

2D-3D Lifting for Shape Reconstruction

Liangliang Nan¹ Andrei Sharf² Baoquan Chen³

¹Shenzhen VisuCA Key Lab/SIAT, China

²Ben Gurion University, Israel

³Shandong University, China



Figure 1: Given a partial 3D scan and a 2D photograph (left), we lift 2D shape structures into 3D yielding a faithful textured mesh reconstruction (right).

Abstract

We present an algorithm for shape reconstruction from incomplete 3D scans by fusing together two acquisition modes: 2D photographs and 3D scans. The two modes exhibit complementary characteristics: scans have depth information, but are often sparse and incomplete; photographs, on the other hand, are dense and have high resolution, but lack important depth information. In this work we fuse the two modes, taking advantage of their complementary information, to enhance 3D shape reconstruction from an incomplete scan with a 2D photograph. We compute geometrical and topological shape properties in 2D photographs and use them to reconstruct a shape from an incomplete 3D scan in a principled manner. Our key observation is that shape properties such as boundaries, smooth patches and local connectivity, can be inferred with high confidence from 2D photographs. Thus, we register the 3D scan with the 2D photograph and use scanned points as 3D depth cues for lifting 2D shape structures into 3D. Our contribution is an algorithm which significantly regularizes and enhances the problem of 3D reconstruction from partial scans by lifting 2D shape structures into 3D. We evaluate our algorithm on various shapes which are loosely scanned and photographed from different views, and compare them with state-of-the-art reconstruction methods.

Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—Geometric algorithms, languages, and systems

1. Introduction

Reconstruction from raw 3D scans remains a challenging problem in computer graphics. While state-of-the-art recon-

struction techniques provide effective solutions for computing the shape from a scan, it remains an open question how to recover the shape faithfully from only a partial scan of the object. Nowadays, the prevalence of commercial scanners

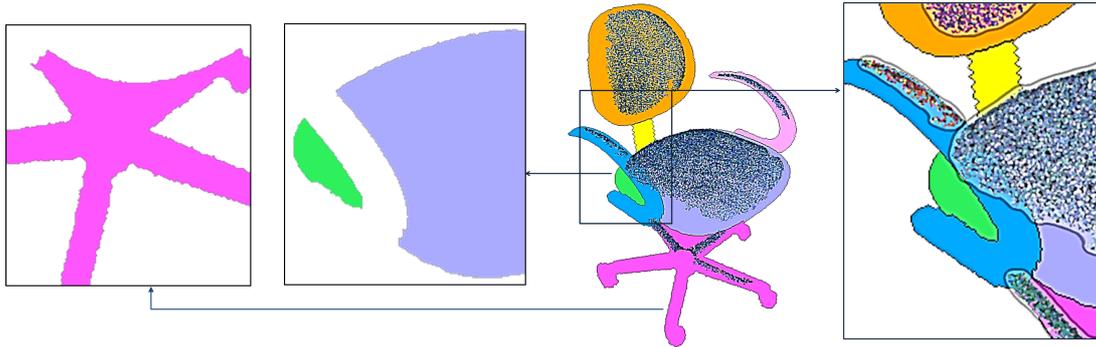


Figure 2: Multi-modal fusion of 3D scans and 2D photographs is challenging. Rightmost zoom, 3D points projecting onto the 2D photograph yield ambiguities: seat points projecting onto armrest, leg points projecting onto seat, and points missing between seat and back. Left zooms, perspective and occlusions yield a disconnected seat patches and distorted legs.

and depth cameras allow end-users to capture 3D objects in matters of seconds in a non-professional fashion. This trend generates increasing amounts of 3D scans which are non-perfect and partial. We observe a growing demand for algorithms that can handle such partial data and reconstruct the actual 3D shape effectively.

A core problem in 3D digital object acquisition is the coverage problem: due to various occlusions and material properties, significant portions of the object are either under-sampled or completely missing [KMHG13]. If scanned data is incomplete, the reconstruction of the 3D shape near missing regions is ill posed as an infinite number of valid surfaces may pass there. Smoothness and watertight constraints usually regularize the problem. However, such constraints are local and low-level, and unable to guide the reconstruction process towards the expected results. High-level constraints, such as symmetry, may provide efficient 3D scan enhancement and completion tools [LA11]. When data is highly incomplete, with large missing parts, it is impossible to infer a faithful 3D completion and reconstruction even with symmetrization (Figure 1 left).

Unlike 3D scans, 2D photographs are easy to acquire and usually have high resolution, thus providing important complementary information to 3D scans. These characteristics naturally lend themselves to detection of local shape properties and suggest the use of a multi-modal acquisition and reconstruction framework. This approach has been investigated lately with reference to urban reconstruction and modeling algorithms [SSS*, KNC*, LZS*11]. Nevertheless, multi-modal fusion is challenging due to the large imperfections in each of the individual modes as well as ambiguities in their fusion together (see Figure 2).

In recent years, the emerging RGB-D sensors (e.g. Kinect) allows simultaneous capturing of depth and color information. In this case, depth and color modalities are captured from nearly the same view point and are inherently regis-

tered together. Such information is not suitable for multi-modal fusion as both modalities capture the same geometry, missing essential complementary traits. From the completion and reconstruction perspective, the complementary traits of 3D scanners and 2D photographs are significant when the two sources are obtained from different views.

In our method, we leverage the 3D reconstruction of point clouds with high-level structures from 2D photographs. We utilize smooth color fragments, boundaries and connectivity information in the 2D photograph, and carefully fuse them with the 3D scan data for reconstruction. In Figure 1 an incomplete scanned chair is enhanced with a 2D photograph, resulting in a faithful 3D reconstruction.

Our input consists of an incomplete 3D scan with large missing parts that provides only a partial coverage of the shape, and an additional 2D photograph capturing the shape from a different view. To demonstrate our technique, it is sufficient to fuse between the 3D scan and a single 2D photograph. Nevertheless an extension to multiple photographs would probably further improve the results.

In a preprocessing step, we segment both 3D point-set and 2D photograph into smooth patches and smooth color fragments respectively. Usually meaningful segmentation is challenging using automatic tools, thus we allow user interaction to refine the segmentation. In the photograph, we detect patch boundaries and compute connectivity information across fragments. We register the point-set with the photograph and fuse depth information together with 2D fragments and boundaries. Utilizing this information, we lift the 2D fragments and boundaries into 3D using depth cues constraints.

Our paper makes the following novel contributions:

- a fusion method between a 3D scan and an additional 2D photograph which enhances sparse depth information with dense 2D fragments and boundaries.

- a scan to photo overlay analysis method which resolves ambiguities in the overlay due to perspective and occlusions.
- a constrained 2D to 3D lifting technique which maps 2D smooth fragments and boundaries onto the 3D scan.

2. Related Work

Given the large volume of work on shape reconstruction, we mostly focus on works that are related to our approach, in particular we discuss reconstruction with respect to prior-based, interactive and 2D-3D multi-modal techniques.

Prior-based surface reconstruction The problem of reconstructing a surface from scans has been researched extensively for more than two decades [HDD*, LPC*, BLN*13]. Many different techniques have been developed assuming a good local sampling density [KBH06, LCOLTE, OGG09]. Our method infers high-level structures from a 2D image to control the global reconstruction process.

High-level priors have been utilized by Pauly et al. [PMG*05] to reconstruct a 3D model using a database of 3D shapes that complete missing parts. Gal et al. [GSH*07] fit a small set of basic shapes at multiple scales via partial matching which augment the scanned data with noise-free samples, and sharp features. Schnabel et al. [SDK09] fill holes in the input by fitting basic primitives to missing data. More recently, the analysis of structural and global relationships in shapes has shown effective for regularization of surface reconstruction [LWC*11, LA13].

Interactive surface reconstruction Automatic reconstruction techniques may fail to faithfully reconstruct the 3D shape if scanned data consists of large missing parts. Interactive methods have recently showed the effective power of incorporating user assistance in the reconstruction loop.

Sharf et al. [SLS*] allow minimal user interaction to control a watertight reconstruction. Law and Aliaga [LA11] reconstruct a 3D shape from a single image by letting the user to define symmetry classes. In Nan et al. [NSZ*], the user loosely places architectural blocks in 3D space, and the algorithm snaps them to their position to reconstruct an urban scene. Similarly, Arikian et al. [ASF*13] allow interactive refinement of an architectural reconstruction process using 2D modification strokes. Habbecke and Kobbelt [HK09] present an interactive image-based modeling tool that allows the user to guide the reconstruction of a 3D model from stereo images.

2D-3D multi-modal fusion A majority of algorithms that fuse 2D and 3D modes are aimed at architectural modeling and reconstruction. The *Facade* system by Debevec et al. [DTM] generates 3D architectural models by manually building geometry proxies and linking related edges in several images.

In the realm of fusion based modeling techniques, Sinha et al. [SSS*] reconstruct piecewise planar 3D structures

by allowing the user to sketch 2D lines over photographs which are then mapped into 3D. Chen et al. [CKX*08] identify junctions, edges, and faces in free-hand 2D sketches for reconstructing 2.5D geometry. Xiao et al. [XFZ*] assume planar rectangular facades and compute details as their 2.5D elevations. Jiang et al. [JTC] reconstruct 3D architectural objects from a single image based on reflective symmetry. Recently, [ZYFY] propose a data-driven approach for flower petal modeling from partially scanned data. With priors from Botany, it fits a parametric model to the incomplete point cloud. At the same time when this work was underway, few methods were proposed to learn depth information from a set of 3D models [SHM*] or a set of images with known depth information [LSH14].

To reconstruct high-resolution range data, Diebel and Thrun [DT05] combine low-resolution range images and photos using Markov random fields. Stamos et al. [SLC*08] register 2D images with 3D range scans by matching linear features. Images are augmented with 3D information in the work of DeepPhoto [KNC*]. Recently, Li et al. [LZS*11] fused 2D photographs and 3D scans, for enhanced depth-layer decomposition, repetition detection and denoising. Similarly, Shen et al. [SFCH] utilize depth and RGB information for object reconstruction, by inferring structure information from the 2D image. Following in this path, we fuse arbitrary 2D photos and 3D scans, to infer smooth geometry, edges, and topological structures.

3. Overview

Our input consists of a sparse 3D point-set sampling of a physical object. We are given in addition a 2D photograph acquired from a different view.

In a preprocessing step, we segment the 3D point cloud into patches of smoothly varying normals, and segment the 2D photograph into smooth color fragments. The dense nature of 2D photographs allows us to efficiently extract 2D contours and build the fragments connectivity graph (see Figure 3 left). Typically, meaningful segmentation is challenging using automatic tools. Therefore, we allow user to iteratively refine the segmentation by strokes connecting/disconnecting nodes in a graph built on the initial patches.

We project the 3D point cloud onto 2D by trying to match the camera view of the 2D photograph. We register the projected point set and photograph using non-linear Gaussian mixture model matching that is robust to missing data.

We analyze the overlaid points and photograph and exploit mutual information to resolve ambiguities due to occlusions, missing parts, and perspective projection. Thus, we refine the segmentations and the connectivity graph (see Figure 3 middle). This provides with geometrical and topological properties which will constrain the 2D-3D lifting procedure.



Figure 3: Left-to-right: Given an incomplete partial 3D scan (top) and a 2D photo of a chair (bottom), we initially segment them into smooth patches. In 2D we also extract boundaries and compute an adjacency graph. We compute a non-rigid registration between the projected 3D scan and 2D photograph yielding an overlay (middle). In mid-right, the transparent overlay are the 3D points and the colored patches are the lifted 2D segments. We lift 2D structures into 3D using a constrained transformation resulting in a faithful reconstruction (right). Texture is also transferred from the non-rigid registration.

We define the 2D-3D lifting as a transform of fragments and boundaries from the 2D plane into 3D. This transformation is obviously underconstrained and therefore, we regularize it using 3D points as depth constraints and 2D connectivity as topological constraints. Our assumption is that the object's underlying geometry is piecewise smooth. Thus, 3D points projecting onto 2D boundaries and fragments serve as depth constraints in the piecewise smooth transformation. We triangulate the 2D fragments and use a *Thin Plate Spline* deformation to lift the planar fragment into 3D. Additionally, our connectivity graph forces adjacent 2D fragments to stay connected in 3D. The result is a reconstruction of the 3D scan with consistent polygonal boundary edges and triangulated interiors.

4. Technical Details

4.1. Preprocessing

Given a raw 3D point cloud, denote P as the uniformly down-sampled input, and $\{p_i(x, y, z), n_i(\partial x, \partial y, \partial z)\} \in P$ as the 3D point's position and orientation. I denotes the 2D photograph, where a pixel $\{q_i(u, v), c_i(r, g, b)\} \in I$ is defined by its position and color.

We initially oversegment the scan P into a set of piecewise smooth patches $S = \bigcup_k (s_k)$ with smooth varying normals by propagating k -closest neighbors for each point. We use a normal variation threshold of $n_i \cdot n_j > 0.8$ and $k = 7$ under a distance threshold $d < 1cm$. To extract smooth color patches with consistent polygonal boundaries in the 2D photograph, we apply the watershed algorithm using a non-parametric marker-based segmentation [Mey92]. We use a stroke thickness of 6 pixels in all our examples.

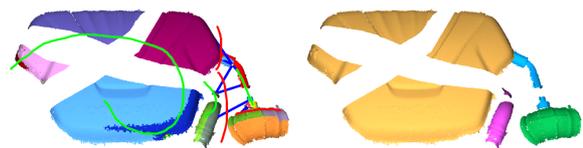


Figure 4: User draws strokes (left) which further merge (by green strokes) or split (by red strokes) 3D patches resulting in a meaningful segmentation (right).

User Interaction. Since automatic segmentation cannot guarantee separation into high-level meaningful parts, we allow manual user refinement in our preprocessing. In 2D, we utilize Grabcut [RKB], where the user draws strokes over the 2D patches to further split and merge them. In 3D, the user iteratively refines the segmentation by drawing two types of strokes over the 3D point cloud that merge and split the 3D patches (see Figure 4). For more detail about the segmentation of a 3D point cloud, please refer to the accompanying video.

4.2. 2D-3D registration

We register the 3D points and the 2D photograph by searching for the 2D perspective projection which best aligns the two data sets. This involves searching for the perspective transformation that corresponds to the photograph's camera view. Finding the exact solution to this problem is difficult since the search space is highly dimensional, and data is incomplete. Nevertheless, we only require an approximate registration between the two data sets.

To reduce the search space, we allow the user to coarsely

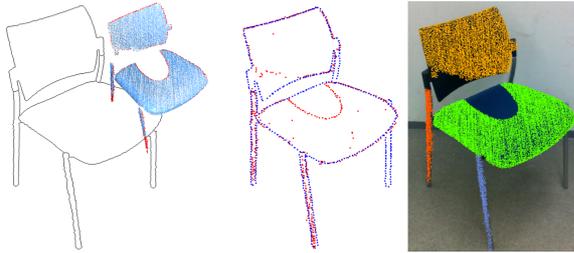


Figure 5: 2D-3D registration. 2D edges (blue) and 3D boundary points (red) are registered (mid), resulting in an overlay of the projected scan onto the image (right).

pre-align the 3D scan and 2D image simply by positioning them close together. Note the alignment is not necessary to be accurate as we refine the registration by a following step. We refine the coarse alignment between the projected point set and photograph by performing a non-linear registration of mixture of Gaussians. We represent both data sets with GMMs and compute their non-linear registration by finding the alignment between the two mixtures [JV11]. The L_2 distance between registered GMMs measures their registration quality. This approach handles our problem well as it is robust to large missing parts and non-linear deformations in the data (see Figure 5).

To speed up the search process, we reduce the data size by considering only object boundaries in the 2D photograph and 3D scan. Boundaries in the 3D scan are computed by analyzing the covariance matrix of the neighborhood of each point. Specifically, given a point p_i and its k neighbors $N(p_i)$, the covariance matrix of $N(p_i)$ is formed:

$$CV = \sum_{p_j \in N(p_i)} \{(p_j - o_i) \otimes (p_j - o_i)\}$$

where \otimes denotes the outer product of two vectors. $\lambda_i^0 < \lambda_i^1 < \lambda_i^2$ denote the eigenvalues of CV , then p_i is regarded as a boundary point if $\lambda_i^1 / \lambda_i^2 < \mu$ [GWM01]. k and μ are not sensitive to noise since we uniformly down-sampled the scan. In our implementation, we choose $k = 30$ and $\mu = 0.4$.

Note that the computed perspective projection is not accurate nor it is required to be. We define our lifting method in a robust manner to account for the approximate registration as well as noise, outliers, and missing regions.

4.3. 2D-3D overlay analysis

We analyze the overlay between the 2D segmented photograph and the projected 3D point cloud. The goal of this step is to resolve ambiguities and inconsistencies in the mutual 2D and 3D information and compute a mapping between 2D and 3D patches.

We observe the following inconsistencies and resolve

them: a 2D patch can be distorted or partial due to perspective projection, a 3D patch can project incorrectly due to occlusions and 3D points may be completely missing (see Figure 2). Using our points-to-photograph overlay, we resolve these inconsistencies through analysis of mutual information, and computing a consistent correspondence between depth and photograph patches.

Patch Voting. Our goal is having consistent depth values projecting per 2D patch. Since color and normal variation typically correlate, we aim at finding a surjective mapping from 2D to 3D patches. Nevertheless, 3D points can project incorrectly onto 2D. Specifically, multiple 3D patches may project together onto a single 2D patch, resulting in depth ambiguities per 2D patch. In Figure 2 (right), armrest and seat 3D points project together onto the armrest patch.

We compute a surjective mapping using a voting scheme. For each 2D patch r_i , each 3D patch s_j projecting onto it yields a voting calculated as the percentage of 3D points projecting inside the 2D patch $Proj(s_j, r_i)$. Thus for given image patch r_i , its best matching 3D patch is defined as $s(r_i)$ such that:

$$Proj(s(r_i), r_i) = \max_{s_j \in S} \{Proj(s_j, r_i)\}$$

2D photographs are typically dense, with accurate local connectivity information between patches. Thus, we build a connectivity graph $G(V, E)$ with nodes $v_i \in V$ corresponding to segmented patches $r_i \in R$ and edges E connecting adjacent patches.

Connectivity Refinement Occlusions in perspective projection yield holes and possibly an incorrect connectivity graph. We define three automatic operations to consolidate the graph structure using mutual 2D-3D information.

- **Connect:** given two disconnected 2D patches $\{r_i, r_j\}$ whose corresponding 3D patches $\{s_i, s_j\}$ are close: $d_{min}(s_i, s_j) < \epsilon$, we connect them by an edge $e \in E$. Here $d_{min}(\cdot)$ is the minimum distance between all point pairs of the two patches.
- **Merge:** In the case that a single 3D patch is mapped to several 2D patches, we merge them into one 2D patch and update the graph nodes respectively. The result of this step is a mapping where each 3D patch is assigned to one patches.
- **Disconnect:** given two image patches $\{r_i, r_j\}$ connected by an edge $e \in E$, we analyze their corresponding patches in the 3D scan $\{s_i, s_j\}$. If the closest Euclidean distance between the two patches is above a threshold $d_{min}(s_i, s_j) > \epsilon$, we disconnect this edge.

In all our experiments, ϵ is set to 10cm. Thus, our graph structure will be consistent with the 3D information in the scan. Note that since the scan is relatively sparse, it is possible that some parts are completely missing. We utilize the graph structure, to infer 3D values and constraints in the 2D-to-3D transformation following.

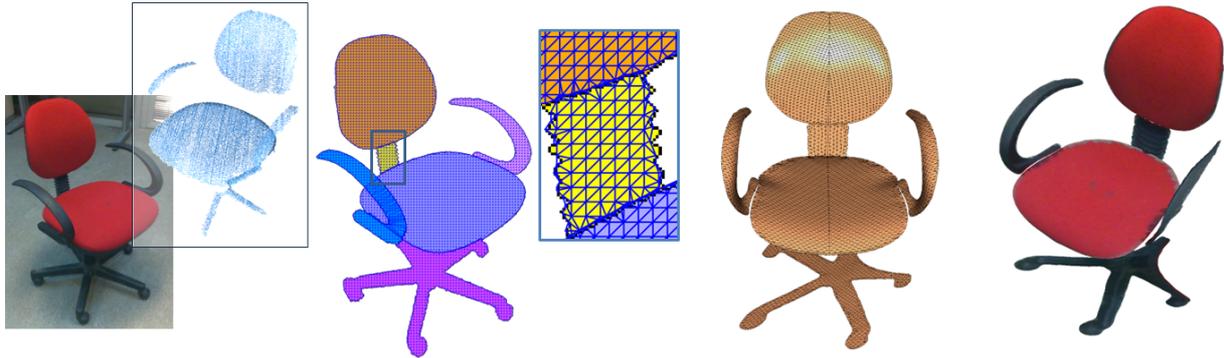


Figure 6: Reconstruction of a 3D chair with completely missing seat and back connector that is automatically inferred from the 2D photograph. From left to right: the input photograph and partial scan, 2D patch triangulation, reconstructed mesh, and textured model.

4.4. 2D-3D lifting transform

We transform planar 2D patches into 3D by computing a non-linear transformation optimization. We initially triangulate the 2D patches thus when lifted, they generate an explicit 3D mesh reconstruction.

Given a 2D patch, we compute its Delaunay triangulation by subsampling the patch boundary and inner part. Furthermore, for each triangulated 2D patch, we extract its 2D boundary polyline separating it from other patches. We use resolution of 5 pixels and 3 pixels to sample image patches and their boundaries respectively (see Figure 6 mid-left).

Given a 2D patch triangulation, computing its texture mapping is straightforward by mapping each vertex to its corresponding pixel in the photograph (see Figure 6 right)

Since lifting from 2D to 3D is underconstrained, we introduce three types of constrains with which we compute the transformation optimization.

- **Boundary constraints** 3D points projecting onto 2D image edges are used as boundary constraints. We analyze the 3D-to-2D registration and collect 3D points that register within an ω radius from the polyline L . We use a point to line segment distance in 2D $d_{pointLine2D}(\cdot)$ to find the corresponding 3D points as follows:

$$\{p'_i \in P \mid d_{pointLine2D}(p'_i, L) < \omega\}$$

For depth consistency reasons, we analyze the adjacency graph and select both patches sharing polyline L (if not on the shape boundary). Similarly, we select all 3D points projecting within ω from the polyline L and add them to the boundary constraints. Note that boundary constraints are sufficient for transforming a 2D patch when other information is not available. In Figure 6, the yellow patch connecting the back and seat is completely missing in the 3D scan. Nevertheless, it is lifted using boundary constraints from adjacent back and seat patches.

- **Inner constraints** Similar to boundary constraints we define also inner constraints. Given a triangulated 2D patch t_i , the projected 3D points registering with the vertices V of t_i are defined as inner constraints:

$$\{p'_i \in P \mid d_{pointPoint2D}(p'_i, V) < \omega\}$$

ω is set to 2 pixels in all our experiments.

- **Topological constraints** We infer topological constraints from boundary polylines and patch triangulation. Specifically, two adjacent 2D patches in the graph, sharing a boundary polyline will be lifted into 3D sharing the same boundary constraints for their mutual polyline. To resolve depth ambiguities, we only choose to use the boundaries that are connected in the scan. Thus, connectivity relations in 2D will be transferred into 3D, restricting connected patches to stay connected. Similarly, the 2D patch triangulation yields additional local connectivity constraints, restricting adjacent 2D vertices to transfer smoothly in 3D.

With the above constraints $\{p'_i, i = 1, 2, \dots, k\}$ as control points, we compute a 3D Thin Plate Spline (TPS) [Boo89] constrained transformation for each triangulated 2D patch t_i .

$$f(x) = \sum_{i=1}^k c_i \cdot \varphi(\|x - p'_i\|)$$

where $\varphi(\rho) = \rho^2 \log \rho$ is the radial basis kernel. The TPS fits a mapping function $f(x)$ between constraint pairs $\{y_i\}$ and $\{x_i\}$ defined as:

$$f_{tps} = \arg \min_f \left(\sum_{i=1}^k \|y_i - f(x_i)\|^2 + \lambda \iint [(\frac{\partial^2 f}{\partial x^2})^2 + 2(\frac{\partial^2 f}{\partial x \partial y})^2 + (\frac{\partial^2 f}{\partial y^2})^2] dx dy \right)$$

where λ is the regularization term for TPS deformation. We set $\lambda = 0.3$ to make the deformation smooth enough against noise in the scan (see Figure 7 right). The transformed 2D patches gives us the reconstruction of the 3D surfaces.

5. Results and Discussion

We have experimented with several data sets representing objects of different properties. Each data set is a point cloud and image pair, where the image is taken from a different view. Furthermore, we loosely scanned the objects in a natural manner, allowing gaps and occlusions to occur.

Reconstruction results Our results show plausible and faithful reconstructions even when the scanned data is highly incomplete. In Figure 6 we show a stool with a completely missing connector between back and seat. Our connectivity constraints lift the corresponding 2D patch (in zoom), generating a complete 3D chair. Note the distortion in the reconstructed legs. Since the scan is completely missing there, 3D lifting of the leg patch was not sufficiently constrained.

To further evaluate our method, we demonstrate our reconstruction algorithm on several objects in Figure 10. Left-to-right, are the input 2D photograph and its segmentation, input and segmented 3D scan, lifted mesh, and textured reconstruction respectively.

Figure 1 and Figure 10 (a) demonstrate two chairs with large missing parts in the back and seat. Note the lack of connectivity between scanned parts. We transfer boundaries and patches from the photograph using 3D constraints, thus completing the missing part and reconstructing a connected chair.

Figure 10 (b) shows a cart with thin wheel rims and spokes which are sparsely captured. Since missing parts are large, it is impossible to infer them with additional knowledge or priors as in our method. We lift thin spoke patches and complete wheel rims with their connectivity from 2D to 3D resulting in a complete and detailed wheel and cart reconstruction.

Figure 10 (c-e) show our technique on a set of challenging scans of mechanical objects with large holes and significant missing parts. From (c) to (e), a set of pipes which exhibit holes at important junctions and bifurcations, a car engine with missing important connectors and another T-shape pipe with holes. We infer high-level structures from the 2D photo which in turn generate shape completion and faithful reconstruction of parts and their connectivity.

In Figure 7 we show an evaluation of the reconstructed model. As can be seen from the figure, the surfaces of the reconstructed model are accurate with respect to the input scan.

In Figure 8 we evaluate the robustness of our photo-scan registration algorithm by varying the initial viewpoints of the scan. Although the initial alignment varies, our registration and reconstruction is robust to such misalignments.

Performance We have run our experiment on a PC with Windows 7, 3.1GHz processor and 64GB RAM. Table 1 summarizes the performance of our algorithm showing scan sizes and reconstruction times for all our objects. As can be seen from the table, our 2D-3D lifting process is

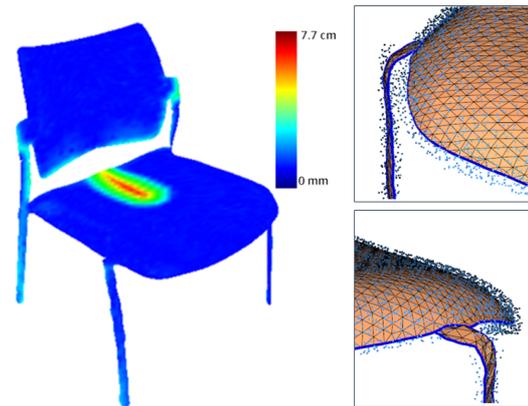


Figure 7: Evaluation of the reconstructed mesh model with respect to original 3D scan. Coloring denotes distance of 3D mesh to closest scan point. Right are zooms of the reconstructed mesh passing close by the 3D points.

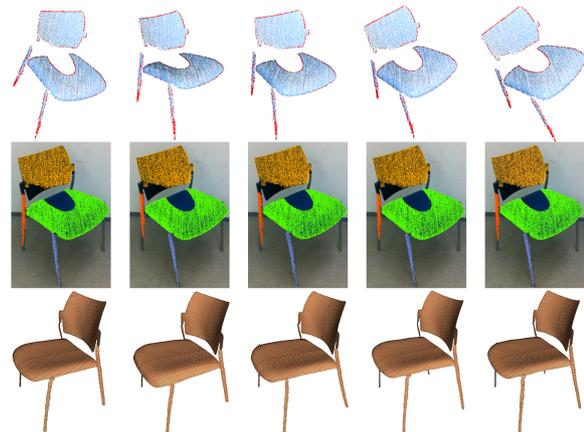


Figure 8: Robustness of our algorithm to 2D-3D misalignment. Top-to-bottom, are five cases showing initial coarse 2D-3D alignment, their fine registration, and reconstruction.

faster than the preprocessing segmentation step. The high efficiency of the 2D-3D lifting benefits from the segmentation of the two inputs. In our reconstruction pipeline, the most computation-intensive process is the TPS fitting, which involves an LU factorization of a large matrix of size $(n + d + 1) \times (n + d + 1)$, where n is the number of control points and $d = 3$ is the dimension of the problem. By segmenting the object into meaningful parts, the size of the matrix is significantly reduced, yielding efficient reconstruction. In addition to the efficiency gain, the boundary and topological constraints discovered from the segmentation of the 2D photograph enable us to recover sharp features of an object. Figure 9 shows the reconstruction of a synthetic me-

chanical part. Note how well the sharp features are preserved using the 2D image structure and segmentation.

Comparison We also conduct a comparison with state-of-the-art automatic reconstruction methods. Figure 11 shows several data set reconstructed using *Algebraic Point Set Surfaces* (APSS) [GG07] (2nd row), *Poisson Reconstruction* [KBH06] (3rd row), and our 2D-3D lifting method (last row). Since the APSS method is defined based on Moving Least Squares fitting, it can not handle severe data missing. Similarly, although [KBH06] can obtain very impressive reconstruction in smooth regions where adequate sampling of the object's surface is available, it usually connects distinct parts (e.g. the region between the legs of the chair) or fills in real holes (e.g. the windows of the car). While our method, by using additional topological constraints inferred from images, successfully completes large missing regions and generates the most faithful results.

Symmetry An efficient scan enhancement and completion tool is symmetrization, where missing data is inferred from symmetric parts in the shape. In Figure 10 (a), we pre-

Table 1: Detailed summary of input scan sizes and reconstruction times (seconds) for our experiments.

Model							
# Points	20.5K	16.5K	12K	55.2K	342.7K	140.2K	138K
Symmetry	3.2	3.7	1.5	-	28.4	-	-
Register	6	5	7	18	38	22	34
2D Segment	62	38	49	175	325	158	274
3D Segment	12	15	13	19	124	32	118
Overlay	1.9	1.4	1.5	5.9	10.2	3.1	11.1
Lifting	62	53	31	12	284	267	105
Total time	147.1	116.1	103	229.9	809.6	482.1	542.1

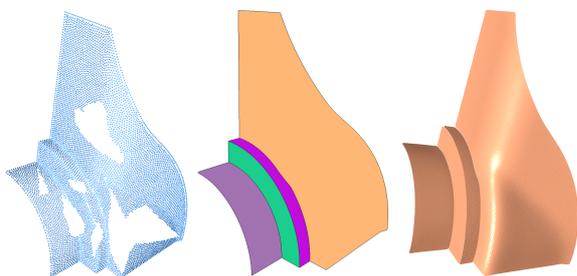


Figure 9: The reconstruction of a synthetic mechanical object consisting of a free-form surface and a few cylinders and planes. Left to right: input point cloud, segmented image, and our result.

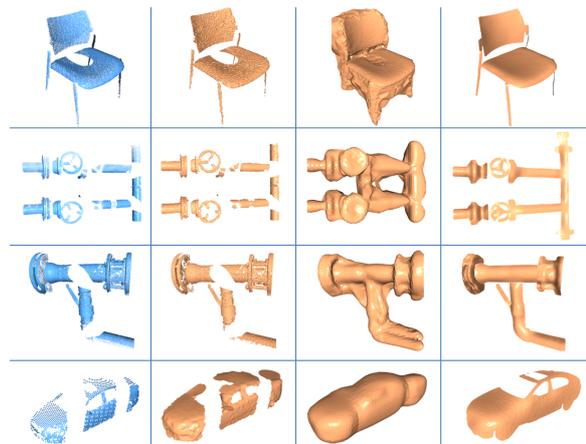


Figure 11: Comparison with state-of-the-art reconstruction methods. Left-to-right, are the input, the results of Algebraic Point Set Surfaces [GG07], Poisson Reconstruction [KBH06], and our 2D-3D lifting method.

process the scan using a similar symmetrization algorithm of Mitra et al. [MGP]. Nevertheless, the symmetrized data is still incomplete with holes in the seat and unconnected legs. Similarly, scans in Figures 1, 6 and 10 (b) were symmetrized, but holes and imperfections persist and remain noticeable.

6. Limitations and Conclusions

Lifting a 2D patch into 3D is essentially an ill-posed problem. Our multi-modal approach fuses between 2D photographs and 3D scans and generates a 2D-3D constrained lifting optimization. Nevertheless, if data is completely missing, i.e. no 3D points mapping to a 2D patch nor to its neighbors, the transformation optimization may become under constrained. For example, in Figure 6, a large part of the chair legs is missing, thus legs patch deforms during 3D lifting. Similarly, lifting may be incorrect if the 2D and 3D patch matching is incorrect. This may occur when the perspective projection yields large occlusions, thus finding a consistent 2D to 3D patch mapping is ambiguous (see Figure 2).

Due to the smooth nature of our TPS fitting, our only assumption is that the object's underlying geometry is piecewise smooth. To recover sharp features of the object, we require the object is segmented into piecewise patches in the segmentation step. By doing this, sharp edges are given as the patches boundaries and are preserved during the 2D-3D lifting step.

In this work we focused on fusing a 3D scan with a single 2D photograph. Thus, it provides with scan enhancement only from the photograph view. For example, the chair legs in Figure 1 and Figure 10 (a) are thin sheets instead of round

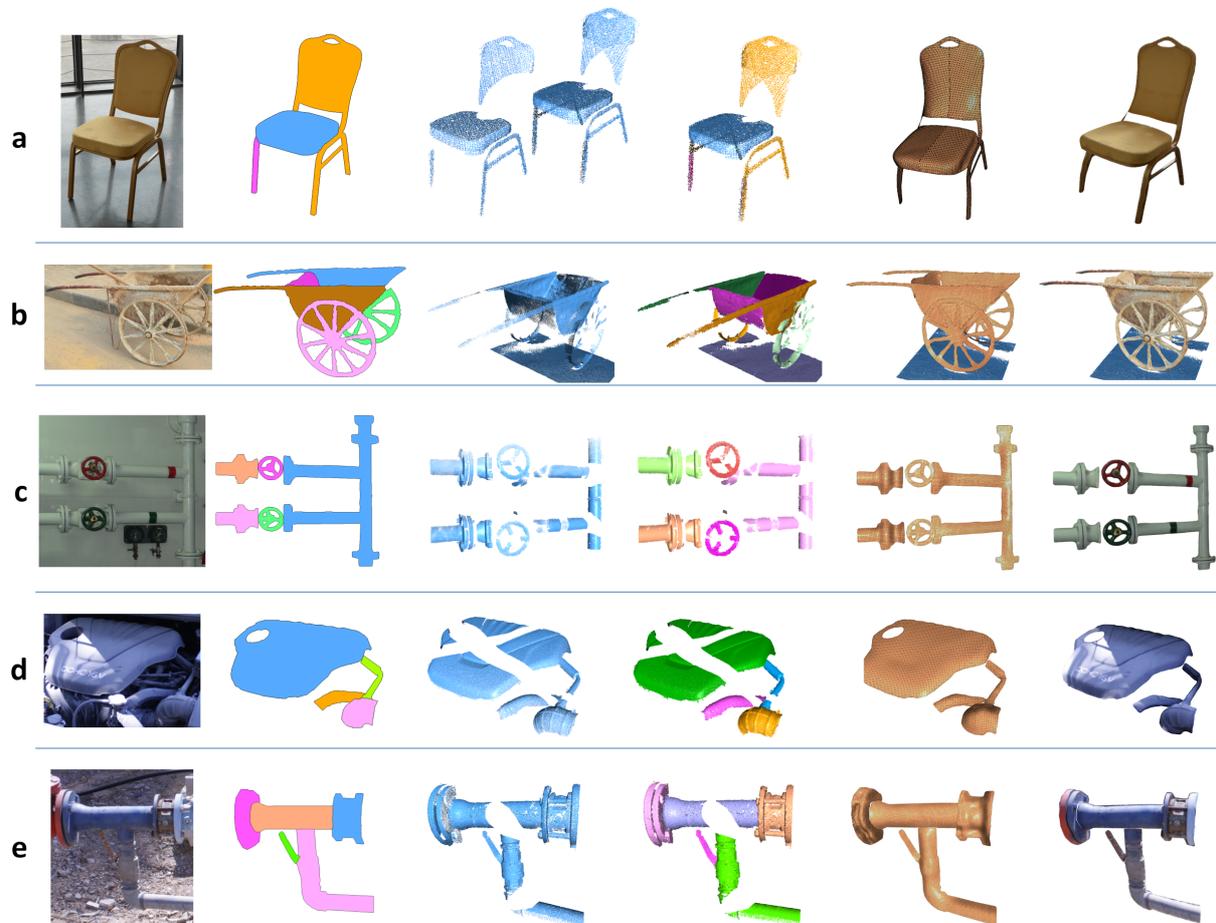


Figure 10: 2D-3D lifting reconstruction algorithm demonstrated on various objects. Left-to-right: input photograph, segmented photograph, input partial scan, segmented scan, reconstructed mesh, and textured model.

cylinders. In future work we plan to further develop the 2D-3D multi-modal framework by incorporating multiple images fusing 2D structures from different views together.

To conclude, we present an algorithm for 3D shape reconstruction from a partial 3D scan utilizing 2D structures from a corresponding photograph. We do not require strict photo-to-scan correspondence, thus they can be in arbitrarily different views. We fuse the 3D and 2D modes taking advantage of their complementary characteristics. Our key observation is that with a registration, we can transfer 2D structures into 3D in a principled way. We utilize depth cues, local and patch-wise adjacency information to constrain 2D-to-3D lifting and transfer shape structures such as geometry and connectivity (i.e. topology). We reason that 2D boundary and connectivity regularize and significantly reduce the complexity of reconstruction problem. Using our method, we were able to reconstruct a variety 3D objects from partial scans with large missing parts in a plausible manner.

Acknowledgements

We thank the anonymous reviewers and Prof. Daniel Cohen-Or, Niloy Mitra, and Peter Wonka for their valuable comments and suggestions. We also thank Ke Xie and Pinxin Long for scanning some objects as the input for our experiments. This work was supported by National Natural Science Foundation of China (61272327, 61232011, 61025012) and Shenzhen Innovation & Technology Program (CXB201104220029A, ZD201111080115A). Andrei Sharf was supported by the European FP7 and the Israel Science Foundation.

References

- [ASF*13] ARIKAN M., SCHWÄRZLER M., FLÖRY S., WIMMER M., MAIERHOFER S.: O-snap: Optimization-based snapping for modeling architecture. *ACM Trans. Graph.* (2013), 6:1–6:15. 3
- [BLN*13] BERGER M., LEVINE J. A., NONATO L. G., TAUBIN

- G., SILVA C. T.: A benchmark for surface reconstruction. *ACM Trans. Graph.* (2013), 20:1–20:17. 3
- [Boo89] BOOKSTEIN F. L.: Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on pattern analysis and machine intelligence* 11, 6 (1989), 567–585. 6
- [CKX*08] CHEN X., KANG S. B., XU Y. Q., DORSEY J., SHUM H. Y.: Sketching reality: Realistic interpretation of architectural designs. *ACM Trans. Graph.* (2008), 1–15. 3
- [DT05] DIEBEL J., THRUN S.: An application of markov random fields to range sensing. In *Proceedings of Advanced Neural Information Processing Systems* (2005), pp. 291–298. 3
- [DTM] DEBEVEC P. E., TAYLOR C. J., MALIK J.: Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach. In *ACM SIGGRAPH 1996*, pp. 11–20. 3
- [GG07] GUENNEBAUD G., GROSS M.: Algebraic point set surfaces. In *ACM Transactions on Graphics (TOG)* (2007), vol. 26, ACM, p. 23. 8
- [GSH*07] GAL R., SHAMIR A., HASSNER T., PAULY M., COHEN-OR D.: Surface reconstruction using local shape priors. *SGP 2007*, pp. 253–262. 3
- [GWM01] GUMHOLD S., WANG X., MACLEOD R.: Feature extraction from point clouds. In *Proceedings of 10th international meshing roundtable* (2001), vol. 2001. 5
- [HDD*] HOPPE H., DE ROSE T., DUCHAMP T., McDONALD J., STUETZLE W.: Surface reconstruction from unorganized points. In *ACM SIGGRAPH 1992*, pp. 71–78. 3
- [HK09] HABBECKE M., KOBELT L.: An intuitive interface for interactive high quality Image-Based modeling. *Computer Graphics Forum* 28, 7 (2009), 1765–1772. 3
- [JTC] JIANG N., TAN P., CHEONG L.-F.: Symmetric architecture modeling with a single image. In *ACM SIGGRAPH Asia 2009*. 3
- [JV11] JIAN B., VEMURI B. C.: Robust point set registration using gaussian mixture models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 33, 8 (2011), 1633–1645. 5
- [KBH06] KAZHDAN M., BOLITHO M., HOPPE H.: Poisson surface reconstruction. *SGP 2006*, pp. 61–70. 3, 8
- [KMHG13] KIM Y., MITRA N., HUANG Q., GUIBAS L. J.: Guided real-time scanning of indoor environments. *Comput. Graph. Forum (Proc. Pacific Graphic 13)* 32, 7 (2013). 2
- [KNC*] KOPF J., NEUBERT B., CHEN B., COHEN M., COHEN-OR D., DEUSSEN O., UYTENDAELE M., LISCHINSKI D.: Deep photo: model-based photograph enhancement and viewing. In *ACM SIGGRAPH Asia 2008*, pp. 116:1–116:10. 2, 3
- [LA11] LAW A. J., ALIAGA D. G.: Single viewpoint model completion of symmetric objects for digital inspection. *Comput. Vis. Image Underst.* (2011), 603–610. 2, 3
- [LA13] LAFARGE F., ALLIEZ P.: Surface reconstruction through point set structuring. In *Computer Graphics Forum* (2013), vol. 32, pp. 225–234. 3
- [LCOLTE] LIPMAN Y., COHEN-OR D., LEVIN D., TAL-EZER H.: Parameterization-free projection for geometry reconstruction. In *ACM SIGGRAPH 2007*. 3
- [LPC*] LEVOY M., PULLI K., CURLESS B., RUSINKIEWICZ S., KOLLER D., PEREIRA L., GINTZON M., ANDERSON S., DAVIS J., GINSBERG J., SHADE J., FULK D.: The Digital Michelangelo Project: 3D scanning of large statues. In *ACM SIGGRAPH 2000*, pp. 131–144. 3
- [LSH14] LIU M., SALZMANN M., HE X.: Discrete-continuous depth estimation from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2014), pp. 716–723. 3
- [LWC*11] LI Y., WU X., CHRYSANTHOU Y., SHARF A., COHEN-OR D., MITRA N. J.: Globfit: Consistently fitting primitives by discovering global relations. *ACM Transactions on Graphics* 30, 4 (2011), to appear. 3
- [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 *Proceedings of the 2011 International Conference on Computer Vision* (2011), ICCV '11, pp. 882–889. 2, 3
- [Mey92] MEYER F.: Color image segmentation. In *Image Processing and its Applications, 1992., International Conference on* (1992), IET, pp. 303–306. 4
- [MGP] MITRA N. J., GUIBAS L. J., PAULY M.: Partial and approximate symmetry detection for 3d geometry. In *ACM SIGGRAPH 2006*, pp. 560–568. 8
- [NSZ*] NAN L., SHARF A., ZHANG H., COHEN-OR D., CHEN B.: Smartboxes for interactive urban reconstruction. In *ACM SIGGRAPH 2010*, pp. 93:1–93:10. 3
- [OGG09] OZTIRELI C., GUENNEBAUD G., GROSS M.: Feature preserving point set surfaces based on non-linear kernel regression. In *Eurographics 2009* (2009), pp. 493–501. 3
- [PMG*05] PAULY M., MITRA N. J., GIESEN J., GROSS M., GUIBAS L. J.: Example-based 3d scan completion. *SGP 2005*. 3
- [RKB] ROTHER C., KOLMOGOROV V., BLAKE A.: Grabcut: interactive foreground extraction using iterated graph cuts. In *ACM SIGGRAPH 2004*, pp. 309–314. 4
- [SDK09] SCHNABEL R., DEGENER P., KLEIN R.: Completion and reconstruction with primitive shapes. *Computer Graphics Forum (Proc. of Eurographics)* 28, 2 (2009), 503–512. 3
- [SFCH] SHEN C.-H., FU H., CHEN K., HU S.-M.: Structure recovery by part assembly. *ACM SIGGRAPH Asia 2012* 31, 6, 180:1–180:11. 3
- [SHM*] SU H., HUANG Q., MITRA N. J., LI Y., GUIBAS L.: Estimating image depth using shape collections. *ACM SIGGRAPH 2014*. 3
- [SLC*08] STAMOS I., LIU L., CHEN C., WOLBERG G., YU G., ZOKAI S.: Integrating automated range registration with multiview geometry for the photorealistic modeling of large-scale scenes. *Int. J. Comput. Vision* (2008), 237–260. 3
- [SLS*] SHARF A., LEWINER T., SHKLARSKI G., TOLEDO S., COHEN-OR D.: Interactive topology-aware surface reconstruction. In *ACM SIGGRAPH 2007*. 3
- [SSS*] SINHA S. N., STEEDLY D., SZELISKI R., AGRAWALA M., POLLEFEYS M.: Interactive 3d architectural modeling from unordered photo collections. In *ACM SIGGRAPH Asia 2008*, pp. 159:1–159:10. 2, 3
- [XFZ*] XIAO J., FANG T., ZHAO P., LHUILLIER M., QUAN L.: Image-based street-side city modeling. In *ACM SIGGRAPH Asia 2009*, pp. 114:1–114:12. 3
- [ZYFY] ZHANG C., YE M., FU B., YANG R.: Data-driven flower petal modeling with botany priors. In *CVPR 2013*, pp. 636–643. 3