Graphical Models 110 (2020) 101073

Contents lists available at ScienceDirect

Graphical Models

journal homepage: www.elsevier.com/locate/gmod

# Human-centric metrics for indoor scene assessment and synthesis

Qiang Fu[a], Hongbo Fu[b], Hai Yan[a], Bin Zhou[c], Xiaowu Chen[c], Xueming Li[a,*]

[a] School of Digital Media and Design Arts, Beijing University of Posts and Telecommunications, Beijing, China
[b] School of Creative Media, City University of Hong Kong, Hong Kong
[c] School of Computer Science and Engineering, Beihang University, China

ARTICLE INFO

*Keywords:*
3D modeling
Environment assessment
Indoor scene synthesis

ABSTRACT

In recent years, many approaches have been proposed to analyze the affordance of 3D objects and to facilitate the synthesis of aesthetic and realistic indoor scenes. However, how to assess the quality of such synthesized 3D scenes is still a challenging problem. To address this, we present a novel approach through so-called *Human-Centric Metrics* (HCMs) for quantitatively evaluating the layout quality of certain objects, and thus to facilitate indoor scene synthesis. Our HCMs consider both the human-object factors to assess the functional accessibility of indoor objects, and the human-environment factors to assess the subjective comfort in the scene. Given a 3D scene with arranged furniture objects, our method automatically places human agents and then use HCMs to assess both the object arrangement and indoor environment based on the position/direction of a certain human agent. The conducted user study shows that the assessment capability of HCMs well conforms to the interior design knowledge. We also use the HCMs based assessment results to synthesize 3D indoor scenes by suggesting indoor objects and their layout given an empty room. A series of experimental results and comparisons are presented to validate the usability and effectiveness of our HCMs for virtual scene assessment and synthesis.

## 1. Introduction

Indoor layout and environment influence the subjective feeling of human beings. This has been noticed and widely studied in ergonomics and environmental psychology for a long time [1,2]. Some recent studies revealed that illumination, ventilation, and space are the key design interventions for both a healthier and more comfortable home, and a happier and more productive workplace [3,4]. The recent literature in human-centric analysis for 3D scenes have been mainly focused on understanding the relationships between human activities and scenes [5,6], and using such relationships for furniture arrangement [7–9]. These techniques perform well in predicting the likelihood of actions in 3D environments, and synthesizing functionally plausible scenes under the pre-defined guidelines. However, a fundamental issue, i.e., whether a synthesized scene will make a human user comfortable or not, has not been addressed. A key challenge is to design a human-like evaluator with metrics that can measure the human subjective feelings in the virtual scene and give reliable feedback. Even though it is not an easy task to assess whether an indoor layout conforms to ergonomics. One of the simple ways is to let a human user enter the room, directly use the arranged objects, and feel the environment from perspectives such as space, ventilation and illumination. Based on the user feedback, we can assess the current indoor scene and even improve the layout quality by rearranging objects.

To design such a human-like evaluator, both human-object and human-environment factors should be considered in an integrated manner. The former could consist of ergonomic priors such as the functional accessibility of indoor objects, which can be directly calculated by using the scope of touch and view. The latter could consist of human subjective feelings about indoor environments like space, ventilation and illumination. Defining quantitative metrics to measure these subjective feelings is challenging. Fortunately, well-designed indoor layouts (e.g., 2D floor plans of residence and workspace) can be used as the positive examples to reveal the relations of indoor layouts and better subjective feelings through a data-driven method.

In this paper we propose the Human-Centric Metrics or HCMs for short, to facilitate indoor scene assessment and synthesis. More specifically, we use virtual human agents with certain actions (i.e., sit, lay, stand) to define metrics for the human-object factors from a database of 2D floor plans. We also use per capita area to describe the indoor space, and use the responses of human agents in the room energy fields (based on the window/door locations) to describe the ventilation and illumination. In this way, the statistics of these parameters calculated from the 2D floor plan database provide the metrics for the human-environment factors. Given a 3D indoor scene, we first place certain human agents to interact with the objects, and then use the

* Corresponding author
*E-mail address:* lixm@bupt.edu.cn (X. Li).

https://doi.org/10.1016/j.gmod.2020.101073
Received 25 February 2020; Received in revised form 3 May 2020; Accepted 5 May 2020
Available online 21 May 2020

HCMs to assess object arrangement based on the functional accessibility, and the comfort of the environment in the places where the human agents stay (Fig. 1-Top). Moreover, the assessment result can also guide the indoor scene synthesis by rearranging objects with better layouts (Fig. 1-Bottom).

In summary, we introduce novel human-centric metrics for indoor scene assessment and synthesis. Important characteristics of our work which set it apart from previous approaches include:

- Indoor scenes are quantitatively assessed according to the human-object factors measuring the functional accessibility of objects and the human-environment factors measuring the human feeling for the environment.
- Indoor scenes are synthesized guided by HCMs-based assessment, instead of pre-defined layout guidelines.

We conduct a user study to validate that HCMs have the capability of indoor environment assessment which well conforms to the interior design knowledge. We also demonstrate the usability of human-centric metrics in scene assessment and synthesis with various experimental results and comparisons.

## 2. Related work

In this section, we first examine the existing research on human-centric analysis for man-made objects and indoor scenes. Then we review the relevant approaches on indoor scene understanding and synthesis.

**Human-Centric Analysis.** Based on the concept of affordance, which describes what the environment offers the individual, human-centric analysis for man-made objects and indoor scenes has been a hot topic in recent years. For example, in man-made object modeling, human pose priors or ergonomics-based guidelines can be used for pose estimation [10], shape editing [11], and shape synthesis [12]. For indoor scenes, the human-object interactions can also be used for scene analysis [6,13], scene reconstruction [14], and scene synthesis [15]. There have also been a variety of efforts to describe and model the affordance. Such as the approaches provided by Hu et al., including an *interaction context* descriptor to explicitly represent the geometry of object-object interactions [16], a co-analysis method for learning a *functionality model* for an object category [17], and a method for learning a model for the *mobility* of parts in 3D objects [18]. Similarly, Pirk et al. [19] has proposed a novel general representation for proximal interactions among physical objects, which are agnostic to the type of objects or interaction involved. Our work does not intend to describe various activities and human-object interactions. Different interactions/activities in our work are implicit in the pre-defined object groups. Our motivation is very similar to [20] in which the object geometry is quantitatively evaluated based on the posed human body. A key difference is that our work focuses on the arrangement between multiple objects, rather than the geometry of a single object. Moreover, since the human-environment relations are harder to measure than the human-object relations, one of our major tasks is to explore proper metrics for indoor environment assessment is still a rather unexplored problem.

**Scene Understanding.** Understanding the hierarchical structure of indoor scene is important to indoor object organization and scenes synthesis. This is always achieved by analyzing object relations from a large indoor scene dataset. For example, Liu et al. [21] proposed to use a probabilistic grammar learned from examples, for hierarchical decomposition of a scene into semantic components. The *Imagining the Unseen* system [22] was proposed to hallucinate geometry in the occluded regions of the scanning scene by globally analyzing the physical stability of the resultant arrangements of the cuboids. Xu et al. [23] proposed a set of extracted focal points which are used to represent substructures in a scene collection for organizing a heterogeneous scene collection. Fu et al. [24] proposed to extract the activity-associated object relations from 2D floor plans to facilitate object exploration when synthesize indoor scenes. Besides, several approaches were proposed to learn spatial relationships of objects, and encode semantic scene structures from indoor scene datasets (e.g., Fisher and Hanrahan [25], Fisher et al. [26], Liang et al. [27]). In our work, we adopt the same way as [24] which collected the indoor object relations from a set of indoor 2D layouts. The major difference is, our work aims at finding proper metrics to assess the object arrangement in terms of ergonomics, rather than only object relation detection.

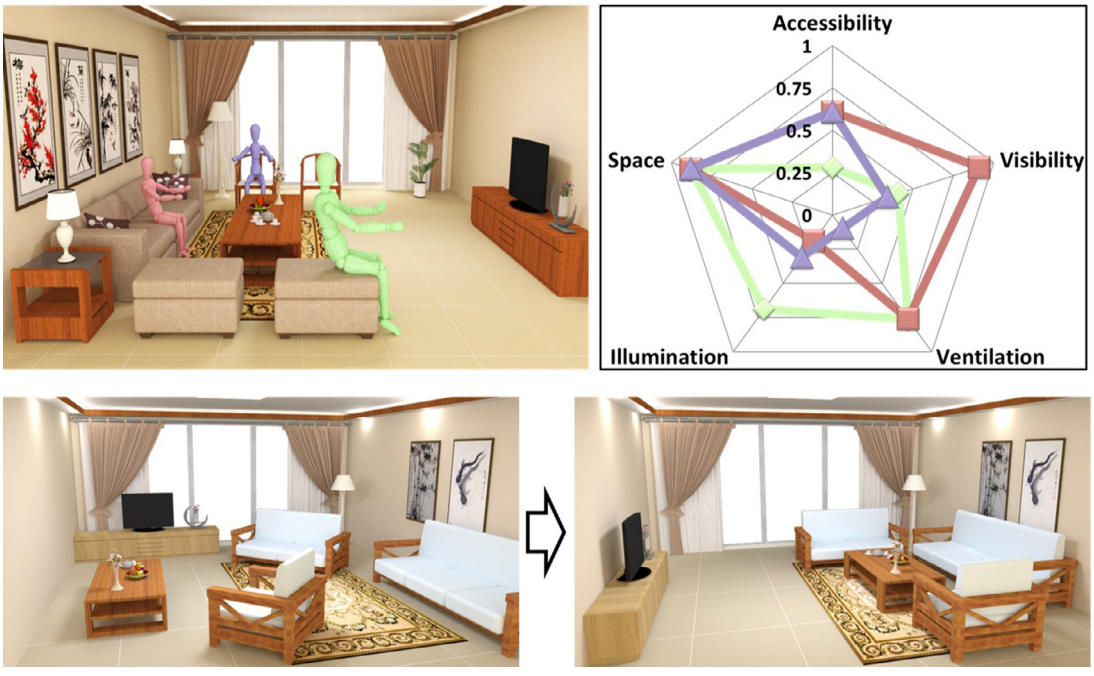


**Fig. 1.** The proposed human-centric metrics can be used to evaluate the given indoor scene in terms of each human agent about both human-object and human-environment factors (Top). Based on the assessment result, our method can generate plausible layouts to synthesize an indoor scene (Bottom).

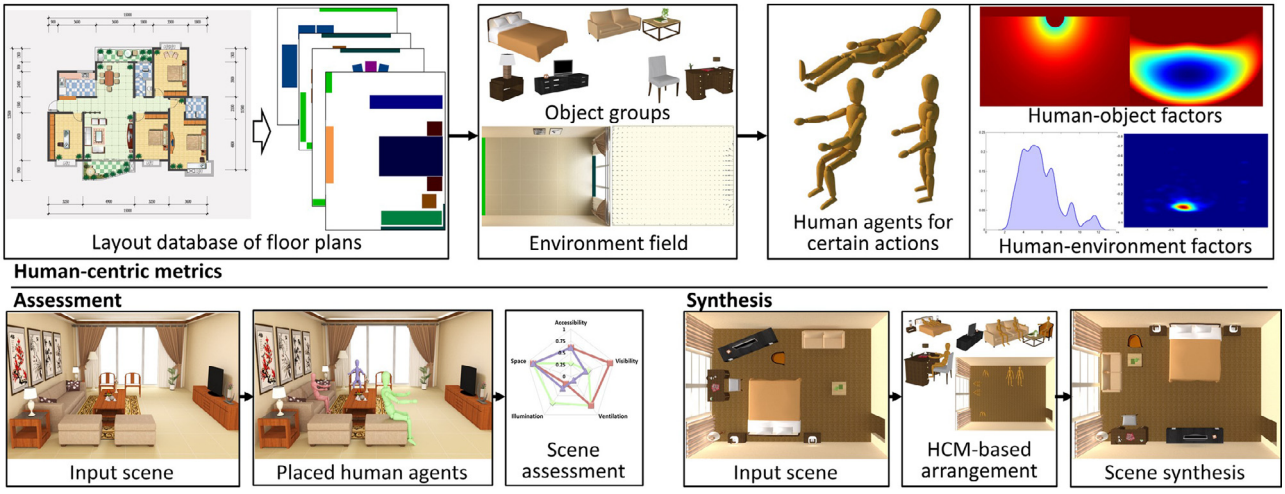


**Fig. 2.** An overview of our approach. We leverage the dataset with well-designed indoor layouts to extract the interaction-related object groups and the priors of environment fields, thus to construct the HCMs that describe both human-object and human-environment relations with respect to different actions (top). Given a scene, the HCMs can be used to assess the layout with respect to each interactive position/direction (bottom-left) and guide the layout refinement as well (bottom-right).

**Scene Synthesis.** Indoor scene synthesis has become a popular topic of computer graphics in the last decade [28]. Various systems have been developed by arranging objects to synthesize indoor scenes (e.g., Yu et al. [29], Merrell et al. [30], Wu et al. [31]). Some methods are proposed to use examples to guide the indoor scene synthesis. For instance, Fisher et al. [32] presented an example-based synthesis method to produce a diverse set of plausible new scenes, with Bayesian networks and Gaussian mixtures as a probabilistic model learning from the 3D scene database. It is also an intercrossed research direction that involves techniques of computer vision [33], e.g., Chen et al. presented an *Image2Scene* system [34] that transforms a 3D room into one that resembles the style of a scene in a given photograph. Xi and Chen proposed an approach that leverages multi-view regularization to enhance the capability of piecewise planar scenes reconstruction [35]. The recent works begin to leverage deep neural networks for indoor scene synthesis. For instance, Wang el al. [36] leveraged convolutional neural network to learn priors from the database for indoor scene synthesis, and Li et al. [37] leveraged generative neural network to generate plausible 3D indoor scenes in large quantities and varieties. To obtain plausible indoor layout for the given scene, our method adopts the human-object factors similar to the human-centric indoor scene synthesis methods. Besides, we also consider the human-environment factors (i.e., space, ventilation and illumination) in the human-centric metrics, thus to insure the synthesized scene to have a comfortable indoor environment.

## 3. Overview

As illustrated in Fig. 2, our technique consists of an offline stage to obtain the HCMs for different kinds of actions from scene database, and an online stage to use the HCMs to assess an input 3D scene. We focus on the interactive objects that the layout of 2D floor plans contain adequate prior for guiding the arrangement of such objects in 3D scenes. The core of our method is to use a 2D indoor layout database to create the metrics, thus to enable the human agents to be evaluators for assessing both the accessibility of objects and the comfort of the environment.

As our work has a similar motivation with [24], which uses a database with commercial floor plans to extract priors of interior design. These floor plans have accurate scale information, and the layout of furniture in each plan already conforms to the interior design knowledge. All objects in the database floor plans have been segmented with semantic labels. Some interactive objects have also been annotated with human agents of certain actions (i.e., sit, lay, stand). We use the same database in our work, and enrich it by counting the numbers of agents in each interactive object, and placing the agents on the interactive objects without overlapping. The database provides examples for gathering objects into groups in which objects are related to the same human agent(s), and provides plausible layouts for the human agents in the room for certain activities.

In the offline stage, we use ergonomics-based priors to form the metrics for the human-object factors, and use statistic model to form the metrics for human-environment factors from the database. In the online stage, given a 3D scene with arranged furniture, we first estimate the number of human agents in the scene by the areas of interactive objects. Then these agents are automatically put into the scene with proper positions and directions. Afterwards, each agent works as an evaluator to assess both the arrangement of the associated objects and the environment based on the HCMs (Section 5.1). Since our HCMs are able to assess object arrangement and indoor layout in the view of human feeling, it can also give suggestions to refine the original layout or even synthesize a functionally plausible scene given an empty room (Section 5.2).

## 4. Human-centric metrics

In this section, we introduce how to use ergonomic priors and statistical model learned from the floor plan database, to define the human-centric metrics for both human-object factors and human-environment factors.

### 4.1. Human-object factors

We follow the idea of [24] that gathers indoor objects into groups, that each group has objects interacted with the same human agent(s). In our implementation, three kinds of actions including sitting, lying and standing, and two kinds of interactions including touching and watching are considered. For simplicity, we pre-define the responding relations between object categories and the actions or interactions. For example, in a group with sofa, tea table and TV set, the sofa has the action label of sitting, the tea table has the interaction label of touching, and the TV set has the interaction label of watching.

For an object group, the quality of all objects' arrangement depends on the ergonomic relations with the agent. Therefore, to establish the relationship between objects in a group with respect to the human agent (i.e., the human-object factors), the HCMs should be able to assess the affordance of touching and watching based on the ergonomic priors.

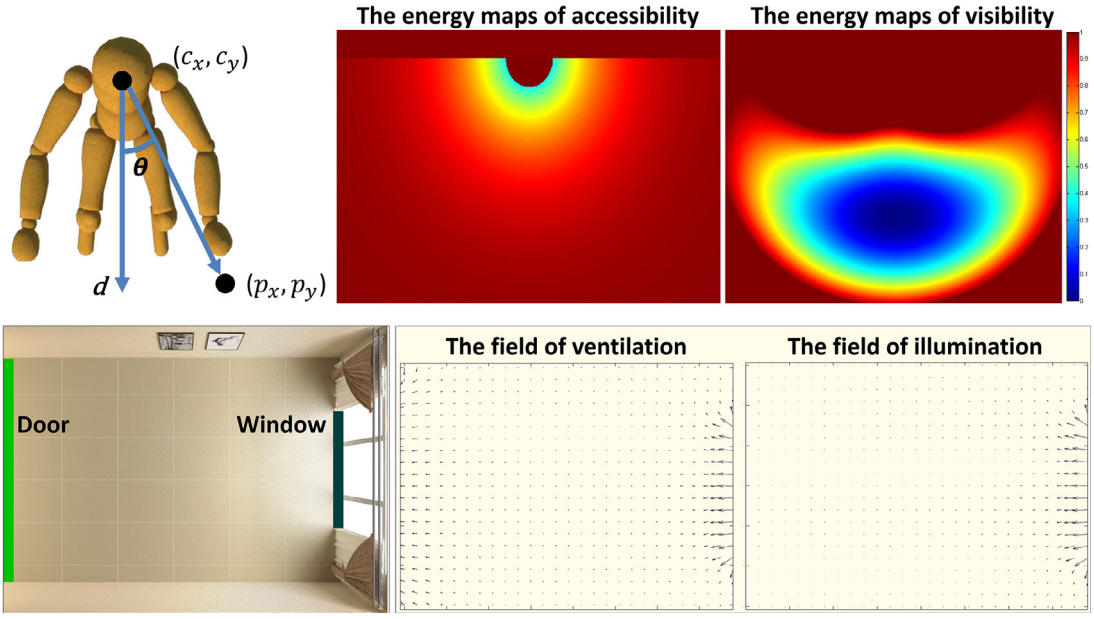


**Fig. 3. Top:** The human agent (left) and its energy maps of accessibility and visibility (right). **Bottom:** A room (left) and its gradient fields of ventilation and illumination with the window and door as the sources (right).

The position/direction of a human agent with a certain action can be determined based on the interactive object in a group. Motivated by the concept of anthropometry, which plays an important role in ergonomics where statistical data about the distribution of body dimensions in the population are used to optimize products [38], we attempt to turn the fixed agent's arms reach and field of view to a quantitative assessment function, as illustrated in Fig. 3 (Top-Left). In the top-down view 2D projection of a 3D human agent, let $\mathbf{c} = (c_x, c_y)$ be the center of shoulders representing the agent's position, and $\mathbf{d}$ be a unit normal vector representing the agent's front direction. We define the accessibility energy as follows:

$$E_a(\mathbf{c}, \mathbf{d}, \mathbf{p}) = \begin{cases} 1, (\mathbf{p} - \mathbf{c}) \cdot \mathbf{d} < 0 \text{ or } D(\mathbf{p}, \mathbf{c}) < \omega_1 \\ 1 - \exp(-\frac{D(\mathbf{p}, \mathbf{c})}{2\omega_1}), \text{otherwise} \end{cases}, \tag{1}$$

where $\mathbf{p} = (p_x, p_y)$ is a position for touching, and $D(\mathbf{p}, \mathbf{c}) = ||\mathbf{p} - \mathbf{c}||_2^2$ measures the Euclidean distance between $\mathbf{c}$ and $\mathbf{p}$. In this way, the position within the arm reach has low energy, and the energies of the areas out of arm reach would soon increase to maximum. For the areas which are inconvenient to access, i.e., the position behind ($(\mathbf{p} - \mathbf{c}) \cdot \mathbf{d} < 0$) or too close ($D(\mathbf{p}, \mathbf{c}) < \omega_1$) to a human agent, We also set their energies to maximum ($E_a(\mathbf{c}, \mathbf{d}, \mathbf{p}) = 1$). For simplify, here we adopt a rough distance threshold $\omega_1$ ($\omega_1 = 0.5$ in our implementation) based on anthropometry to calculate the arm reach.

Similarly, since the view distance and angle impact the visual acuity, there have been researches (e.g., Tam et al. [39]) demonstrating that proper distance exists between human users and objects for watching like a TV set. In our work, we accept the empirical principle that the viewing distance should be between 3 and 4 m, which can be formulated as a distance constraint $\tilde{D}(\mathbf{p}, \mathbf{c}) = (D(\mathbf{p}, \mathbf{c}) - 3.5)/\omega_2$ ($\omega_2 = 2.5$ in our implementation). The field of view is described by a direction function: the position directly in front of the human has low energy. Let $\theta$ be the angular difference between the viewing and front directions. We then define visibility energy as follows:

$$E_v(\mathbf{c}, \mathbf{d}, \mathbf{p}) = \begin{cases} 1, (\mathbf{p} - \mathbf{c}) \cdot \mathbf{d} < 0 \\ \min(1, \sin(\theta)^2 + \tilde{D}(\mathbf{p}, \mathbf{c})^2), \text{otherwise} \end{cases} \tag{2}$$

We also set the positions which are behind the human agent ($(\mathbf{p} - \mathbf{c}) \cdot \mathbf{d} < 0$) to the maximum energy $E_v(\mathbf{c}, \mathbf{d}, \mathbf{p}) = 1$. Fig. 3 (top-right) illustrates the energy maps of for an agent about the accessibility and visibility.

### 4.2. Human-environment factors

Environmental comfort is another crucial index for scene quality assessment. For a well-designed indoor scene, the factors such as space, ventilation and illumination are well-considered. However, designing a function measuring the human feelings about these environmental factors is not an easy task. To address the problem, we propose to learn such measure functions via statistical analysis from the floor plan database, which has been designed to follow the interior design knowledge. We first represent these environmental factors by per capita area (for space) and energy fields (for ventilation and illumination), and then statistically analyze the relations between these environmental factors, and the number, position and direction of the human agent.

The space feeling of human for indoor scenes is very complex, it always involves factors like the room size, layout, color theme, etc. For simplicity, we adopt a similar strategy to [24], which leverages the ratio of occupancy to explore proper furniture for the given room based on the room size. In our work, we use per capita area of the residents in a room, which is also associated with both the number and size of furniture in the room to measure the human space feeling. The per capita area of an indoor scene (denoted as $Ar$) influences the human feelings about the space, too large $Ar$ will make human feel the room empty and too small $Ar$ will make human feel crowded. We divide the total area of the room by the total number of human agents, to obtain the per capita area of the room. The number of agents in a scene depends on the objects associated with certain actions, such as couch, bed, chair. More specifically, for each category of objects with certain actions in a given scene, the summation areas of their projections on the floor are divided by the required area per agent for a certain action (e.g., sitting requires 0.4 square meter). We then sum the numbers of agents for each object category in the scene to obtain the total number of human agents.

On the other hand, environmental factors such as ventilation and illumination have been widely studied in interior design [1,40]. According to these researches, windows and doors always play important roles in ventilation and illumination. The air often flows from windows to doors [3], and the natural light often comes from windows. These motivate us to assume that windows have high energy while doors have low energy, hence the energy field in a room can be used to represent its ventilation and illumination. We adopted the definition of the electric

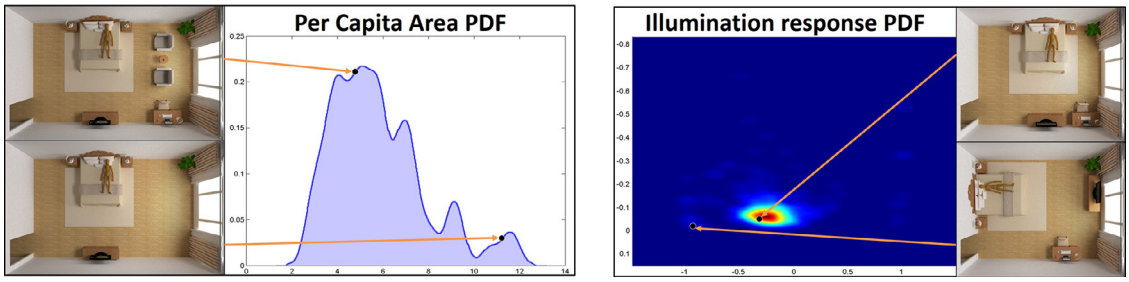


**Fig. 4.** The 1D PDF of the per capita area (left) and the 2D PDF of the response to illumination (right) for bedroom. In each PDF, we give two scene examples and their positions in the chart.

field formed by point charges to construct the energy field. The energy field depends on the position and size of the windows and doors. We use the superposition principle to combine energies generated by all source points. For simplicity, sampled points in the floor plan with the labels of windows and doors are regarded as the source points. For point $\mathbf{p}$ in the floor plan of a room, its energy can be calculated as follows:

$$E_f(\mathbf{p}) = \sum_i \frac{Q(i)}{4\pi||\mathbf{p} - P(i)||_2^2}, \tag{3}$$

where $P(i)$ is the coordinate of the $i$th source point in the room, and $Q(i) = \pm 1$ indicates the type of source, which determines the directions of the energy field. In our implementation, we set $Q(i) = -1$ for source points of windows and $Q(i) = 1$ for doors, to ensure the gradient of the energy field to flow from windows to doors (e.g., air flowing) or just directly inside (e.g., sunlight). Note that both windows and doors are involved in representing the ventilation, while the illumination only involves windows (see Fig. 3 (bottom)). In this way, we obtain the energy field $E_f(\mathbf{p})$ and its gradient $E'(\mathbf{p})$. For a human agent placed in the field, we use the field energy at position $\mathbf{c}$ of the agent and the direction difference between the agent and the field gradient, to jointly represent the response of such an agent to the energy field, denoted as $R(\mathbf{c}, \mathbf{d}) = (E_f(\mathbf{c}), E'_f(\mathbf{c}) \cdot \mathbf{d})$. We can use this definition to calculate the response of human agent to the energy fields of both ventilation and illumination.

After that, we propose to learn the metrics by statistically analyzing these factors based on the floor plan database. Similar to [41] in which the probability density function (PDF) with Gaussian kernel is proposed to describe the pair-wise relations of various object geometric parameters, we also employ the PDF to define the assessment metric functions about all the above human-environment factors. With a set of data $X$ with $n$ members, the PDFs with Gaussian kernel are represented by a 1D kernel density estimator $K_1(r)$ and 2D kernel density estimator $K_2(r_1, r_2)$ as follows:

$$K_1(r) = \frac{1}{n}\sum_{l=1}^{n} g(r - x_l, h),$$
$$K_2(r_1, r_2) = \frac{1}{n}\sum_{l=1}^{n} g(r_1 - x_l, h_1) g(r_2 - \tilde{x}_l, h_2), \tag{4}$$

where the Gaussian kernel $g(t, h) = \frac{\exp(-t^2/2h^2)}{\sqrt{2\pi h^2}}$, and the bandwidth $h := \sigma \cdot (\mathrm{perc}_{95}X - \mathrm{perc}_5 X)$, in which $\mathrm{perc}_5 X$ denotes the 5th percentile of $X$, and $\sigma = 0.05$. With the 1D kernel density estimator, we statistically analyze the per capita area of the room by using $K_1(Ar)$ for scenes from the database. For the response of human agent to the energy fields of both ventilation and illumination, which are jointly represented by dual data, we use the 2D kernel density estimator $K_2(R(\mathbf{c}, \mathbf{d}))$ for statistical analysis. Note these calculations are performed on the database indoor scenes of the same types. We show two examples of 1D and 2D PDFs as illustrated in Fig. 4, which can be used for indoor scene assessment.

## 5. Assessment and synthesis

In this section, we first introduce how to assess indoor scenes by the proposed HCMs, and then give more details about the application of HCMs in indoor scene synthesis.

### 5.1. Scenes assessment

Based on the HCMs, each human agent can be used as an evaluator to assess the local object arrangement and global indoor layout in the view of human feelings. We use the human-object relations for assessment of object arrangement mainly because they perform better than object-object relations in explanation of why a certain object arrangement is proper. For example in Fig. 5(b), the human agents on a couch can have two different directions to watch TV. Only using object-object relations (e.g., Yu et al. [29]) is not adequate to assess which direction is better, while using HCMs can give an assessment for such a case.

Given a 3D scene with arranged furniture, let $\mathcal{G}$ be one object group in the scene. The number of the associated agents $N$ for each group is calculated by the 2D projection areas of the interactive objects and the required per capita area for certain actions (described in Section 4.2). To determine proper positions/directions of human agents in $\mathcal{G}$, a human agent with a certain action (i.e., sit, lay, stand) is represented by the 2D bounding box of its projection from the top-down view, with its front direction. We use a sitting agent to test the projection edge for objects like chair and couch, and a lying agent to test the projection area for objects like bed, while standing agents are directly put in front of objects like wardrobe and cabinet. Since the positions $\mathbf{c}_n$ and directions $\mathbf{d}_n$ of the human agents might be influenced by how the objects are arranged in the group, especially for a large interactive object like a sofa which might have more than one human agent, we employ the following optimization that leverages the interactive objects for touching/watching to place the agents in $\mathcal{G}$ by searching the projection edge/area of objects for sitting/lying, respectively.

$$\arg\min_{\mathbf{c}_n, \mathbf{d}_n} \sum_{n,k} E(\mathbf{c}_n, \mathbf{d}_n, \mathbf{p}_k) + \sum_{i,j} W(\mathbf{c}_i, \mathbf{c}_j), \tag{5}$$

where $n, i, j \in \{1, \cdots, N\}$, $i \neq j$, $k = 1, \cdots, K$ and $\mathbf{p}_k$ is one of $K$ points within the projection area of the interactive object for touching or watching in object group $\mathcal{G}$. We also set a penalty term $W(\mathbf{c}_i, \mathbf{c}_j)$ to avoid overlapping between the $i$th and $j$th placed agents, namely, $W(\mathbf{c}_i, \mathbf{c}_j) = \infty$ if the 2D bounding boxes of two agents overlap, and $W(\mathbf{c}_i, \mathbf{c}_j) = 0$ otherwise. Since we assess the accessibility and visibility separately, we use $E(\mathbf{c}, \mathbf{d}, \mathbf{p})$ to represent both of the energies ($E_a(\mathbf{c}, \mathbf{d}, \mathbf{p})$ and $E_v(\mathbf{c}, \mathbf{d}, \mathbf{p})$) in two scenarios, discretize the positions in the detected area (i.e., projection edge/area) of each interactive object, and sample the direction into four types (up, down, left and right). Then due to the limited positions and directions, we solve this optimization problem by enumeration. For example in Fig. 5(a), the red regions of the sofa edges are used for detection. Fig. 5(b) shows two different solutions of human agent placement: the left one would be adopted if an object for watching is on the right

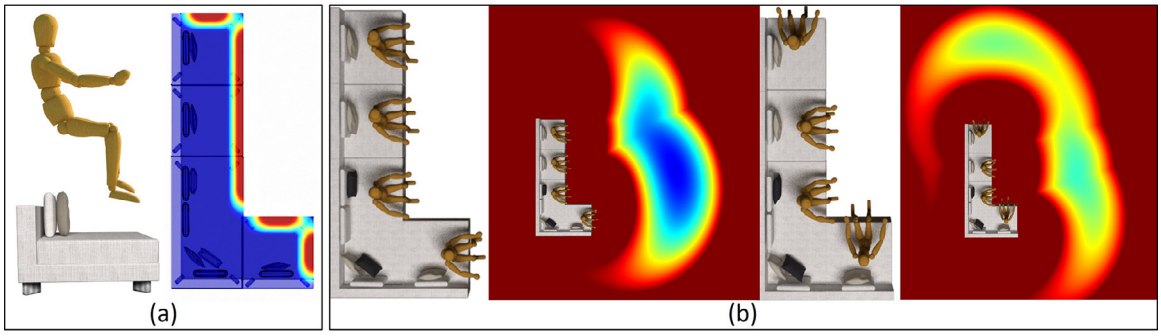


**Fig. 5.** **(a)** We use a human agent to detect the interactive area on the projection of an object. **(b)** We visualize the energy maps of visibility for different positions and directions of the agents.

side of the sofa, while the right one would be adopted if the watching object is on the top side.

After settling all agents, we attempt to leverage the HCMs defined in Section 4 for scene assessment. We directly switch the accessibility and visibility energies to the assessment scores of human-object factors. For the human-environment factors, we use the probability of the per capita area of the room $Ar$ and the response of human agent to the energy fields of both ventilation and illumination $R(\mathbf{c}_n, \mathbf{d}_n)$, by the integral of the PDFs obtained in Section 4.2. More specifically, we define all assessment scores as follows:

$$\begin{aligned} SO(\mathbf{c}_n, \mathbf{d}_n, \mathcal{G}) &= 1 - \min_k(E(\mathbf{c}_n, \mathbf{d}_n, \mathbf{p}_k)), \\ SR(Ar) &= \textstyle\int_{Ar-\rho}^{Ar+\rho} K_1(Ar), \\ SE(r1, r2) &= \textstyle\int_{r_1-\rho}^{r_1+\rho} \int_{r_2-\rho}^{r_2+\rho} K_2(r1, r2), \end{aligned} \tag{6}$$

where $r_1$ and $r_2$ are the elements of $R(\mathbf{c}_n, \mathbf{d}_n)$, $\mathbf{p}_k$ is the same in Eq. (5) for object group $\mathcal{G}$, and we set $\rho = 0.1$. In this way, we obtain the assessment scores $SO$ for accessibility and visibility, $SR$ for space, and $SE$ for ventilation and illumination. Note that since all agents have the same $Ar$, the space assessments for all agents in a room are same. In this way, we can obtain the assessment of a given indoor scene in terms of a certain placed human agent, or obtain a comprehensive assessment by the weighted average assessment scores of all human agents in the scene.

### 5.2. Assessment-based scene synthesis

The process of our indoor scene synthesis consists of two stages. First, we gather the existing objects of the given scene into groups based on the interactive relations (described in Section 3) and determine the local object arrangement (i.e., relative positions/directions) of the objects within a group based on the human-object assessment. We then explore a proper global layout for these gathered object groups in the room based on the human-environment assessment. Besides, we can also leverage the metrics for room space to suggest plausible object combinations for indoor scene synthesis.

**Local Object Arrangement.** As illustrated in Fig. 6-Top, in the first stage we use the human-object assessment score $SO$ to determine the relative position/direction of the objects within a group for local object arrangement. More specifically, the human agents are firstly placed on the interactive regions of the objects with activities like sitting or lying in the group (as discussed in Section 5.1). Since all objects have not been arranged into groups yet, we expect the agents on the same objects should have as the same directions as possible to determine their placement rather than using Eq. (5). Then we calculate the assessment scores $SO$ of accessibility and visibility for areas around these objects with respect to their associated agents (Eq. (6)), and choose the areas where there has the maximum touching/watching assessment score for object placement. To predigest the solving process, the objects are calculated in turn based on their sizes, and the objects for watching are tackled in the end. That is, if the largest object is for touching, we fix its position/direction first, then pick the largest object with placed human agents and determine its relative position/direction based on $SO$, vice versa. Note that the front directions of objects for touching/watching could always be opposite to the front directions of the interacted human agent. This is useful to determine the object orientations in the arranged groups.

**Global Layout Generation.** The purpose of the second stage is to explore a global layout to place the arranged object groups in the room. Since the human-environment assessment can indicate where the human agents should be under a certain room configuration, similar to the local object arrangement determination, the object groups are placed in the room where we can get the maximum sum of the human-environment assessment score $SE$ for all agents about both ventilation and illumination. Fig. 6-Bottom illustrates an example of the global layout generation. The placement order also depends on the sizes of the object groups. Given an empty room (Fig. 6(a)), we fix the relative positions/directions of the associated human agents and then go through all the sampled positions in the room with four directions (i.e., up, down, left and right) to explore the absolute positions/directions of the agents where they have the maximum scores $SE$ (Fig. 6(b)). Note the scores $SE$ for ventilation and illumination are used in combination here. After that, the global layout of such an object group is determined. We then masked out the occupied areas for the next group (Fig. 6(c)). Finally, all groups are placed in the room after repeating this procedure (Fig. 6(d)). Note that, the positions of certain objects (e.g., sofa and TV set) might need to be slightly adjusted to force them against the nearby wall.

**Object Group Suggestion.** Moreover, we also attempt to suggest proper object groups based on the scene type and the per capita area of the room. That is, given an empty room with a known scene type, we first estimate the number of residents $N$ according to the statistics on the per capita area of such a type of room. Namely, for the area of the given room, the per capita area with $N$ residents would lead to the maximum space assessment score $SR$. Then we give a series of suggestions on the plausible object groups $\sum_i GN(\mathcal{G}_i) = N$ where $GN(\mathcal{G}_i)$ is the number of users for object group $\mathcal{G}_i$. For each suggestion result, we calculate $P(\mathcal{G}_i)$ which is the frequency that group $\mathcal{G}_i$ occurs in the floor plan database, as the priority to order all suggestion results from the maximum $P(\mathcal{G}_i) = 1$ to the minimum $P(\mathcal{G}_i) = 0$. With the suggested object groups, we perform the scene synthesis process to obtain the final 3D scenes.

## 6. Results and discussions

In this section, we first show some representative assessment and synthesis results of various indoor scenes. Then we discuss the results by comparisons with the state-of-the-art indoor scene synthesis methods [29,34] and the ergonomic criterions, and conduct a user study to validate the effectiveness of our HCMs-based assessment.

**Scene Assessment and Synthesis.** The layout dataset we used consists of 472 floor plans across 5 types of indoor scenes, including bedrooms (187), living rooms (113), dining rooms (113), office rooms (46), and conference rooms (13). We considered and used 12 categories of

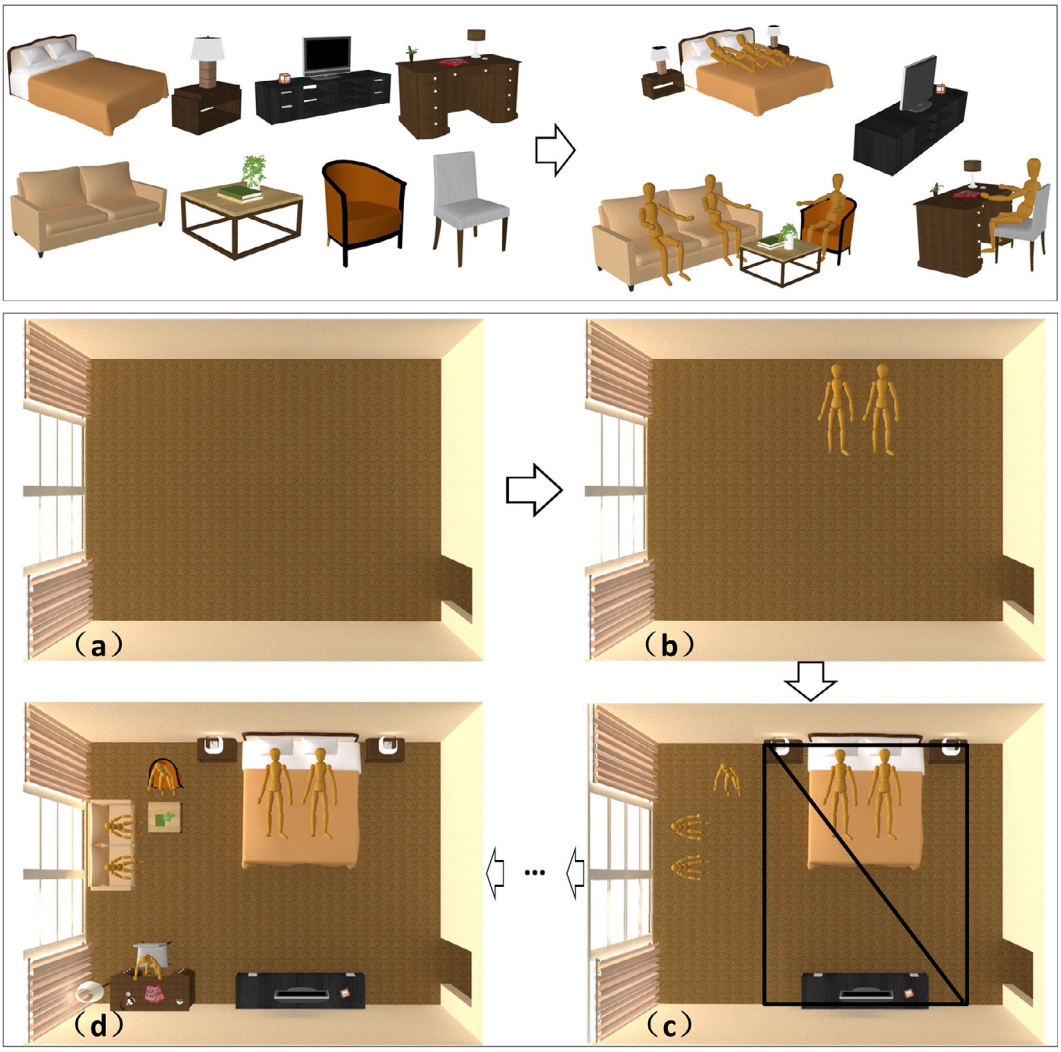


**Fig. 6. Top:** the local arrangements for the interaction-related object groups are determined (from left to right), based on the accessibility and visibility energies calculated by our HCMs. **Bottom:** based on the environment assessment by HCMs for the given room (a), the global layouts for the arranged groups are determined in turn (b-d): we detect the positions/directions of the associated human agents for a certain object group (b), place the group in the room and mask out the occupied area (c), and repeat this procedure until all group placement is completed (d).

furniture-level indoor objects, including chairs, beds, sofas, coffee tables, TV sets, tables, wardrobes, bedside tables, dining tables, cabinet, conference tables, and working desks. Based on the HCMs, we can assess indoor scenes in the view of each human agent, and give five assessment scores including accessibility, visibility, space, ventilation and illumination. Fig. 7 shows the assessment results of five categories of scenes. Each case has 3 different layouts and been assessed in terms of the same human agent. The positions/directions of human agents in the room are same for scenes #1 and #2, while the relative positions/directions of objects in the group are same for scenes #1 and #3. From the assessment results, we can see that the different scores of accessibility, visibility between scenes #1 and #2 show that the HCMs can assess the object arrangement, and the different scores of ventilation and illumination between scenes #1 and #3 show that the HCMs can assess the indoor layout reflecting the environment feelings.

In Fig. 8, we show some synthesized indoor scenes guided by HCMs-based assessment. The top four cases show that our method is able to provide a proper object arrangement for indoor scenes. The bottom three cases show that given an empty room with a known scene type (bedroom in this case), our method suggests a proper number of human agents and their associated object groups, ensuring the room to have an appropriate per capita area. On the average, our method took less than 1 s for scene assessment and less than 10 s for scene synthesis, tested on a PC with Intel Core i7-4790 3.60GHz PC with 16GB RAM.

**Comparisons and User study.** We compare our method with the state-of-the-art indoor scene synthesis methods [29,34]. For small rooms with less than 20 objects, the average time cost of [29], Chen et al. [34] and our method for completing indoor scene synthesis are 40, 5 and 5 s, respectively. We show an example in Fig. 10-Left, in which the same indoor scene inputs are used for indoor scene synthesis by the three methods. We can see that all these methods generated plausible indoor layouts. However, Yu et al. [29] requires more time to fulfill the iterative computation, while [34] requires a similar scene photo as the reference to guide the synthesis. In Fig. 10-Right, we show two plausible object arrangement results, where both our method and [29] can generate the left arrangement result, while our method can also generate the right one benefited from the flexible object-object relations. Since the HCMs leverage human agents to establish object-object relations, our method can obtain more plausible indoor layouts such as the right one, which is more proper for a small input room.

We also validate the effectiveness of the five assessment scores given by our HCMs via comparisons with ergonomic criterions. As illustrated in Fig. 9 (a), we align the arms reach and binocular field (within 120°) of human beings to the energy maps of accessibility and visibility of

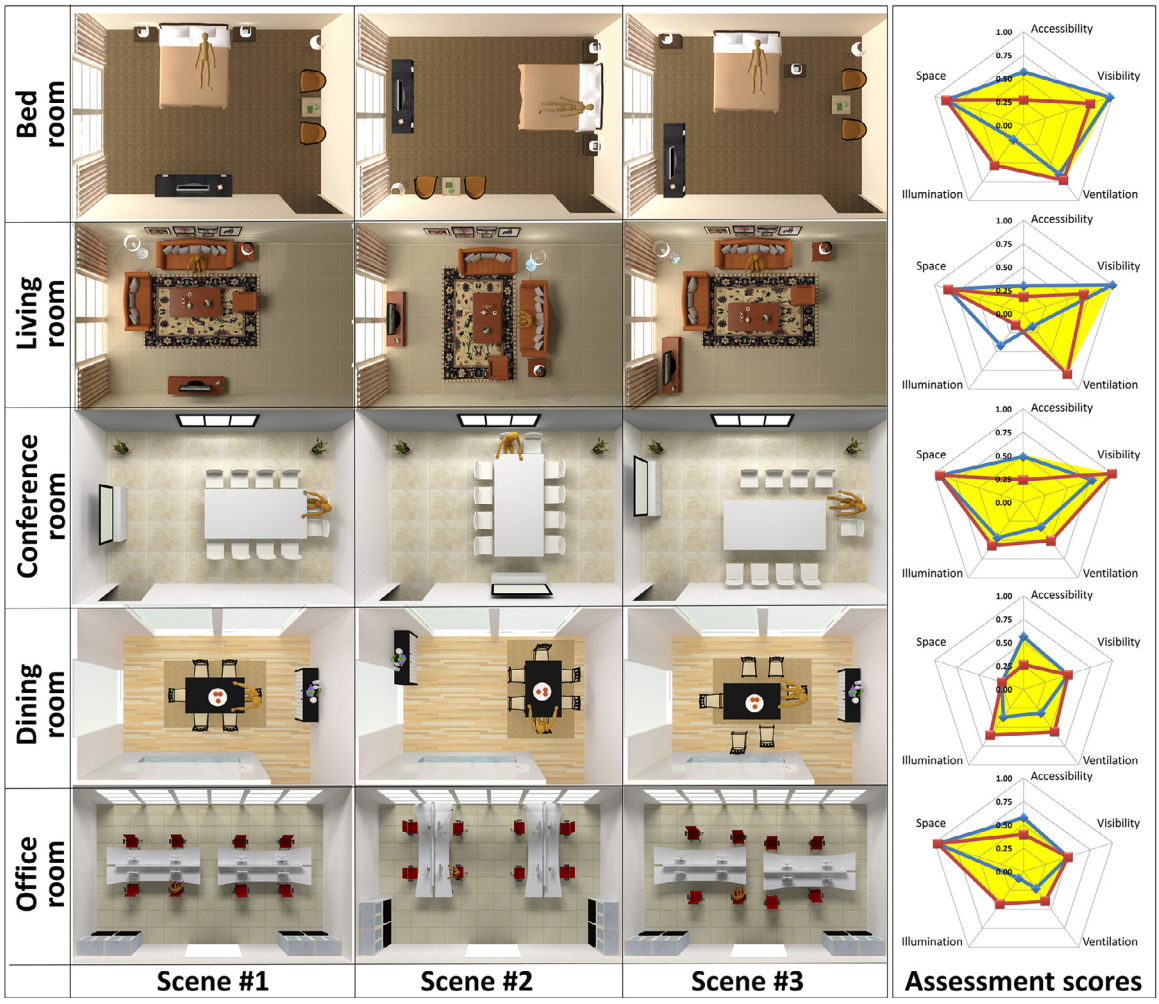


**Fig. 7.** Five groups of 3D scenes with different layouts and the selected human agent (left), and the corresponding assessment scores (right). The coordinate of each radar map is the same as Fig. 1, and the yellow area, blue line and red line are the assessment scores of scenes #1, #2 and #3, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

our HCMs as comparisons. In Fig. 9 (b) we also provide the peaks of PDFs for the space of the five scene categories: they are from 3 to 6 $m^2$ which is an empirical range obeyed by interior designers. It indicates that the assessment results of our HCMs about the above three indexes well conform to the ergonomic criterions. The assessments of ventilation and illumination are more subjective, usually without explicit criterion. Hence a user study, in which indoor scene examples are assessed both by our method and interior designers, was conducted to demonstrate the usability of the HCMs in ventilation and illumination assessment.

The user study was conducted with 8 professional interior designers with rich specialized knowledge. To prepare for the user study, we collected 8 groups of 3D indoor scenes from 5 different types (2 groups for bedroom, living room and conference room, and 1 group for dining room and office room). Each group has 3 scenes with different layouts whose qualities decrease in turn (see 3 layouts of one scene group at the top of Fig. 11). We selected one human agent and calculated the average value of the associated assessment scores about the ventilation and illumination based on our HCMs. Afterwards, we asked each participant to assess the comfort of the selected human agent with respect to the ventilation and illumination, by giving scores in the range from 0 (poorest) to 1 (best) with one effective figure retained after the decimal point. Since the assessment scores are subjective, we align the mean values of the assessment scores of both participants and our HCMs, by subtracting the mean values from all assessment scores in each group, in order to compare whether they are consistent. The assessment results are summarized in Fig. 11 (bottom), where we can see that the assessment by HCMs has a consistent trend as the interior designers, which means the assessment of indoor environment by HCMs conforms to the subjective feelings of human interior designers.

**Limitations.** Our current approach has several limitations. First, since we use all human agents in each database object group to calculate the probability density function, the metrics for human-environment factors might be biased and impacted by the styles of both object shape and layout in the floor plan database. Moreover, the assessment is conducted by each human agent rather than the whole room. Even though we can comprehensively consider all human agents in the room to assess the whole indoor scene, finding proper weights to balance the impact of each agent is still not easy. For example, a bedroom with bed and sofa may not have correct assessment result, if the agents on bed and sofa make the same contributions to the comprehensive assessment. Second, our method focuses on furniture-level objects, for small objects like cups, laptops, etc., method of [8] can be used after performing our method to place the interactive furniture. Third, since we place the object groups in turn for indoor scene synthesis, if the placement of one group is improper, it might lead to the improper placement for the rest groups. To address this, we also provide some alternative placement solutions of the object group (i.e., positions/directions of associated human agents with suboptimal assessment scores) for users to choose. Lastly, our current

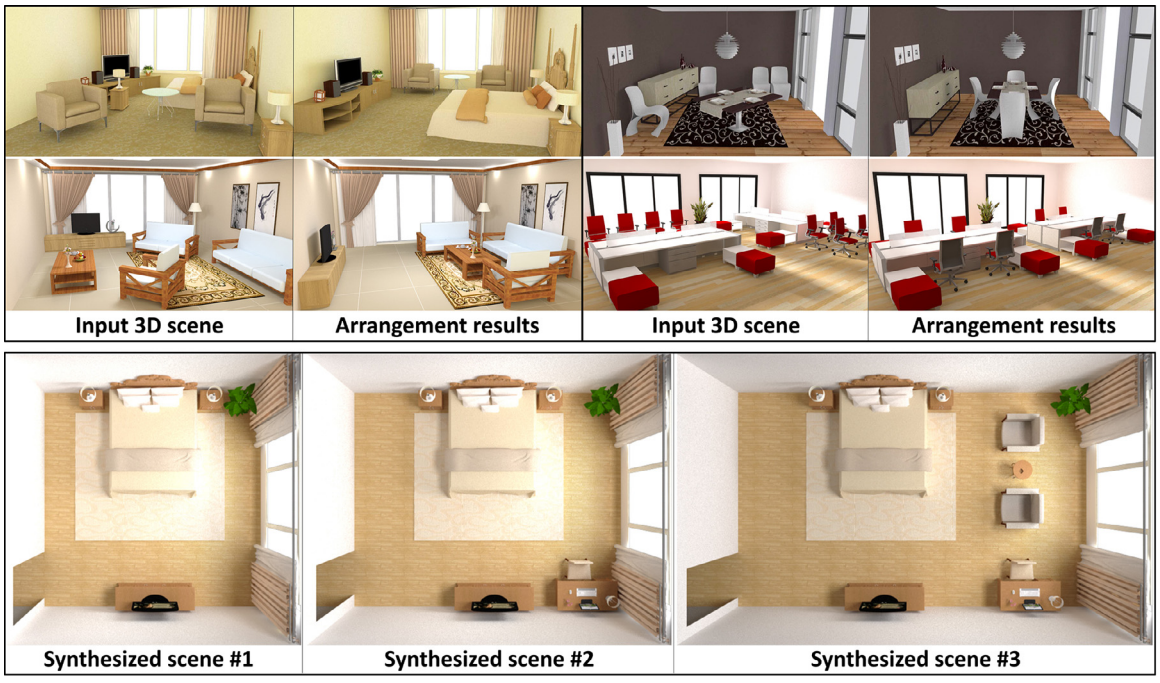


**Fig. 8.** Based on the HCMs, we can not only refine the object arrangement (top four cases), but also synthesize scenes by suggesting the indoor object groups with respect to the size of the room (bottom).

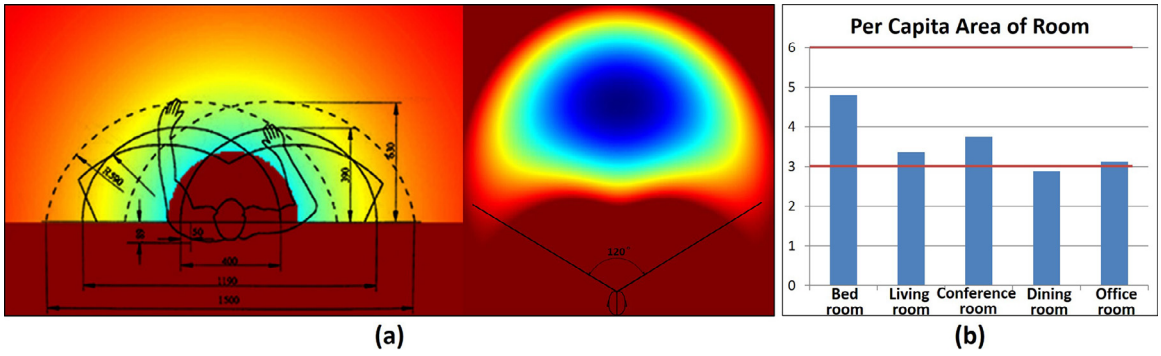


**Fig. 9.** **(a)**: comparing the arms reach and binocular field with the accessibility and visibility energy maps of our HCMs. **(b)**: the peaks of PDFs for the per capita areas of the five scene types, and the desired range between two red lines. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

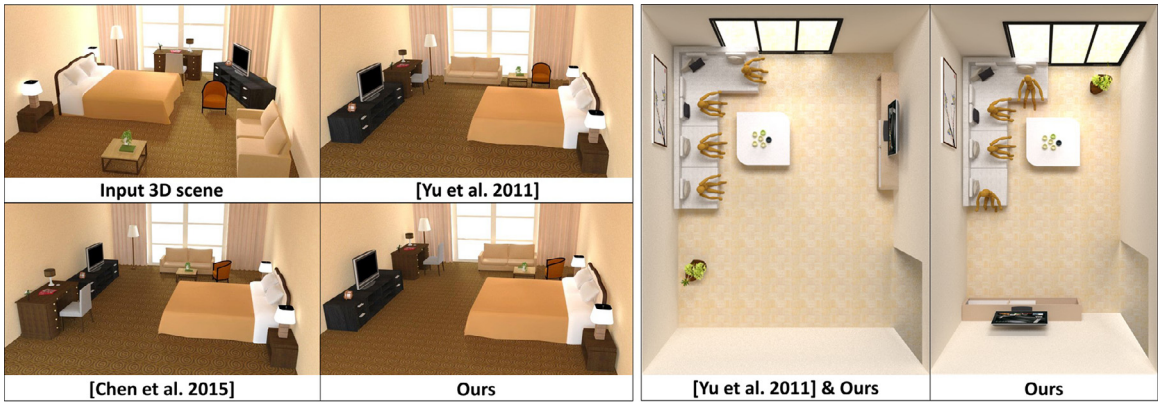


**Fig. 10.** Comparisons with the works of Yu et al. [29] and Chen et al. [34].

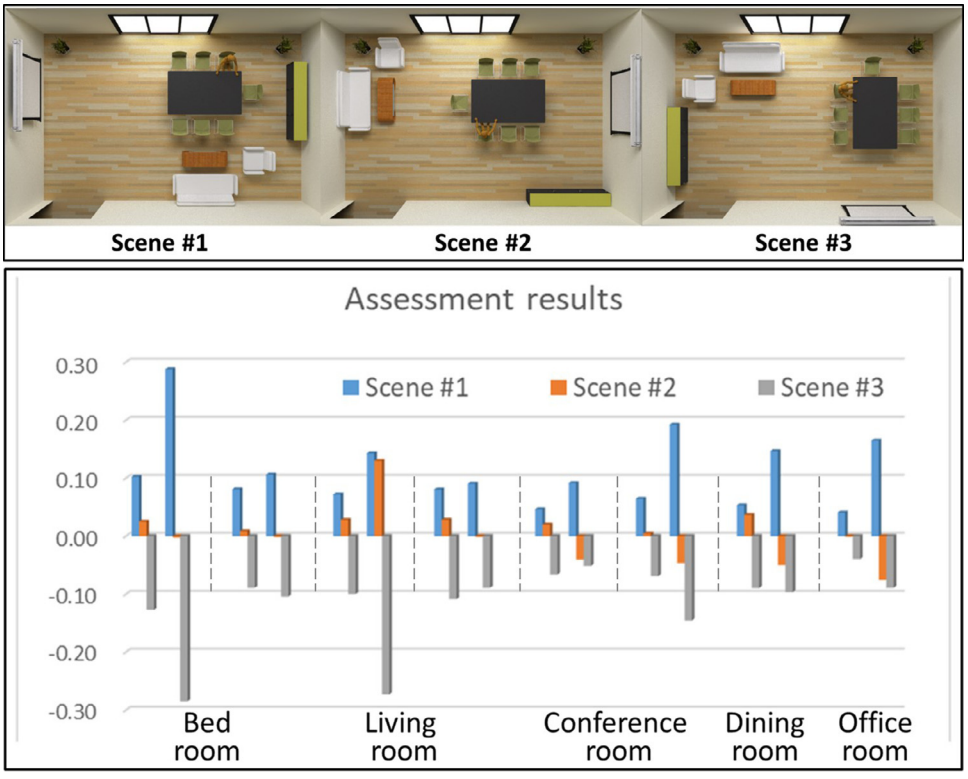


**Fig. 11.** **Top**: one scene group (3 scenes) with different layouts for assessment comparison. **Bottom**: 8 pairs of assessment results (separated by the grey dotted lines) for scene groups in 5 different scene types, with each pair having 3 assessments by the user study participants (left) and 3 assessments by our HCMs (right).

method has made several simplifications and assumptions, including using rough and empirical parameters to assess the touching and watching activities, using the energy field to represent the ventilation and illumination of the room. Therefore, the assessment scores calculated by the proposed HCMs are still very qualitative. However, since we do not intend to establish an evaluation system for indoor scenes with respect to various indoor indexes which might require more complex and precise evaluation models, these assessment scores in our method are more proper to be used to distinguish which layout is better for a room, thus to facilitate the indoor scene synthesis.

## 7. Conclusions

This paper has presented a novel method that assesses both the object arrangement and indoor layout for the given scene, thus to guide the indoor scene synthesis. We tackle this problem by introducing human-centric metrics which leverage ergonomics-based priors to evaluate the human-object factors, and the statistic models learned from a floor plan database to evaluate human-environment factors. Based on the proposed human-centric metrics, our method is able to grade the intra-group object arrangement in terms of the associated human agents, and the indoor layout in the room as well. In this manner, our method can assess the given indoor scene with quantitative scores, with a quality most often comparable to those manually produced by artists, as demonstrated by a user study. Moreover, our method can also use assessment scores to synthesize plausible indoor scenes.

Our work still has much room for improvement. In the future, we plan to further consider more personalized factors like age, gender and body shape of human agents to enrich our HCMs and facilitate both the assessment and synthesis of indoor scenes. The aesthetical factors can also be important to influence human emotion, e.g., what kind of object arrangement will feast residents' eyes when they stay somewhere. We believe that using human-centric metrics for indoor scene assessment and synthesis can open up new opportunities towards the more ergonomic and personalized computer aided interior design.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## CRediT authorship contribution statement

**Qiang Fu:** Conceptualization, Methodology, Writing - original draft, Software. **Hongbo Fu:** Conceptualization, Methodology, Writing - review & editing. **Hai Yan:** Data curation, Visualization, Validation. **Bin Zhou:** Methodology, Investigation. **Xiaowu Chen:** Conceptualization, Supervision. **Xueming Li:** Supervision.

## Acknowledgements

This work was partially supported by grants from the NSFC (No. 61532003, No. 61902032), City University of Hong Kong (No. 7004915), and the Open Project Program of the State Key Laboratory of Virtual Reality Technology and Systems, Beihang University (VRLAB2018C11, VRLAB2019B01) and Shenzhen Research Institute, City University of Hong Kong.

## Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.gmod.2020.101073.

## References

[1] L. Huang, Y. Zhu, Q. Ouyang, B. Cao, A study on the effects of thermal, luminous, and acoustic environments on indoor environmental comfort in offices, Build. Environ. 49 (2012) 304–309.
[2] J. Kim, T. Hong, J. Jeong, C. Koo, M. Kong, An integrated psychological response score of the occupants based on their activities and the indoor environmental quality condition changes, Build. Environ. 123 (2017) 66–77.
[3] F.D. Ching, C. Binggeli, Interior Design Illustrated, John Wiley & Sons, 2012.

[4] J. Kim, R. De Dear, Workspace satisfaction: the privacy-communication trade-off in open-plan offices, J. Environ. Psychol. 36 (2013) 18–26.
[5] A. Gupta, S. Satkin, A.A. Efros, M. Hebert, From 3d scene geometry to human workspace, IEEE Computer Vision and Pattern Recognition, 2011.
[6] M. Savva, A.X. Chang, P. Hanrahan, M. Fisher, M. Nießner, Scenegrok: inferring action maps in 3d environments, ACM Trans. Graph. 33 (6) (2014) 212:1–212:10.
[7] M. Fisher, M. Savva, Y. Li, P. Hanrahan, M. Nießner, Activity-centric scene synthesis for functional 3d scene modeling, ACM Trans. Graph. 34 (6) (2015).
[8] R. Ma, H. Li, C. Zou, Z. Liao, X. Tong, H. Zhang, Action-driven 3d indoor scene evolution., ACM Trans. Graph. 35 (6) (2016) 173–181.
[9] M. Savva, A.X. Chang, P. Hanrahan, M. Fisher, M. Nießner, Pigraphs: learning interaction snapshots from observations, ACM Trans. Graph. 35 (4) (2016) 139.
[10] V.G. Kim, S. Chaudhuri, L. Guibas, T. Funkhouser, Shape2pose: human-centric shape analysis, ACM Trans. Graph. 33 (4) (2014) 120:1–120:12.
[11] Y. Zheng, H. Liu, J. Dorsey, N. Mitra, Ergonomics-inspired reshaping and exploration of collections of models, IEEE Transactions on Visualization and Computer Graphics PP (99) (2015). 1–1
[12] Q. Fu, X. Chen, X. Su, H. Fu, Pose-inspired shape synthesis and functional hybrid, IEEE Trans. Vis. Comput. Graph. 23 (12) (2017) 2574–2585.
[13] H. Grabner, J. Gall, L. Van Gool, What makes a chair a chair? in: IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 1529–1536.
[14] K. Xu, H. Huang, Y. Shi, H. Li, P. Long, J. Caichen, W. Sun, B. Chen, Autoscanning for coupled scene reconstruction and proactive object analysis, ACM Trans. Graph. 34 (6) (2015) 177:1–177:14.
[15] S. Qi, Y. Zhu, S. Huang, C. Jiang, 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.
[16] R. Hu, C. Zhu, O. van Kaick, L. Liu, A. Shamir, H. Zhang, Interaction context (icon): towards a geometric functionality descriptor, ACM Trans. Graph. 34 (4) (2015) 83:1–83:12.
[17] R. Hu, O. van Kaick, B. Wu, H. Huang, A. Shamir, H. Zhang, Learning how objects function via co-analysis of interactions, ACM Trans. Graph. (TOG) 35 (4) (2016) 47.
[18] R. Hu, W. Li, O. Van Kaick, A. Shamir, H. Zhang, H. Huang, Learning to predict part mobility from a single static snapshot, ACM Trans. Graph. 36 (6) (2017) 227.
[19] S. Pirk, V. Krs, K. Hu, S.D. Rajasekaran, H. Kang, Y. Yoshiyasu, B. Benes, L.J. Guibas, Understanding and exploiting object interaction landscapes, ACM Trans. Graph. 36 (4) (2017).
[20] A. Mao, H. Zhang, Z. Xie, M. Yu, Y. Liu, Y. He, Automatic sitting pose generation for ergonomic ratings of chairs, IEEE Transactions on Visualization and Computer Graphics (2019). 1–1
[21] T. Liu, S. Chaudhuri, V.G. Kim, Q. Huang, N.J. Mitra, T. Funkhouser, Creating consistent scene graphs using a probabilistic grammar, ACM Trans. Graph. 33 (6) (2014) 211:1–211:12.
[22] T. Shao, A. Monszpart, Y. Zheng, B. Koo, W. Xu, K. Zhou, N.J. Mitra, Imagining the unseen: stability-based cuboid arrangements for scene understanding, ACM Trans. Graph. 33 (6) (2014) 209:1–209:11.
[23] K. Xu, R. Ma, H. Zhang, C. Zhu, A. Shamir, D. Cohen-Or, H. Huang, Organizing heterogeneous scene collections through contextual focal points, ACM Trans. Graph. 33 (4) (2014) 35:1–35:12.
[24] Q. Fu, X. Chen, X. Wang, S. Wen, B. Zhou, H. Fu, Adaptive synthesis of indoor scenes via activity-associated object relation graphs, ACM Trans. Graph. 36 (6) (2017) 201.
[25] M. Fisher, P. Hanrahan, Context-based search for 3d models, ACM Trans. Graph. 29 (6) (2010) 182:1–182:10.
[26] M. Fisher, M. Savva, P. Hanrahan, Characterizing structural relationships in scenes using graph kernels, ACM Trans. Graph. 30 (4) (2011) 34:1–34:12.
[27] Y. Liang, F. Xu, S. Zhang, Y. Lai, T. Mu, Knowledge graph construction with structure and parameter learning for indoor scene design, Comput. Vis. Media 4 (2) (2018) 123–137.
[28] S.H. Zhang, S.K. Zhang, Y. Liang, P. Hall, A survey of 3d indoor scene synthesis, J. Comput. Sci. Technol. 34 (3) (2019) 594–608.
[29] L.-F. Yu, S.K. Yeung, C.-K. Tang, D. Terzopoulos, T.F. Chan, S. Osher, Make it home: automatic optimization of furniture arrangement, ACM Trans. Graph. 30 (4) (2011) 86.
[30] P. Merrell, E. Schkufza, Z. Li, M. Agrawala, V. Koltun, Interactive furniture layout using interior design guidelines, ACM Trans. Graph. 30 (4) (2011) 87.
[31] W. Wu, L. Fan, L. Liu, P. Wonka, Miqp-based layout design for building interiors, in: Computer Graphics Forum, vol. 37, Wiley Online Library, 2018, pp. 511–521.
[32] M. Fisher, D. Ritchie, M. Savva, T. Funkhouser, P. Hanrahan, Example-based synthesis of 3d object arrangements, ACM Trans. Graph. 31 (6) (2012) 135:1–135:11.
[33] M. Cheng, Q. Hou, S. Zhang, P.L. Rosin, Intelligent visual media processing: when graphics meets vision, J. Comput. Sci. Technol. 32 (1) (2017) 110–121.
[34] X. Chen, J. Li, Q. Li, B. Gao, D. Zou, Q. Zhao, Image2scene: transforming style of 3d room, in: Proceedings of the 23rd Annual ACM Conference on Multimedia Conference, ACM, 2015, pp. 321–330.
[35] W. Xi, X. Chen, Reconstructing piecewise planar scenes with multi-view regularization, Comput. Vis. Media 5 (4) (2019) 337–345.
[36] K. Wang, M. Savva, A.X. Chang, D. Ritchie, Deep convolutional priors for indoor scene synthesis, ACM Trans. Graph. 37 (4) (2018) 70.
[37] M. Li, A.G. Patil, K. Xu, S. Chaudhuri, O. Khan, A. Shamir, C. Tu, B. Chen, D. Cohen-Or, H. Zhang, Grains: generative recursive autoencoders for indoor scenes, ACM Trans. Graph. 38 (2) (2019) 12.
[38] Tilley, R. Alvin, The Measure of Man and Woman: Human Factors in Design, John Wiley & Sons, 2001.
[39] W.J. Tam, F. Speranza, S. Yano, K. Shimono, H. Ono, Stereoscopic 3d-tv: visual comfort, IEEE Trans. Broadcast. 57 (2) (2011) 335–346.
[40] S. Hygge, I. Knez, Effects of noise, heat and indoor lighting on cognitive performance and self-reported affect, J. Environ. Psychol. 21 (3) (2001) 291–299.
[41] N. Fish, M. Averkiou, O. van Kaick, O. Sorkine-Hornung, D. Cohen-Or, N.J. Mitra, Meta-representation of shape families, ACM Trans. Graph. 33 (4) (2014) 34:1–34:11.