Fast 3D Indoor Scene Synthesis by Learning Spatial Relation Priors of Objects

Song-Hai Zhang, *Member, IEEE,* Shao-Kui Zhang, Wei-Yu Xie, Cheng-Yang Luo, Yong-Liang Yang, and Hongbo Fu

Abstract—We present a framework for fast synthesizing indoor scenes, given a room geometry and a list of objects with learnt priors. Unlike existing data-driven solutions, which often learn priors by co-occurrence analysis and statistical model fitting, our method measures the strengths of spatial relations by tests for complete spatial randomness (CSR), and learns discrete priors based on samples with the ability to accurately represent exact layout patterns. With the learnt priors, our method achieves both acceleration and plausibility by partitioning the input objects into disjoint groups, followed by layout optimization using position-based dynamics (PBD) based on the Hausdorff metric. Experiments show that our framework is capable of measuring more reasonable relations among objects and simultaneously generating varied arrangements in seconds compared with the state-of-the-art works.

Index Terms—3D Indoor Scene Synthesis, Furniture Objects Arrangement, Complete Spatial Randomness.

1 Introduction

3D indoor scene arrangement is to automatically arrange furniture objects, which benefits various applications [1], [2], [3] including video game, virtual reality, home decoration, or even creating datasets for 3D scene understanding [4]. With the emergence of various datasets for 3D indoor scenes [5], [6], [7], techniques of arranging furniture objects have shifted toward data-driven approaches [4], i.e., learning priors expressing strategies of existing layouts of furniture objects.

However, inherent difficulties of 3D indoor scene synthesis still exist in various aspects. First, it is inevitable for dealing with furniture layouts parameterized continuously or discretely, which distribute in complex high-dimensional spaces [8]. A few works (e.g., [9], [10], [11], [12]) attempt to simplify layouts into independent cliques or subsets e.g., [10], [11]. Their underlying metric largely depends on "co-occurrence", which merely counts co-existence frequencies from existing layouts. However, co-occurrence is not sufficient to fully indicate the relationship between furniture objects. For example in Figure 2, 'nightstand' often co-exists with 'chair' in one room, but they are rather independent in terms of layout arrangement. On the other hand, 'nightstand' has high dependency with 'bed' not only due to the co-existence, but also the spatial closeness and consistency across layouts. This observation motivates us to learn stronger spatial relation priors beyond co-occurrences towards more plausible arrangements.

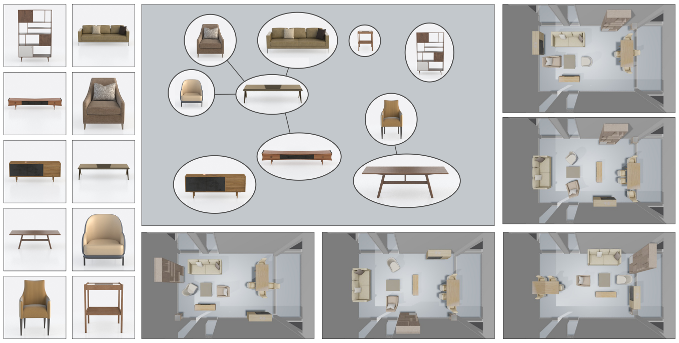
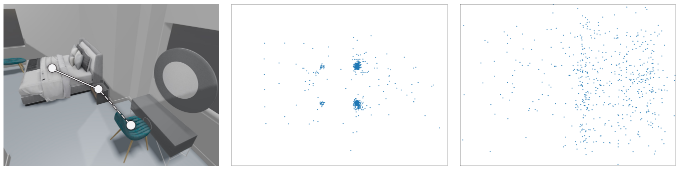


Fig. 1. Given a list of furniture objects (Left), we decompose them into disjoint groups (Top-Middle) with coherence for each individual group and freedom among groups. By incorporating discrete templates learned from datasets [5] as priors to guide syntheses, our method generates various plausible layouts in seconds.

Second, due to innumerable arrangement choices, it is hard to exhaustively list all possible spatial relations among objects [13], [14], [15], [16], [17] or to mathematically formulate unified and accurate models for them [11], [18], [19], [20]. For example, Chang et al. [13] dictate a specific set of possible relations such as "support", "right", "front", etc, which fundamentally limit the variety of possibly synthesized scenes. To model relations with multiple patterns, a common approach is to fit observed layouts with models. While allowing comprehensive exploration of a continuous layout space, the "fitted models" could potentially introduce unexpected results that are suboptimal, especially when the underlying layout patterns do not satisfy the model assumptions. Figure 3 shows less successful examples of sampling relative positions from a Gaussian Mixture Model (GMM) and a Convolutional Neural Network (CNN) [20]. We argue that when the dataset of 3D Scenes is of sufficient size, the exact cases by observing samples (without fitting continuous statistical models) already offer adequate layout variations while ensuring layout quality. This is particularly desirable for practical applications

- *Song-Hai Zhang is with the Department of Computer Science and Technology, Tsinghua University, Beijing, China and Beijing National Research Center for Information Science and Technology (BNRist), Tsinghua University, Beijing, China. E-mail: shz@tsinghua.edu.cn*
- *Shao-Kui Zhang and Wei-Yu Xie are with the Department of Computer Science and Technology, Tsinghua University, Beijing, China. E-mail: zhangsk18@mails.tsinghua.edu.cn; ervinxie@qq.com*
- *Cheng-Yang Luo is with the Department of Mathematical Sciences, Tsinghua University, Beijing, China. E-mail: luocy16@mails.tsinghua.edu.cn*
- *Yong-Liang Yang is with the Department of Computer Science, University of Bath, United Kingdom. E-mail: y.yang@cs.bath.ac.uk*
- *Hongbo Fu is with the School of Creative Media, City University of Hong Kong, Hong Kong. E-mail: hongbofu@cityu.edu.hk*

Manuscript received - -, 2020; revised - -, -.



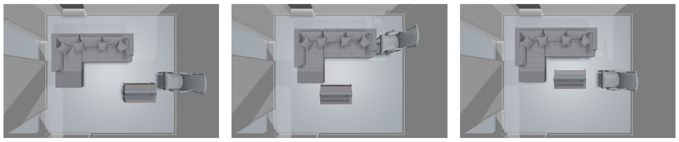
(a) A bedroom. (b) A stand & a bed. (c) A stand & a chair.

Fig. 2. Illustrating the problems of co-occurrence. With similar frequencies, two relative positions of two pairs of objects are shown in 2a. The points in 2b and 2c represent the relative positions between two given furniture objects. Axes are aligned to walls, and bed/chair is centered. In 2b, the double bed and the nightstand are obviously spatially related, while there is no obvious spatial relation between the nightstand and the chair.

where the robustness of the results is a major concern rather than the unpredictable variousness.

To address the above difficulties, we propose a method to measure the strength of spatial relations between objects by utilizing tests for complete spatial randomness (CSR) [21]. A test for CSR (Section 4) describes how likely a set of events are generated *w.r.t* a homogeneous Poisson process. Intuitively, it measures how obvious certain patterns exist in a set of points. Therefore, objects with high value of test for CSR tend to be grouped and arranged together. Objects that fail to pass tests for CSR are ignored, even if they have high co-occurrence (Section 4).

Furthermore, we present an approach for extracting representations of various shapes of layout strategies. Unlike existing frameworks [11], [19], [22], which fit continuous priors and might cause the sampling of inappropriate transformations of furniture objects, our approach first removes outliers inside datasets and then directly takes the remaining data as “discrete priors”, with each datum expressing an “exact” transformation incorporating density peak clustering (DPC) [23]. Finally, we present a framework for automatically synthesizing various arrangements of given furniture objects *w.r.t* an input room geometry, by partitioning the input objects into disjoint groups according to the learnt priors, followed by an optimization. Instead of using Markov Chain Monte Carlo (MCMC) [22], which typically takes more than thousands of iterations to converge, we optimize furniture arrangements based on the Hausdorff metric to cope with the learned discrete priors, and are able to complete the entire process in seconds.



(a) GMM (b) CNN [20] (c) Ours

Fig. 3. Fitting continuous statistical models to represent spatial relations between furniture objects has inherit difficulties. These models can potentially lead to unexpected results that are not optimal. Instead, our method can extract and reproduce exact patterns without fitting such models while ensuring synthesis quality.

In summary, our work makes the following contributions:

1) We incorporate spatial relation prior learning based on CSR and DPC, which is more effective than simply measuring co-occurrences, thus leading to plausible results consistent with common sense.
2) Our method improves the robustness and efficiency of indoor scene arrangement due to the usage of discrete and exact priors, which ensure predictable and quality object relationships, and enable efficient layout optimization based on a Hausdorff metric in seconds.

2 Related Works

TABLE 1
Qualitative characteristics of similar works to ours.

Method	Spatial Measurement	Layout Strategy
Yu et al. [22]	User Suggestions	MCMC
Qi et al. [10]	Co-Occurrence	MCMC
Wang et al. [20]	-	CNN
Wang et al. [19]	-	CNN
Ours	Tests for CSR	PBD w/ Hausdorff metric

3D Indoor Scene Synthesis aims at generating appropriate and well-aligned layouts of furniture objects for rooms. Various solutions considering different input settings and tasks have been proposed. For example, [24], [25], [26], [27] generate room layouts based on RGB-D images or 3D scans. Human language [12], [13], [14], hand-drawn sketches [18], semantic bounding boxes [28] have also been explored as additional inputs to guide scene synthesis. Table 1 lists similar works compared to ours. A full review of existing works on indoor scene synthesis is beyond the scope of this paper. Please refer to an insightful survey in [1]. Our work focuses on furniture layout synthesis within a single room. Please refer to the recent works [29], [30] and the references therein for floorplan synthesis with multiple rooms.

To synthesize a room layout, typically, two stages are required: “selecting” a list of appropriate furniture objects and “arranging” them. One characteristic that classifies different works is whether or not the two stages are coupled with each other. For example, [19], [20], [31], [32] iteratively infer the next objects to be included into rooms, i.e., placing objects depends on each pending layout. [9], [10], [22] and ours firstly create a list (graph) of objects of interest and arrange them. It is hard to compare which class of methods is better. However, making object arrangement decoupled with object selection gives flexibility to swap or combine different ways for object selection and arrangement.

As discussed in Section 1, the representations of layout strategies play an important role in 3D indoor scene synthesis. To encode prior knowledge, [15], [16], [33] attempt to quantify interior design rules, i.e., mathematically modeling how we arrange furniture objects according to designers or common senses. In contrast, our framework is data-driven due to the emerging availability of 3D indoor scene datasets, which enable various data-driven approaches. For example, Chang et al. [14] model spatial relations between objects using semantics such as “left”, “right”, “front”, etc. However, since it is difficult to enumerate all potential semantics between objects, our discrete priors are learnt to express as many exact patterns as possible according to datasets. To fit observed distributions of objects, Gaussian mixture models (GMMs) are adopted by [11], [18], [34], but [11], [18] do not have the same input to ours and the priors of [34] considers only the “XOZ” plane without the “Y” axis (height). [22], [35] model contexts for objects, e.g., average orientations and distances between objects, orientations *w.r.t* the nearest walls, etc. Furthermore, Wang el al. [19], [20] train convolutional neural

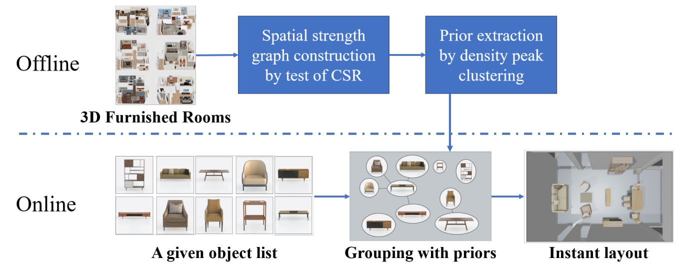
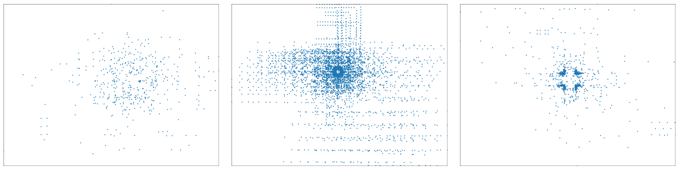

Offline
3D Furnished Rooms
Spatial strength graph construction by test of CSR
Prior extraction by density peak clustering
Online
A given object list
Grouping with priors
Instant layout

Fig. 4. The pipeline of our system.



(a) $d^{wa,ct} = 1.12$. (b) $d^{dt,ch} = 2.03$. (c) $d^{be,ni} = 2.47$.

Fig. 5. Several results of tests for CSR. 5a plots relative positions between a wardrobe cabinet and a coffee table, 5b plots relative positions between a dining table and a chair, and 5c plots relative positions between a bed and a nightstand. Axes are aligned to walls, and the former object is centered.

networks (CNNs) for placing furniture objects. Instead, we do not fit any models. Our discrete priors are from exact cases inside observed data. We assemble a subset of discrete ground truth which already includes sufficient and exact layout patterns from datasets after denoising. Graph structures are constructed by [9], [10] for encoding priors, which still resort to co-occurrence but measure spatial relations inappropriately. Based on contextual models, for optimizing transformations of furniture objects, Yu et al. [22] and Qi et al. [10] incorporate MCMC, which is experimentally verified being inefficient [33]. The reason to incorporate MCMC is due to the complicated fitted models are not differentiable, while MCMC enables updating transformations of furniture objects by proposal functions, i.e., sampling based on previous samples. The stochastic nature of the sampling strategy often leads to non-optimized samples being rejected. To ensure convergence, a large number of iterations is often needed, and each iteration usually requires costly evaluation to accept/reject samples, making the whole process computationally expensive. By incorporating the discrete priors, we measure the loss of arrangements using the Hausdorff metric, so MCMC is no longer needed, thus accelerating the entire optimization. A full discussion of MCMC is beyond the scope of this paper. Please refer to a detailed experiment in [33].

Our task partially resembles [22], but takes an automatic approach to extract constraints from existing layouts. Instead, the framework of Yu et al. [22] requires manual assignment of spatial relations between furniture objects, while learning clustered means of distances and orientations from a given dataset. Our task also partially resembles [33], but their approach is not data-driven. We are also inspired by the works of [11] and [36]. However, the former requires exemplar scenes as input, while the latter focuses on the re-arrangement of existing scenes. In contrast, we aim to learn general patterns for pairs of objects from existing layout examples for synthesizing new scenes.

Tests for Complete Spatial Randomness (CSR) is a classical topic [37]. Given a series of points distributed on a plane, a test for CSR is typically used to answer how likely the points are placed randomly. Formally, it describes how likely a set of events are generated *w.r.t* a homogeneous Poisson process (planar Poisson process). Previously, most applications of CSR are confined to ecology [38], e.g., to investigate whether or not a set of observed plants are located with patterns. Rosin [39] is probably the first to bring the concept of CSR into computer vision to handle the problem of how to detect white noises inside images. To the best of our knowledge, our work is the first to introduce tests for CSR to solve the problem of 3D indoor scene synthesis.

3 Overview

As illustrated in Figure 4, our pipeline is split into an offline stage for spatial relation prior learning, and an online stage for automatic scene synthesis based on a given list of furniture objects and the learnt priors. The pattern of spatial relations are extracted from datasets in the offline stage. We first learn a specific spatial strength graph model $\bar{G}$ indicating how objects are spatially related with each other (Section 4). In this graph, vertices represent objects, and edges are associated with weights to encode the spatial strengths between objects. This is more powerful than simply counting co-occurrence. We then extract versatile patterns of layout strategies as discrete "templates" by reducing noises within datasets such as SUNCG [5] using Density Peak Clustering [23] (Section 5). Given the learned priors, an empty room, and a set of user-specified objects, during the online stage, our method first groups spatially coherent objects into groups (e.g., a bed, a night stand and a TV Stand, as illustrated in Figure 11b). Next, we do an instant arrangement for each group by heuristically using the learned templates. Finally, we adjust the overall layout by optimizing a consistent loss function (Section 6).

Existing datasets for scene understanding and synthesis, such as SUNCG [5] and 3D Front [40], typically contain a set of furniture objects and a set of existing arrangements (rooms), and reuse each furniture object from several to thousands of arrangements with other objects. This motivated us to use such datasets to extract relations between furniture objects instead of arrangements. Therefore, to make it object-centric, we convert a given dataset into a multigraph $G = (V, E)$, which is conceptually a direct mathematical representation of the original dataset: each vertex corresponds to an object instance and each directional edge encodes the relative position and orientation between a pair of objects. More specifically, a vertex $v^i \in V$ contains a set of attributes $\{(d^{i,\omega}_{wall}, \theta^{i,\omega}_{wall}, t^{i,\omega}_{wall})|\omega = 1, 2, 3 \ldots, \Omega\}$, i.e., the row values of distances, orientations and translations of an object *w.r.t* its nearest walls. For each pair of objects, there are often multiple edges connecting them, because they may co-exist in different scenes. In the following prior learning stage, we will remove edges that suggest an implausible transformation (relative translation and rotation) between the two furniture objects and consider the remaining edges as "discrete" and "exact" priors since we do not intend to fit any statistical model.

Centering an object o_i, the k-th edge $e^{i,j,k} \in E$ from v^i to v^j is valued by a quadruple $(p^{i,j,k}_x, p^{i,j,k}_y, p^{i,j,k}_z, p^{i,j,k}_\theta)$ representing the k-th relative translation and orientation of o_j *w.r.t* o_i. We leverage $E^{i,j}$ to indicate the set of edges formed from v^i to v^j, where v^i is the corresponding vertex in V of object o_i.

As far as we know, 3D-Front [40][1] is the only suitable large-scale data set for the research of scene analysis and synthesis, besides SUNCG. 3D-Front contains 34 categories, 9,992 3D models, 70,000+ rooms, and 1,260,168 co-occurrences between furniture objects. However, most of objects have only fewer than 10 co-occurrences with another instance. This is too sparse to extract reliable patterns between two objects. On the contrary, SUNCG contains 175 categories, 2,266 models (non-furniture objects such as doors and windows are not included), 520,000+ rooms, and 54,844,805 co-occurrences. Therefore, the dataset we utilized in this paper is a combination of 3D-Front and SUNCG. More specifically, for 3D-Front, 30 categories of 9,317 objects have their corresponding categories in SUNCG. Thus, based on these categories, we first coarsely cluster objects. Then, based on visual similarities as illustrated in Figure 9, we enable relation sharing from objects of SUNCG to objects of 3D-Front. After combining them, 175 categories of 11,583 objects are achieved with 55,889,558 co-occurrences, in which only 9,317 objects from 3D-Front have geometric models. Eventually we construct G with this combined dataset. Based on G, we measure spatial relations between objects in Section 4, and learn layout priors in Section 5.

4 Spatial Strength Graph

Before actually extracting a template from datasets for each pair of objects, a question naturally arises: do we require templates for all pairs? As shown in Figure 2, the plots of relative translations of two objects with high co-occurrence could be very messy, with the transformations between them rather independent of each other. This motivates us to learn a spatial strength graph (SSG) so that a multitude of pairs of objects that have low relations of spatial strength is ignored when arranging furniture objects. This helps us synthesize more plausible scenes but also accelerates the synthesis process.

Formally, an SSG is a weighted graph defined as $\bar{G} = (\bar{V}, \bar{E})$, where $\bar{G}$ denotes an entire graph with $\bar{V} = V$ representing all objects in the dataset and $\bar{E}$ being the edges with the associated weights to encode the spatial strength between objects. Here the question becomes how to measure the weights from a large-scale dataset with highly diverse co-occurrences regarding spatial relations, leading us to assemble the aforementioned tests for CSR. There exist several methods of tests for CSR, including using the Diggle's function [21], [41], distance-based methods [37], [42], [43], angle-based method [44], [45], etc. In this paper, we follow [44] to test CSR by means of angles. If P and Q are two nearest points to point O, angle θ of point O is defined as the smaller of two possible angles between OP and OQ, clockwise and counterclockwise, and it is thus always between 0 and π. Therefore, we measure the weights of $\bar{E}$ by Equation 1, which is the "d-value" in [44] within the domain of tests for CSR [21]:

$$d = \sqrt{m} \sup |F_c(\theta) - F_e(\theta)|. \tag{1}$$

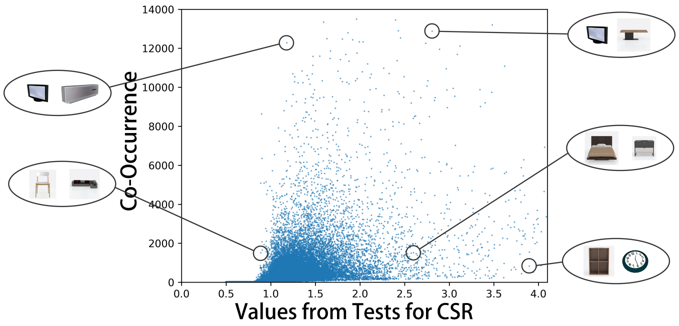


Fig. 6. The diagram plots the CSR value and co-occurrence of every pair of objects. Two objects might co-occur in many rooms, while the strength of their spatial relation could be low, vice versa. For example, the bed and the nightstand have low co-occurrence, but they are spatially related according to human intuitions.

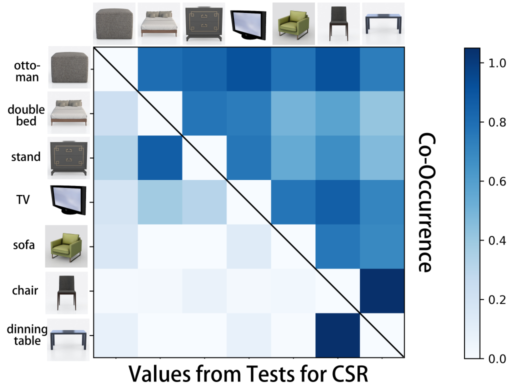


Fig. 7. A comparison between values from tests for CSR and co-occurrence. The CSR test values better reflect the dependency between objects compared with co-occurrence.

Here, F_c and F_e are respectively a cumulative distribution function (CDF) and an empirical distribution function (EDF) *w.r.t* angle θ, which is subject to uniform distribution [44]. m is the number of points formulating F_e. For each pair of objects o_i and o_j, the weights $\bar{E}^{i,j}$ are set to $d^{i,j}$ subject to random samples from $E^{i,j}$ in a ratio of 10%, as suggested in [44] and [46]. If all points are randomly distributed (i.e., following a plannar Possion distribution), the upper bound of the EDF minus the CDF of uniform distribution should be close to zero. As shown in Figure 5a, a wardrobe and a coffee table are spatially independent, so their d-value is low. Although considerable noises exist in Figure 5b, the d-value of a dining table and a chair is still reasonably high. Finally, Figure 5c shows clear patterns between a bed and a nightstand.

Figure 6 statistically suggests the differences between tests for CSR and co-occurrences, where we plot the two measurements for all pairs of objects, where pairs including an air-conditioner typically co-occur frequently but air-conditioners are placed independently to most of other furniture objects according to the common sense.

Figure 7 shows a quantitative and intuitive comparison between using co-occurrence and using tests for CSR to measure the strengths of relations between several common furniture objects. The results are normalized respectively due to different scales. The upper triangular part depicts co-occurrence and the lower part corresponds to the results from tests for CSR, which alleviate the unreasonableness caused by co-occurrence. It is obvious that placing a sofa is independent of arranging a double-bed, but they have a high frequency of co-existence in different rooms of various types. Such unreliable relations potentially confuse scene synthesis algorithms. Applying tests for CSR for them is able to decouple them spatially. It is a similar case for many other objects

[1] https://tianchi.aliyun.com/specials/promotion/alibaba-3d-scene-dataset

preferring independent layouts with most of the others, such as a white dryer, a wardrobe and a brown stand.

5 Prior Learning

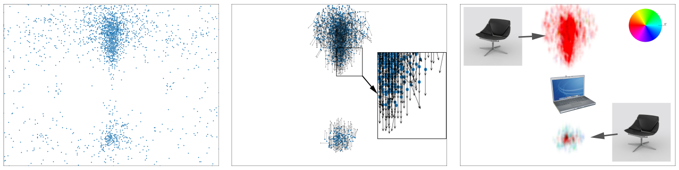
Patterns are priors suggesting how we arrange objects in real-life layouts. Figure 8c shows a pattern of a laptop *w.r.t* an office chair. Since the relative translations are incorporated, patterns can inherently avoid unreasonable situations such as collisions. However, it is obvious that we cannot adopt a unified model for all patterns, since the patterns can have arbitrary shapes. To extract arbitrarily-shaped patterns in a discrete representation, we adopt the approach in [23], which clusters vectors of dimension D, where $D >= 1$, according to ρ (Equation 2) and δ (Equation 3). The indicating function $I_{\{d \le d_c\}}$ returns 1 if $d \le d_c$ and 0 otherwise.

$$\rho_k = \sum_{k'} I_{\{d \le d_c\}}(d_{k,k'}), d_c = d_{(\eta K^2)}, \qquad (2)$$

$$\delta_k = \min_{k':\rho_k<\rho_{k'}} (d_{k,k'}). \qquad (3)$$

Given a set of edges $E^{i,j}$ from v^i to v^j in G where all relative translations, i.e., attributes of edges (Section 3), are plotted as shown in Figure 8a, we first calculate pairwise Euclidean distances $d_{k,k'}$ between all edges $e^{i,j,k} \in E^{i,j}$ using their translations $tr_k = (p_x^{i,j,k}, p_y^{i,j,k}, p_z^{i,j,k})$, i.e., $d_{k,k'} = \|tr_k - tr_{k'}\|$. For each edge $e^{i,j,k}$, ρ_k is counted as the number of other edges with their distances to it less than d_c. Taking $K = |E^{i,j}|$ edges, d_c is the ηK^2-greatest value among all pairwise distances with $\eta = 0.015$ as suggested by [23]. δ_k represents the minimal distance from a set of $e^{i,j,k'}$ with higher $\rho_{k'}$ than ρ_k. As a result, despite arbitrary shapes, merely edges with high ρ_k form a potential pattern, and each edge with high ρ_k and high δ_k indexes to a potential pattern, which is analogous to a cluster center in [23]. In contrast, noises tend to have high values of δ while their local density ρ is distinctly low. As a result, we only discard noises and keep the remaining patterns $\tilde{E}^{i,j}$, as illustrated in Figure 8b. The rest of accurate patterns form a discrete templates $\tilde{E}^{i,j}$, which are already fully usable to our framework. Although we do not fit models and we use the discrete and exact priors for scene synthesis in our framework, our learnt priors are scalable to other works for 3D indoor scene synthesis. To incorporate our model in other works, e.g., MCMC [10], [22], our priors can be easily fitted to distributions such as using non-parametric kernel density estimation based on Gaussian kernels, as shown in Figures 8c. Similar to the visualization of dense optical flows [47], we apply the system of hue, saturation and value (HSV) to represent orientations, where angles are normalized within $(0, 2\pi)$ as hue, probability densities are represented as saturation, and values are all set to 1. Figure 10 shows some other representative results of learnt priors. Since height differences for most objects do not vary significantly, we plot the two channels $(p_x^{i,j,k}, p_z^{i,j,k})$ and orientations $p_\theta^{i,j,k}$ to make the visualizations of learnt priors more intuitive.

We also perform similar prior learning tasks for individual objects with regard to their nearest walls where $d_{k,k'}$ becomes the differences of scalars. In doing so, we keep the values t_w^k and θ_w^k with both high values of ρ_k and δ_k, where t_w^k and θ_w^k are respectively plausible distances and orientations to the nearest walls of furniture object k. Consequently, each furniture object is assigned a set of t_w^k and a set of θ_w^k plausibly as attributes to its corresponding vertex in G.



(a) Input. (b) De-noised. (c) Fitted.

Fig. 8. The overall process of prior learning. (a) is the input with considerable noises. (b) is the de-noised result, which is readily to use in our framework. (c) depicts the further generalization of our templates into fitted models, which are applicable for other frameworks such as MCMC. Different colors in the HSV color space represent different orientations of objects, as shown in the inset.

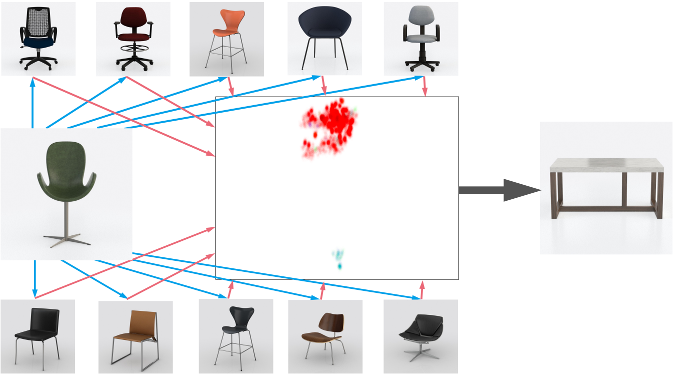


Fig. 9. Reusing existing templates for new objects of similar geometry. Given a previously unseen office chair (Left), we achieve the layout strategy of it *w.r.t* a desk (Right) by merging templates of objects geometrically similar to the chair (Top and Bottom).

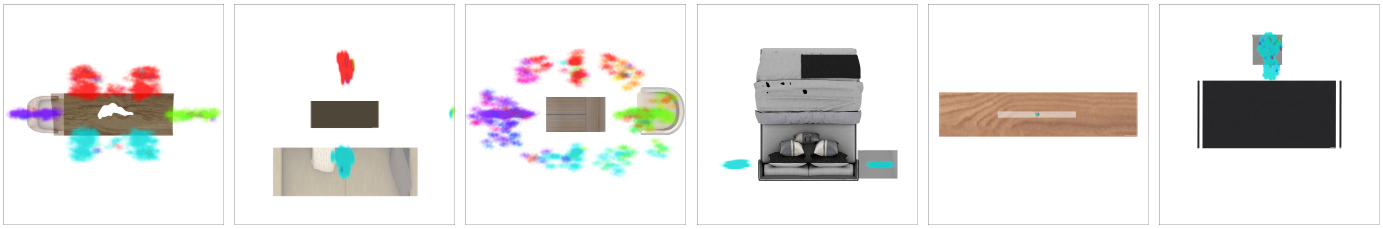
Next, we further generalize our templates to make them reusable and extensible. We observed that objects with the same semantics and similar geometries share layout strategies. As shown in Figure 9, given a new object without the corresponding priors learnt from our datasets, we find its similar models by comparing 3D shapes of models using [48], which uses the shape edit distance $s_{shed}^k \in [0, 1]$ to measure the degree of similarity, where $s_{shed}^k = 0$ indicates two identical models. We select the top-K results and take the union of the K templates as the template for the new object.

6 Scene Synthesis

In this section, we incorporate the learnt SSG and priors to synthesize room layouts. Our synthesis process is a two-step approach: a heuristic arrangement, followed by an optimization. Given a set of input objects $\hat{O}$, we first decompose them into several groups according to the SSG, and arrange objects within each group, where relative transformations are immediately indexed by the templates. Finally, we apply a global optimization to satisfy layout strategies of objects in $\hat{O}$. Note that our framework is capable of expansion by easily incorporating methods of object selections such as [10], [49], [50] or user suggestions [22], [33].

6.1 Heuristic Layouts with Formulated Groups

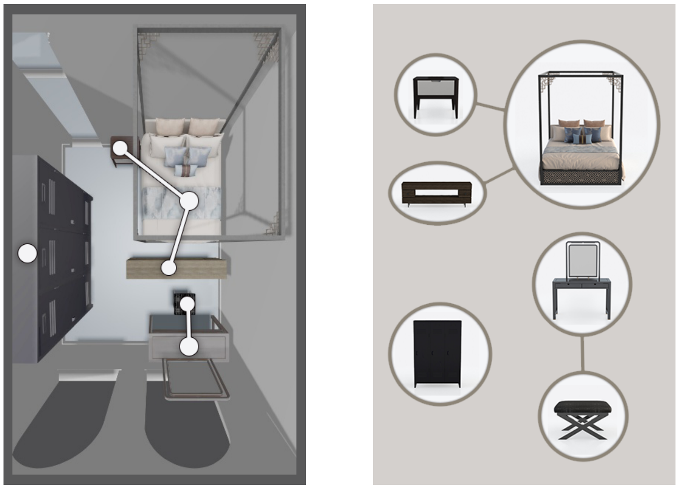
We first construct an unweighted graph, whose vertices correspond to input objects $\hat{O}$. This graph is described by an adjacency matrix M_{adj}, whose entries are determined by $\bar{G}$ in Section 4. More specifically, if d-value $d^{u,v} \ge \epsilon$, where $d^{u,v}$ is the



(a) DiningTable-Chair (b) CoffeeTable-Sofa (c) CoffeeTable-Sofa (d) DoubleBed-Stand (e) TVStand-TV (f) Desk-Chair

Fig. 10. Several representative results of learnt priors. The color coding for orientation is the same as Fig. 8.

result of a test for CSR, then we set $M^{u,v}_{adj} = 1$, where ϵ is typically equal to 1.628 as suggested in [44]. After constructing M_{adj}, we iteratively create disjoint groups $g \in Gr$ of objects by finding connected components of the graph represented by M_{adj}. Figure 11 shows examples of resulting groups. It is common to see a group containing only one object, such as wardrobe, cabinet or shelf, since their placement usually does not require the consideration of other objects. Such single-object groups greatly ease the subsequent optimization process.



(a) Marked relations. (b) Disjoint graph.

Fig. 11. Formulating functionally coherent groups of objects using the tests for CSR.

Based on a given room shape (including the position of doors and windows), partitioned groups, and learnt templates, we then generate proposals for pending scenes, i.e., objects are immediately placed and oriented *w.r.t* their groups and walls. Intuitively, we heuristically initialize the scene using the learnt templates instead of totally randomizing it. The discrete priors are considered multinomial distributions where each sample directly suggests an exact transformation between two objects. However, directly sampling the discrete priors only guarantees plausible transformations for each pair of furniture objects while relative transformations among furniture objects require the further optimization in the following subsection.

For each group $g \in Gr$, layouts of g are heuristically generated by sampling a posterior probability distribution $\Psi_{G|\tilde{E}}(g)$ expressed in Equation 4, given templates $\tilde{E}$ (Section 5).

$$\Psi_{G|\tilde{E}}(g) = \frac{\alpha(g) \cdot \Phi_{\tilde{E}|G=g}(\tilde{E}^{\mu})}{\int \alpha(g) \cdot \Phi_{\tilde{E}|G=g}(\tilde{E}^{\mu}) dg}, \tag{4}$$

$$= \frac{\alpha(g) \cdot \sum_{\mu} \prod_{\tau \in g} \phi_{\tilde{E}^{\mu}|T=\tau}(\tilde{E}^{\mu,t}, \tau^{\mu})}{\int \alpha(g) \cdot \Phi_{\tilde{E}|G=g}(\tilde{E}^{\mu}) dg}, \tag{5}$$

where $\alpha(g)$ denotes the probability of each object $\tau \in g$ being the dominant object τ^{μ} in g. Let $deg(\tau)$ denote the degree of τ *w.r.t* M_{adj}, and is essentially the number of objects connected with it according to the tests for CSR (Section 4), and $dmax = \max_{\tau \in g} deg(\tau)$. The likelihood $\phi_{\tilde{E}^{\mu}|T=\tau}(\cdot)$ is a multinomial distribution formed by the given template $\tilde{E}^{\mu,t}$ of τ *w.r.t* τ^{μ}, while it is equal to a constant when $\tau = \tau^{\mu}$.

$$\alpha(g) = \begin{cases} \frac{1}{|\{\tau|\tau \in g, deg(\tau)=dmax\}|} & \text{, if } deg(\tau^{\mu}) = dmax \\ 0 & \text{, otherwise} \end{cases}, \tag{6}$$

When sampling $\Psi_{G|\Theta=\hat{O}}$, we first randomly decide τ^{μ} of g. Equation 5 implies that $\{\phi_{\tilde{E}^{\mu}|T=\tau}(\cdot)|\tau \in g\}$ are independent of each other, so the transformations of objects are sampled according to their own templates, respectively. In practice, if an object has a relatively low d-value to τ^{μ}, we further decompose the group and assign a new dominant object to it. In some cases, this heuristic strategy could sample a sufficiently plausible layout even without a further optimization. However, the heuristic strategy may still results in unreasonable conditions such as collision between groups, objects out of room boundaries, etc. Next we show how we adjust objects so that a plausible layout of objects is eventually presented.

6.2 Template Matching

After the heuristic layout, we do template matching to optimize the placement of furniture objects and thus make their arrangement more plausible. As discussed in the previous subsection, heuristic layout is a way to initialize scenes considering merely transformations between furniture objects. In this subsection, we globally optimize entire rooms to achieve more plausible transformations among them.

Equation 7 mathematically formalizes template matching, where we are trying to minimize the summation of the Hausdorff distances d_H between all objects *w.r.t* their templates. X^i indexes the transformation of object o^i and $\tilde{E}$ is a set of sampled transformations in Section 5.

$$X^* = \arg\min_X L(X, \tilde{E}) \tag{7}$$

$$= \arg\min_X \sum_{i,j} M^{i,j}_{adj} d_H(X^i, \tilde{E}^{i,j}) + Col(X, r), \tag{8}$$

d_H is a Hausdorff metric between an element to a set of transformations, derived by the distance function d_h under the space of translation and rotation. The reason for assembling the Hausdorff distance is that it directly tackles samples instead of distributions. As explained previously, it is unlikely to mathematically express a unified distribution to model arbitrary layout patterns. In contrast, if we could extract samples of arbitrary shape, the Hausdorff metric enables pipelines to skip model fitting and to optimize directly using refined samples.

$$d_H(x,S) = \min_{v \in S} d_h(x,v), \tag{9}$$

$$\begin{aligned} d_h(x,s) &= \left\| x_p - v_p \right\| + \exp(ori(x_\theta, v_\theta)), & (10)\\ ori(\theta,\theta') &= \min(2\pi - \left|\theta - \theta'\right|, \left|\theta - \theta'\right|), & (11) \end{aligned}$$

Equation 12 represents the artifacts among objects and between objects and walls, where $p(\chi, k)$ returns the k-th corner position of the rotated bounding box of furniture object χ. Ideally, if there is no collision and no object out of boundary, $Col(X,r)$ should be equal to 0.

$$\begin{aligned} Col(X,r) &= Col_{wall}(X,r) + Col_{obj}(X) \\ &= \sum_{i,k} \prod_{r} tR(p(X_i,k), p(R,r), p(R,r+1)) \\ &\quad + \sum_{i,k,j} \prod_{l} tL(p(X_i,k), p(X_j,l), p(X_j,l+1)). \end{aligned} \tag{12}$$

Col_{wall} measures whether or not objects are out of walls, whilst Col_{obj} calculates overlaps among objects. Truncated by $tR(\cdot)$ and $tL(\cdot)$, $\gamma(\cdot)$ represents the "to-left" test of computational geometry [51], such as the utilization in [52]. In addition to given objects, we place doors and windows with the fixed transformations to avoid blocking them.

$$\begin{aligned} tR(p_1,p_2,p_3) &= \max(-\gamma(p_1,p_2,p_3), 0), & (13)\\ tL(p_1,p_2,p_3) &= \max(\gamma(p_1,p_2,p_3), 0). & (14) \end{aligned}$$

Since the underlying metrics are factorized as quadratic terms, we optimize Equation 9 by utilizing position-based dynamics (PBD) [53], which is also detailed in [33]. Incorporating heuristic approaches in section 6.1, the synthesis requires 10 iterations to converge on average after heuristic attempts.

7 EXPERIMENTS

In this section, we conduct several experiments including comparisons to the state-of-the-art techniques to verify the effectiveness of our framework. Formulating functional groups using CSR enables us to generate hybrid rooms without manually providing a predefined room type at the beginning of synthesis [19], [20]. Figure 12 shows various synthesized results. Please find more results in the supplementary materials.

7.1 Tests for CSR

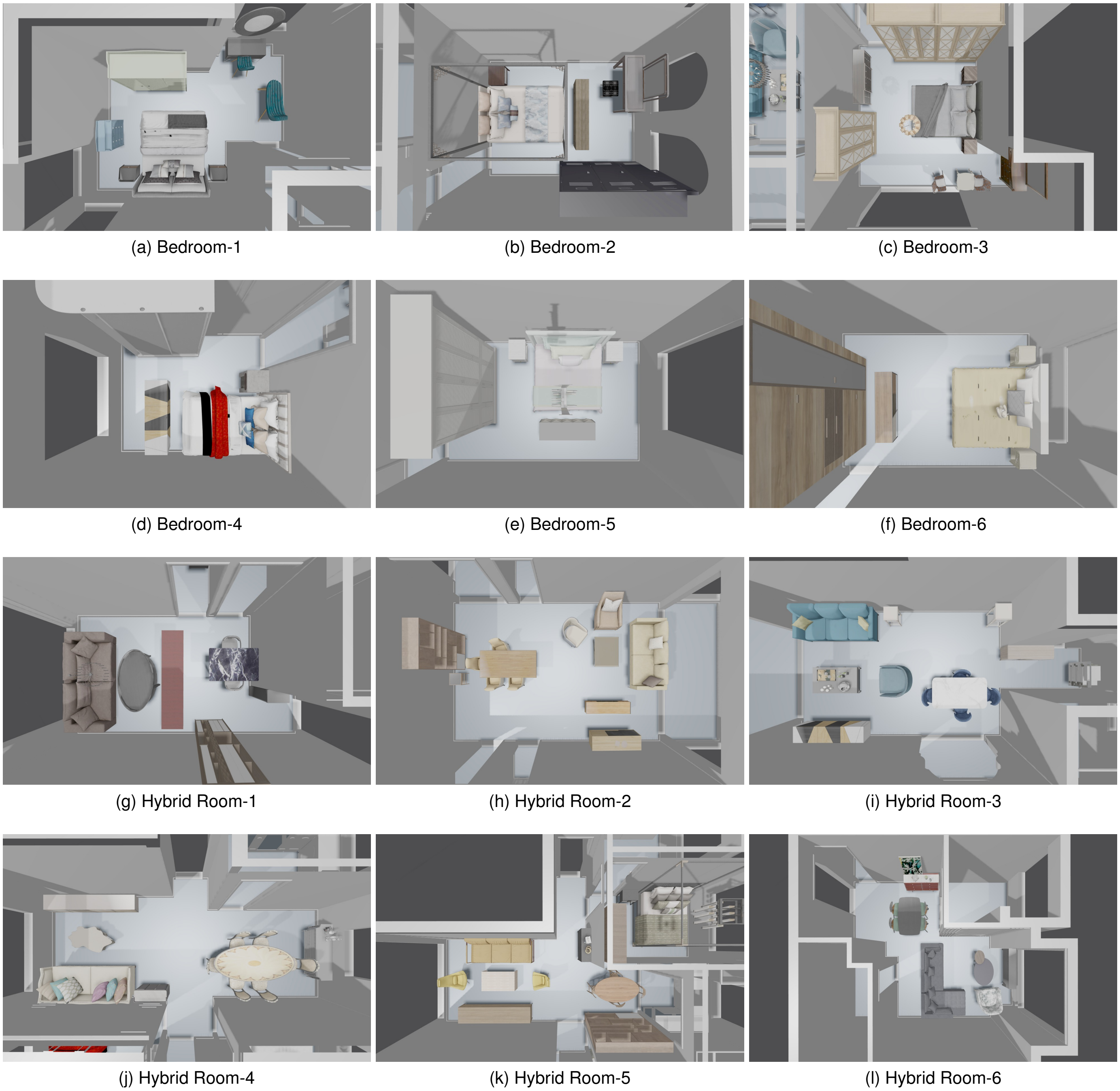
We conducted a user study to measure how tests for CSR are consistent with the intuitions of humans. We sorted pairwise relations by tests for CSR and co-occurrence (COO), respectively. For each sorted list of pairs, from their respective highest values, we systematically sample 500 templates at a fixed interval in order to achieve a set of templates with their values of COO and CSR in complete ranges that the dataset can derive. Typically, for SUNCG [5], we choose the interval $int = 120$. Then four subjects were invited to judge whether or not two presented sets of templates were consistent with real-life layout strategies. All subjects are university students with the typical common sense of arranging furniture. Note that "common sense" refers to daily layouts that are commonly seen, instead of professional interior design. The presented templates are rendered according to Figure 10, where a major furniture object is rendered from a top view and placed in the center of each image. The participants were told to decide if possible transformations of another secondary object in the image could happen in real life. For example, given a coffee table, should we place another chair in the suggested positions and orientations? If the participants are confused about several templates, secondary objects will be also rendered as shown in the supplementary materials. Table 2 lists the proportions of reasonable templates suggested by all the participants, where pairwise relations are classified according to their room types (e.g., "double-bed & night-stand" belongs to the Bedroom) since we want to show the accuracy of tests for CSR in different room types. The results suggest that the values generated by co-occurrence contain more pairs that are rather spatially independent.

7.2 Efficiency

Our work is able to synthesize scenes efficiently due to the usage of the Hausdorff metric and position-based dynamics [53], which is verified [33] to be faster than using MCMC. In this section, we conduct an experiment to show the achieved performance gain. We compared ours with two state-of-art frameworks [10], [22], since they have the same input and output as ours when arranging furniture objects, i.e., to arrange a set of furniture objects in a specified room. Note that "same output" means results only introduce transformations to furniture objects instead of brand new instances. Weiss et al [33] also have the same input and output to ours. However, our work is different from [33] since ours is data-driven and does not require user-specified constraints for each synthesis.

Statistically, arrangements are not guaranteed to be identical with each other. Consequently, since this section compares efficiency, we focus on the time-consumption while arrangements are merely considered "done" or "not done". [19] is also a state-of-art work, but as discussed in Section 2, their object selection and arrangement are coupled with each other. Thus, we will compare ours with [19] in Section 7.3.

All the methods are used to synthesize the examples chosen from Figure 12. For fair comparison, we do heuristic arrangement for both [10] and [22] to speed up their work. The time costs are shown in Table 3, where the values with "greater-than signs" denote examples requiring more than $20,000$ iterations. According to our experiments, the reason why MCMC is slow is three-fold. Firstly, each proposal move of MCMC is randomly performed. It could help to escape from local minima, but might also move away from reasonable results. Hence the number of iterations is usually large. Secondly, MCMC requires to evaluate a costly objective function to judge whether a proposal can be accepted or not. Thus even rejecting a proposal is also expensive. Thirdly, it is hard to find a good termination condition for MCMC in practice. Either simple thresholding or no further decay of the loss function cannot guarantee a good layout. In contrast, our discrete priors suggest more reasonable proposals for each iteration, and thus our method is more efficient.



(a) Bedroom-1 (b) Bedroom-2 (c) Bedroom-3

(d) Bedroom-4 (e) Bedroom-5 (f) Bedroom-6

(g) Hybrid Room-1 (h) Hybrid Room-2 (i) Hybrid Room-3

(j) Hybrid Room-4 (k) Hybrid Room-5 (l) Hybrid Room-6

Fig. 12. Examples of various synthesized results.

TABLE 2
User study: evaluations of tests for CSR and co-occurrence. Each number represents a proportion of a plausible prior.

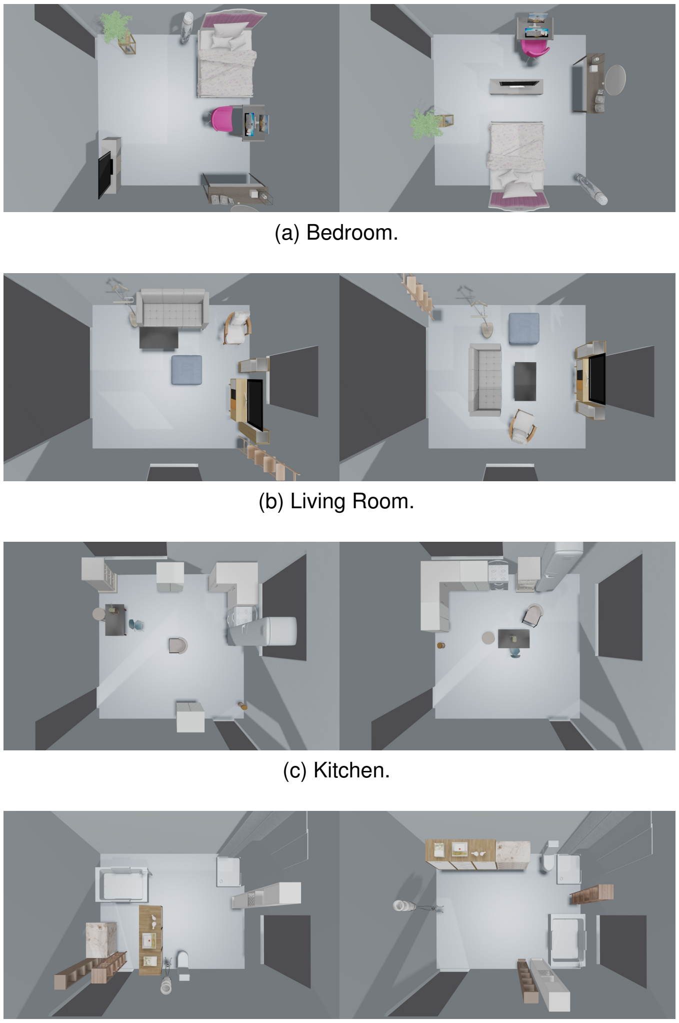
Metric	Bedroom	Living Room	Bathroom	Dining Room	Balcony	Hall	Garage	Total
Tests for CSR	93.31%	85.47%	96.67%	92.42%	86.36%	89.47%	76.17%	88.55%
Co-occurrence	32.26%	43.81%	86.67%	45.76%	23.08%	38.46%	36.84%	43.53%

7.3 Aesthetic and Plausibility

In this subsection, we evaluate the aesthetic and plausibility of the results generated by our method. Two experiments have been conducted to demonstrate the strength of ours compared with the ground truth and PlanIT [19]. First, we re-arrange objects from the ground-truth, i.e., the scenes originated from the dataset [5] and compare our results with the results of the original datasets. Since the evaluations of 3D indoor scenes are subjective, we conduct a perceptual study to analyze them. In the first experiment, 97 subjects were invited from universities and the society. Subjects are all elder than 18 so that they have the necessary appreciation of beauty for room layouts.

TABLE 3
Time consumption (in seconds) of different methods for synthesizing the scenes similar to those in Figure 12.

	# of Objects	Yu et al. [22]	Qi et al. [10]	Ours
Bedroom	9	>299.27s	229.76s	0.28s
Living Room	25	>2135.30s	1790.65s	1.88s
Bathroom	8	>313.54s	216.20s	0.29s
Hybrid-1	13	>714.60s	>481.35s	0.64s
Hybrid-2	35	>2313.38s	>1667.03s	2.31s
Hybrid-3	28	>1351.15s	1122.63s	1.24s



(a) Bedroom.

(b) Living Room.

(c) Kitchen.

(d) Bathroom.

Fig. 13. Comparisons between PlanIT [19] (Left) and our method (Right).

We ask each subject to grade each presented room layout. We show each subject in total 20 rooms, and she/he would see several different layouts of a same room but without knowing which layouts are generated. They rank each layout in the range from level-1 (poor) to level-5 (perfect). As listed in Table 4, the scores of our results and the ground-truth are comparable, indicating the aesthetic and plausibility of our results.

Second, we compare our method with PlanIT [19], a state-of-the-art method for indoor scene synthesis. See the visual comparisons in Figure 13. Similarly, we run a perceptive study to evaluate the results quantitatively and another 49 subjects are invited from the society similar to the first experiment. However, the comparison is inherently difficult. As discussed in Section 2, PlanIT couples the object selection task with the arrangement task. Consequently, to make the same input and output of two frameworks, we take the object selection results by PlanIT as input and re-arrange them using our method. Finally, given 20 rooms, we conduct the second perceptive study similar to the first one. As shown in Table 4, our rearranged results (the last row) receive consistently higher scores than those by PlanIT (the second last row). Note that a fair comparison of computational time with PlanIT would be difficult since PlanIT performs furniture object selection and arrangement together using neural networks, while our work requires furniture objects to be given. The running time of PlanIT depends on how crowded the rooms. It typically requires at least one minute for a single layout generation.

Although deep learning based approaches have demonstrated convincing performance on many problems, they inherently strive to learn the mapping/distribution from training sets, and thus the performance largely depends on how the data is prepared. For PlanIT, the relation graph of its training set is heuristically derived from the SUNCG dataset in the sense that several handcrafted rules with thresholds are defined to extract 'support edges', 'spatial edges', and 'superstructures' (cf. Section 4.3 in [19]). Moreover, to make its neural network model effective on the relation graph, PlanIT prunes a great number of 'insignificant' edges, and keeps only those with strong object co-occurrences (cf. Appendix A.2 in [19]). Hence the resulting training set mainly reflects co-occurrences and some other object-wall relationships, so as the trained neural network. On the other hand, our method explicitly measures the strengths of spatial relations based on CSR tests. This allows us to directly integrate discrete yet more accurate spatial relations into the room layouts, instead of using the relations implicitly learned by a neural network.

8 Conclusion

In this paper, we presented a framework for 3D indoor scene synthesis based on the analysis of patterns. Instead of using co-occurrence, we first incorporate tests for CSR to measure more plausible spatial relations between furniture objects. The state-of-the-art frameworks for 3D indoor scene synthesis typically fit models for representing how to arrange furniture objects, and depend on sampling that causes implausible scenes. To alleviate this, we first learn priors discretely by assembling density peak clustering [23]. The resulting priors are essentially a subset of original data so that they express exact transformations between objects. Our discrete priors subsequently enable the Hausdorff metric without resorting to MCMC, thus accelerating it. In the experiments, we verify the effectiveness and efficiency of our framework.

This work suffers from at least the following limitations. Firstly, our way to learn priors is sensitive to the size and density of datasets. An ideal dataset should be both sufficiently large and dense, where "large" refers to datasets containing a considerable number (at least at the scale of thousands) of rooms such as SUNCG [5], and "dense" refers to each object being used by hundreds of layouts instead of one or two. This is because density peak clustering [23] requires sufficient data in able to detect noises. SUNCG [5] is a sufficiently practical dataset for our method, but it still contains the aforementioned cases as shown in Figure 14. which exhibits the so-called "long-tailed distribution". Tests for CSR also suffer from the long-tail distribution problem, since we can never do CSR tests using only one or two pieces

TABLE 4
User study: aesthetics, where Ours-GT uses our framework to re-arrange furniture objects in the ground truth (SUNCG) [5] and Ours-RE uses our framework to arrange furniture objects from PlanIT [19].

Methods	Bedroom	Living Room	Bathroom	Dining Room	Balcony	Hall	Garage	**Hybrid Room**	Total
Ground Truth	2.911	3.422	3.156	3.589	3.378	2.878	3.511	3.367	3.276
Ours-GT	2.944	3.292	2.989	3.344	3.344	3.061	3.256	3.317	3.194
PlanIT [19]	2.884	2.859	2.739	2.78	-	-	-	-	2.834
Ours-RE	3.396	3.541	3.506	3.559	-	-	-	-	3.522

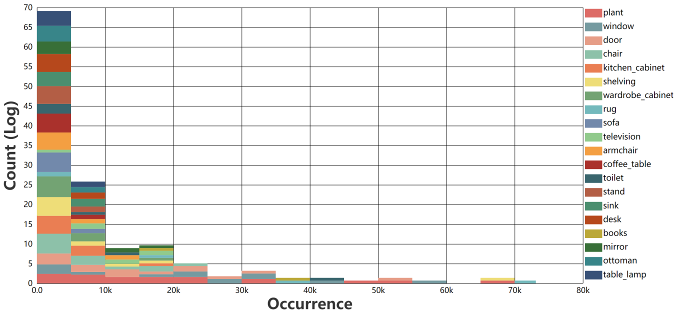


Fig. 14. A histogram shows the long-tailed distribution of objects in SUNCG [5]. It plots top-20 categories in SUNCG, where few instances are used frequently. For example, most of furniture instances are arranged in less than 5000 scenes or even less than 100 scenes, while very few instances such as instances of category rug, shelving or plant occur more than $65,000$ times.

of data. We have attempted to solve the long-tail distribution problem by clustering furniture objects using shape similarities. Nevertheless, shape similarity does not guarantee us to find the most "exactly" similar object. As discussed by Huang et al. [54], measuring the shape similarity is even subjective. Furthermore, the way to learn priors is sometimes "gullible" to datasets. For example, intuitively an office chair is spatially independent of a wardrobe. However, if in an entire dataset the chair is relatively transformed to the wardrobe identically in all rooms, the test for CSR of them could still pass. Note that a real-world dataset is also applicable to our method if its size and density are reasonable to ensure robust prior learning and noise removal. 3D real-world datasets, e.g., SceneNN and ScanNet, still cannot be used in our method, because most of the objects are incomplete or with redundant parts in geometry, leading to inaccurate positions and bounding boxes, and difficulty in shape similarity to share priors. In addition, acquiring labeled data from real-world is extremely expensive and current datasets are far below for prior extraction. For example, SceneNN/ScanNet contains 100/1,513 scenes with 1,482/36,213 objects, and co-occurrences are very limited, and more importantly there is no label of room belonging.

In the future, we are interested in performing finer comparisons of 3D shapes for generalizing our templates (e.g., by adopting 3DMatch [55]). Recently, improvements for density peak clustering are also available [56], [57] for better parameter selection and non-central node allocation. We hope that our pipeline, learnt models and synthesized layouts can contribute to automatic room layouts as well as associated domains such as scene understanding [58]. Besides, the extracted spatial relation priors can also be potentially used for interactive scene modeling tasks such as a suggestive 3D scene modeling interface [59].

Acknowledgments

The authors would like to thank all reviewers for their thoughtful comments. This work was supported by the National Key Technology R&D Program (Project Number 2017YFB1002604), the National Natural Science Foundation of China (Project Numbers 61521002, 61772298), Research Grant of Beijing Higher Institution Engineering Research Center, and Tsinghua–Tencent Joint Laboratory for Internet Innovation Technology. Yong-Liang Yang was partly supported by RCUK grant CAMERA (EP/M023281/1, EP/T014865/1).

References

[1] S.-H. Zhang, S.-K. Zhang, Y. Liang, and P. Hall, "A survey of 3d indoor scene synthesis," *Journal of Computer Science and Technology*, vol. 34, no. 3, p. 594, 2019.

[2] T. Germer and M. Schwarz, "Procedural arrangement of furniture for real-time walkthroughs," in *Computer Graphics Forum*, vol. 28, no. 8. Wiley Online Library, 2009, pp. 2068–2078.

[3] G. H. Lyons, *Ten Common Home Decorating Mistakes & How to Avoid Them*. Blue Sage Press, 2008.

[4] A. Handa, V. Patraucean, V. Badrinarayanan, S. Stent, and R. Cipolla, "Understanding real world indoor scenes with synthetic data," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4077–4085.

[5] S. Song, F. Yu, A. Zeng, A. X. Chang, M. Savva, and T. Funkhouser, "Semantic scene completion from a single depth image," *Proceedings of 30th IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[6] W. Li, S. Saeedi, J. McCormac, R. Clark, D. Tzoumanikas, Q. Ye, Y. Huang, R. Tang, and S. Leutenegger, "Interiornet: Mega-scale multi-sensor photo-realistic indoor scenes dataset," in *British Machine Vision Conference (BMVC)*, 2018.

[7] J. Zheng, J. Zhang, J. Li, R. Tang, S. Gao, and Z. Zhou, "Structured3d: A large photo-realistic dataset for structured 3d modeling," *CoRR*, vol. abs/1908.00222, 2019.

[8] Y. Li, J. Zhang, Y. Cheng, K. Huang, and T. Tan, "Df 2 net: Discriminative feature learning and fusion network for rgb-d indoor scene classification," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[9] Q. Fu, X. Chen, X. Wang, S. Wen, B. Zhou, and H. Fu, "Adaptive synthesis of indoor scenes via activity-associated object relation graphs," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 6, p. 201, 2017.

[10] S. Qi, Y. Zhu, S. Huang, C. Jiang, and S.-C. Zhu, "Human-centric indoor scene synthesis using stochastic grammar," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5899–5908.

[11] M. Fisher, D. Ritchie, M. Savva, T. Funkhouser, and P. Hanrahan, "Example-based synthesis of 3d object arrangements," *ACM Transactions on Graphics (TOG)*, vol. 31, no. 6, p. 135, 2012.

[12] R. Ma, A. G. Patil, M. Fisher, M. Li, S. Pirk, B.-S. Hua, S.-K. Yeung, X. Tong, L. Guibas, and H. Zhang, "Language-driven synthesis of 3d scenes from scene databases," in *SIGGRAPH Asia 2018 Technical Papers*. ACM, 2018, p. 212.

[13] A. Chang, W. Monroe, M. Savva, C. Potts, and C. D. Manning, "Text to 3d scene generation with rich lexical grounding," *arXiv preprint arXiv:1505.06289*, 2015.

[14] A. Chang, M. Savva, and C. D. Manning, "Learning spatial knowledge for text to 3d scene generation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 2028–2038.

[15] P. Merrell, E. Schkufza, Z. Li, M. Agrawala, and V. Koltun, "Interactive furniture layout using interior design guidelines," in *ACM transactions on graphics (TOG)*, vol. 30, no. 4. ACM, 2011, p. 87.

[16] Y.-T. Yeh, L. Yang, M. Watson, N. D. Goodman, and P. Hanrahan, "Synthesizing open worlds with constraints using locally annealed reversible jump mcmc," *ACM Transactions on Graphics (TOG)*, vol. 31, no. 4, p. 56, 2012.

[17] M. Li, A. G. Patil, K. Xu, S. Chaudhuri, O. Khan, A. Shamir, C. Tu, B. Chen, D. Cohen-Or, and H. Zhang, "Grains: Generative recursive autoencoders for indoor scenes," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 2, p. 12, 2019.

[18] K. Xu, K. Chen, H. Fu, W.-L. Sun, and S.-M. Hu, "Sketch2scene: sketch-based co-retrieval and co-placement of 3d models," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 4, p. 123, 2013.

[19] K. Wang, Y.-A. Lin, B. Weissmann, M. Savva, A. X. Chang, and D. Ritchie, "Planit: Planning and instantiating indoor scenes with relation graph and spatial prior networks," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 4, p. 132, 2019.

[20] K. Wang, M. Savva, A. X. Chang, and D. Ritchie, "Deep convolutional priors for indoor scene synthesis," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 4, p. 70, 2018.

[21] P. J. Diggle, "On parameter estimation and goodness-of-fit testing for spatial point patterns," *Biometrics*, pp. 87–101, 1979.

[22] L.-F. Yu, S. K. Yeung, C.-K. Tang, D. Terzopoulos, T. F. Chan, and S. Osher, "Make it home: automatic optimization of furniture arrangement." *ACM Trans. Graph.*, vol. 30, no. 4, p. 86, 2011.

[23] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.

[24] A. Avetisyan, M. Dahnert, A. Dai, M. Savva, A. X. Chang, and M. Nießner, "Scan2cad: Learning cad model alignment in rgb-d scans," *arXiv preprint arXiv:1811.11187*, 2018.

[25] K. Chen, Y. Lai, Y.-X. Wu, R. R. Martin, and S.-M. Hu, "Automatic semantic modeling of indoor scenes from low-quality rgb-d data using contextual information," *ACM Transactions on Graphics*, vol. 33, no. 6, 2014.

[26] M. Fisher, M. Savva, Y. Li, P. Hanrahan, and M. Nießner, "Activity-centric scene synthesis for functional 3d scene modeling," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 6, p. 179, 2015.

[27] T. Shao, W. Xu, K. Zhou, J. Wang, D. Li, and B. Guo, "An interactive approach to semantic modeling of indoor scenes with an rgbd camera," *ACM Transactions on Graphics (TOG)*, vol. 31, no. 6, p. 136, 2012.

[28] M. Liu, K. Zhang, J. Zhu, J. Wang, J. Guo, and Y. Guo, "Data-driven indoor scene modeling from a single color image with iterative object segmentation and model retrieval," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 4, pp. 1702–1715, 2020.

[29] W. Wu, X.-M. Fu, R. Tang, Y. Wang, Y.-H. Qi, and L. Liu, "Data-driven interior plan generation for residential buildings," *ACM Trans. Graph.*, vol. 38, no. 6, 2019.

[30] R. Hu, Z. Huang, Y. Tang, O. Van Kaick, H. Zhang, and H. Huang, "Graph2plan: Learning floorplan generation from layout graphs," *ACM Trans. Graph.*, vol. 39, no. 4, 2020.

[31] D. Ritchie, K. Wang, and Y.-a. Lin, "Fast and flexible indoor scene synthesis via deep convolutional generative models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6182–6190.

[32] R. Ma, H. Li, C. Zou, Z. Liao, X. Tong, and H. Zhang, "Action-driven 3d indoor scene evolution." *ACM Trans. Graph.*, vol. 35, no. 6, pp. 173–1, 2016.

[33] T. Weiss, A. Litteneker, N. Duncan, M. Nakada, C. Jiang, L.-F. Yu, and D. Terzopoulos, "Fast and scalable position-based layout synthesis," *arXiv preprint arXiv:1809.10526*, 2018.

[34] P. Henderson, K. Subr, and V. Ferrari, "Automatic generation of constrained furniture layouts," *arXiv preprint arXiv:1711.10939*, 2017.

[35] Y. Liang, F. Xu, S.-H. Zhang, Y.-K. Lai, and T. Mu, "Knowledge graph construction with structure and parameter learning for indoor scene design," *Computational Visual Media*, vol. 4, no. 2, pp. 123–137, 2018.

[36] H. Xie, W. Xu, and B. Wang, "Reshuffle-based interior scene synthesis," in *Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*. ACM, 2013, pp. 191–198.

[37] P. J. Diggle *et al.*, *Statistical analysis of spatial point patterns*. Academic press, 1983.

[38] J. Gignoux, C. Duby, and S. Barot, "Comparing the performances of diggle's tests of spatial randomness for small samples with and without edge-effect correction: application to ecological data," *Biometrics*, vol. 55, no. 1, pp. 156–164, 1999.

[39] P. Rosin, "Thresholding for change detection," in *Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271)*. IEEE, 1998, pp. 274–279.

[40] H. Fu, R. Jia, L. Gao, M. Gong, B. Zhao, S. Maybank, and D. Tao, "3d-future: 3d furniture shape with texture," *arXiv preprint arXiv:2009.09633*, 2020.

[41] L. Ho and S. Chiu, "Testing the complete spatial randomness by diggle's test without an arbitrary upper limit," *Journal of Statistical Computation and Simulation*, vol. 76, no. 07, pp. 585–591, 2006.

[42] J. BESAGE, "On the detection of spatial pattern in plant communities," *Bull. int. statist. Inst.*, vol. 45, pp. 153–158, 1973.

[43] W. Hines and R. Hines, "The eberhardt statistic and the detection of nonrandomness of spatial point distributions," *Biometrika*, vol. 66, pp. 73–79, 04 1979.

[44] R. Assuncao, "Testing spatial randomness by means of angles," *Biometrics*, pp. 531–537, 1994.

[45] R. M. Assunçao and I. A. Reis, "Testing spatial randomness: a comparison between t^2 methods and modifications of the angle test," *Brazilian Journal of Probability and Statistics*, pp. 71–86, 2000.

[46] P. J. Diggle, J. Besag, and J. T. Gleaves, "Statistical analysis of spatial point patterns by means of distance methods," *Biometrics*, pp. 659–667, 1976.

[47] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion," in *Scandinavian conference on Image analysis*. Springer, 2003, pp. 363–370.

[48] Y. Kleiman, O. van Kaick, O. Sorkine-Hornung, and D. Cohen-Or, "Shed: shape edit distance for fine-grained shape similarity," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 6, p. 235, 2015.

[49] Y. Liang, S.-H. Zhang, and R. R. Martin, "Automatic data-driven room design generation," in *International Workshop on Next Generation Computer Animation Techniques*. Springer, 2017, pp. 133–148.

[50] Y. He, Y. Cai, Y.-C. Guo, Z.-N. Liu, S.-K. Zhang, S.-H. Zhang, H.-B. Fu, and S.-Y. Chen, "Style-compatible object recommendation for multi-room indoor scene synthesis," *arXiv preprint arXiv:2003.04187*, 2020.

[51] M. De Berg, M. Van Kreveld, M. Overmars, and O. Schwarzkopf, "Computational geometry," in *Computational geometry*. Springer, 1997, pp. 1–17.

[52] R. L. Graham, "An efficient algorithm for determining the convex hull of a finite planar set," *Info. Pro. Lett.*, vol. 1, pp. 132–133, 1972.

[53] J. Bender, M. Müller, M. A. Otaduy, M. Teschner, and M. Macklin, "A survey on position-based simulation methods in computer graphics," in *Computer graphics forum*, vol. 33, no. 6. Wiley Online Library, 2014, pp. 228–251.

[54] S.-S. Huang, A. Shamir, C.-H. Shen, H. Zhang, A. Sheffer, S.-M. Hu, and D. Cohen-Or, "Qualitative organization of collections of shapes via quartet analysis," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 4, pp. 1–10, 2013.

[55] A. Zeng, S. Song, M. Nießner, M. Fisher, J. Xiao, and T. Funkhouser, "3dmatch: Learning local geometric descriptors from rgb-d reconstructions," in *CVPR*, 2017.

[56] R. Liu, W. Huang, Z. Fei, K. Wang, and J. Liang, "Constraint-based clustering by fast search and find of density peaks," *Neurocomputing*, vol. 330, pp. 223–237, 2019.

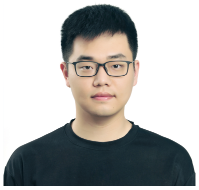
[57] B. Tong, "Density peak clustering algorithm based on the nearest neighbor," in *3rd International Conference on Mechatronics Engineering and Information Technology (ICMEIT 2019)*. Atlantis Press, 2019.

[58] S. Satkin, J. Lin, and M. Hebert, "Data-driven scene understanding from 3d models," 2012.

[59] L.-F. Yu, S.-K. Yeung, and D. Terzopoulos, "The clutterpalette: An interactive tool for detailing indoor scenes," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 2, p. 1138–1148, 2016.

Song-Hai Zhang received the PhD degree of Computer Science and Technology from Tsinghua University, Beijing, in 2007. He is currently an associate professor in the Department of Computer Science and Technology at Tsinghua University. His research interests include image/video analysis and processing as well as geometric computing.

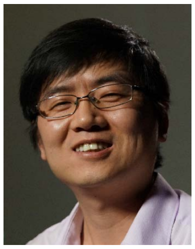
Shao-Kui Zhang is a Ph.D. candidate in the Department of Computer Science and Technology at Tsinghua University, Beijing. His research interests include computer graphics, media analysis, and computer vision. He received a B.S. degree of software engineering from Northeastern University, Shenyang, in 2018.



Wei-Yu Xie is currently an undergraduate in the School of Computer Science at Beijing Institute of Technology, and will be a Ph.D. candidate in the Department of Computer Science and Technology at Tsinghua University, Beijing from 2021. His research interests include computer graphics, computational geometry, and model reconstruction.

Cheng-Yang Luo is an undergraduate student in the School of Economics and Finance at Tsinghua University, Beijing. He also has second major in the Department of Mathematical Sciences. His research interests include computer graphics and computer vision.

Yong-Liang Yang is a Senior Lecturer in the Department of Computer Science, University of Bath. He received the B.S. degree and the Ph.D. degree of Computer Science from Tsinghua University. His research interests are broadly in visual computing and interactive techniques.



Hongbo Fu is a Professor with the School of Creative Media, City University of Hong Kong. He received the B.S. degree in information sciences from Peking University, and the Ph.D. degree in computer science from Hong Kong University of Science and Technology. He has served as an Associate Editor of The Visual Computer, Computers & Graphics, and Computer Graphics Forum. His primary research interests include computer graphics and human computer interaction.