistér, J. M. Frahm, A. Akbarzadeh, P. Mordohai, and et al. Detailed real-time urban 3D reconstruction from video. *Int. J. Comput. Vision*, 78(2-3):143–167, 2008.

[15] F. Schaffalitzky and A. Zisserman. Geometric grouping of repeated elements within images. In *Shape, Contour and Grouping in Computer Vision*, pages 165–181, 1999.

[16] R. Schnabel, R. Wahl, and R. Klein. Efficient ransac for point-cloud shape detection. *Computer Graphics Forum*, 26(2):214–226, 2007.

[17] S. Sinha, D. Steedly, and R. Szeliski. Piecewise planar stereo for image-based rendering. In *Proc. ICCV*, pages 1881–1888, 2009.

[18] S. N. Sinha, D. Steedly, R. Szeliski, M. Agrawala, and M. Pollefeys. Interactive 3D architectural modeling from unordered photo collections. *ACM Trans. Graph.*, 27(5):1–10, 2008.

[19] I. Stamos, L. Liu, C. Chen, G. Wolberg, G. Yu, and S. Zokai. Integrating automated range registration with multiview geometry for the photorealistic modeling of large-scale scenes. *Int. J. Comp. Vis.*, 78(2-3):237–260, 2008.

[20] T. Werner and A. Zisserman. New techniques for automated architecture reconstruction from photographs. In *Proc. ECCV*, volume 2, pages 541–555, 2002.

[21] C. Wu, J.-M. Frahm, and M. Pollefeys. Detecting large repetitive structures with salient boundaries. In *Proc. ECCV*, pages 142–155, 2010.

[22] J. Xiao, T. Fang, P. Tan, P. Zhao, E. Ofek, and L. Quan. Image-based façade modeling. *ACM Trans. on Graphics*, 27(5):1–10, 2008.

[23] J. Xiao, T. Fang, P. Zhao, L. Maxime, and L. Quan. Image-based street-side city modeling. *ACM Trans. on Graphics*, 2009.

[24] Q. Zheng, A. Sharf, G. Wan, Y. Li, N. J. Mitra, D. Cohen-Or, and B. Chen. Non-local scan consolidation for 3D urban scenes. *ACM Transactions on Graphics*, 29(3), 2010.