

# Tailored Reality: Perception-aware Scene Restructuring for Adaptive VR Navigation

ZHI-CHAO DONG, University of Science and Technology of China, China

WENMING WU, University of Science and Technology of China, Hefei University of Technology, China

ZENGHAO XU, University of Science and Technology of China, China

QI SUN, New York University, USA

GUANJIE YUAN, LIGANG LIU, and XIAO-MING FU, University of Science and Technology of China, China

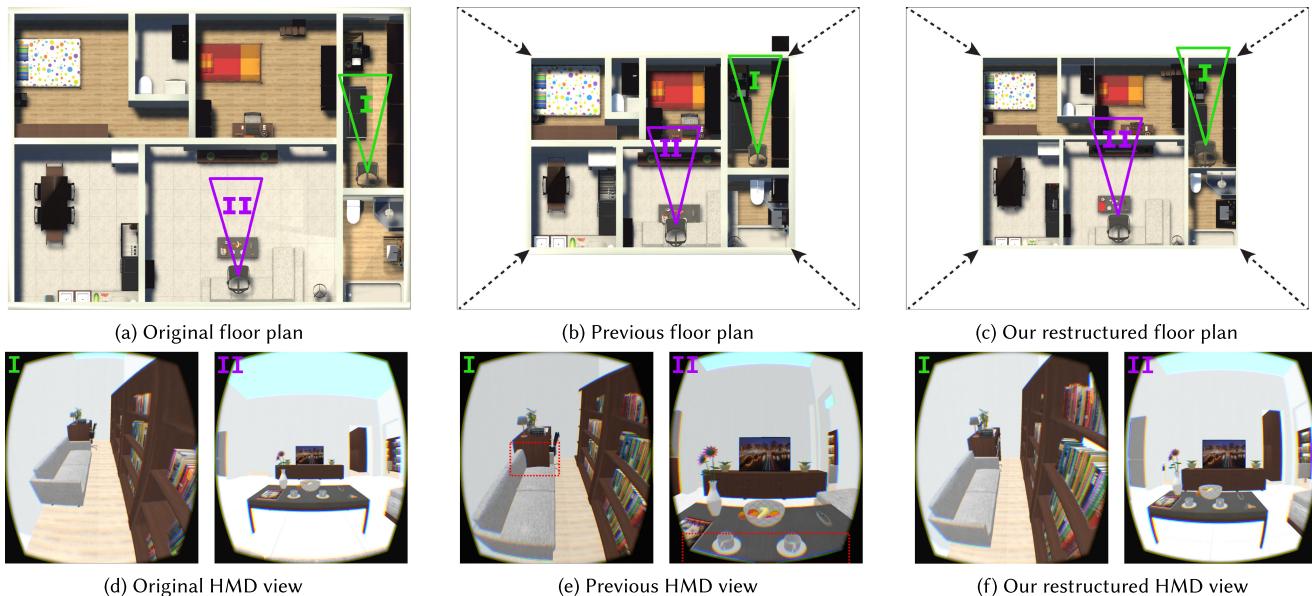


Fig. 1. Illustration of our system. Tailored for first-person view VR applications, we propose a novel perception-aware scene retargeting method. It restructures a given virtual scene (a) for users' small and varied physical spaces (valid area in (b) and (c)). Our technique preserves not only geometric and topological properties but also perceptual fidelity with the immersive first-person view. It addresses the commonly existing collision problems from previous method. For instance the red areas in (e). This can be visualized by comparing the sampled HMD views between Huang et al. [2016] (e) and ours (f).

In virtual reality (VR), the virtual scenes are pre-designed by creators. Our physical surroundings, however, comprise significantly varied sizes, layouts, and components. To bridge the gap and further enable natural navigation, recent solutions have been proposed to redirect users or recreate the virtual content. However, they suffer from either interrupted experience or distorted appearances. We present a novel VR-oriented algorithm that

automatically restructures a given virtual scene for a user's physical environment. Different from the previous methods, we introduce neither interrupted walking experience nor curved appearances. Instead, a perception-aware function optimizes our retargeting technique to preserve the fidelity of the virtual scene that appears in VR head-mounted displays. Besides geometric and topological properties, it emphasizes the unique first-person view perceptual factors in VR, such as dynamic visibility and objectwise relationships. We conduct both analytical experiments and subjective studies. The results demonstrate our system's versatile capability and practicability for natural navigation in VR: It reduces the virtual space by 40% without statistical loss of perceptual identicity.

**CCS Concepts:** • Computing methodologies → Computer graphics; Graphics systems and interfaces; Virtual reality;

**Additional Key Words and Phrases:** Virtual reality, perception, retargeting

#### ACM Reference format:

Zhi-Chao Dong, Wenming Wu, Zenghao Xu, Qi Sun, Guanjie Yuan, Ligang Liu, and Xiao-Ming Fu. 2021. Tailored Reality: Perception-aware Scene Restructuring for Adaptive VR Navigation. *ACM Trans. Graph.* 40, 5, Article 193 (October 2021), 15 pages.

<https://doi.org/10.1145/3470847>

This work is supported by the National Natural Science Foundation of China (61802359 and 62025207), Zhejiang Lab (NO.2019NB0AB03), and USTC Research Funds of Double First-Class Initiative (YD0010002003).

Authors' addresses: Z.-C. Dong, Z. Xu, G. Yuan, L. Liu, and X.-M. Fu (corresponding author), University of Science and Technology of China, China; emails: {dzc206, crazer, yqj}@mail.ustc.edu.cn, {lgliu, fxum}@ustc.edu.cn; W. Wu, University of Science and Technology of China, Hefei University of Technology, China; email: wwmng@ mail.ustc.edu.cn; Q. Sun, New York University, USA; email: qisun0@gmail.com.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2021 Association for Computing Machinery.

0730-0301/2021/10-ART193 \$15.00

<https://doi.org/10.1145/3470847>

## 1 INTRODUCTION

**Virtual reality (VR)** means to visually and naturally simulate varied impossible virtual worlds while users remain in limited physical spaces. Besides the graphical context, the “simulation” shall also include natural interactions such as navigation and locomotion, whereas the limited physical spaces are a fundamental challenge from the broad deployment of VR facilities: Due to the masked vision from the surroundings, VR users commonly experience risks hitting obstacles with navigations driven by natural locomotion.

Recent advancements in **redirected walking (RDW)** techniques have demonstrated promising possibilities for addressing this problem. The main ideas are to subtly redirect camera motion [Langbehn and Steinicke 2018; Nilsson et al. 2018] or alternate geometric appearance [Dong et al. 2017, 2019; Sun et al. 2016]. However, the former solution could introduce discontinuous experience when the virtual scenes are significantly larger: When the users are close to obstacles, the unnoticeable redirection gains may be insufficient. Then, the virtual camera position has to be reset, resulting in sudden and unexpected teleportation. For the latter, the alternated geometries result in visually distorted appearances, especially with regularly shaped objects. Moreover, geometry manipulation is typically only applicable to highly occluded road-alike scenes [Sun et al. 2016].

However, the architectural scene retargeting improves creators’ productivity by automatically transferring a given three-dimensional (3D) scene to varied layouts and sizes. It preserves the creative intention such as spatial connectivity [Huang et al. 2016]. For design purposes, these methods typically apply to third-person bird-views to expose the whole scene details with the least occlusions and number of views on a monitor. VR **head-mounted displays (HMDs)** and renderings, in comparison, are to simulate the users’ own eyes inside the scene. Under such immersive settings, the experience is first-person-view dominated and limited by physical obstacles with users walk to navigate.

So, intelligently re-creating the virtual space while preserving its visual and spatial perception to VR users remains an open yet demanding problem. In this article, we present a perception-aware algorithm that addresses the challenge. Besides topologies and geometries, biopsychosocial studies on human visual behaviors also shed light on our approach that measures factors such as scale [Park et al. 2014], structure [Zelinsky 2013], distance [Henderson et al. 2008], and visibility [Henderson et al. 2011]. These ensure perceptually similar visual/locomotive/interactive experience to be perceived by VR users after the restructuring. Mathematically, we seek the optimized manipulator via a series of analytical modeling, user-centered view sampling, and hierarchical optimization. The optimal restructured scene is then rendered in users’ HMD to enable natural walk-through within a limited physical space. Our system does not introduce shape distortion or interrupting, the two problems currently faced by RDW.

We conducted subjective studies, numerical analysis, usability validations, and pressure tests to evaluate the system’s capability and extendability. They demonstrate that users perceive the restructured scene as similar to the original and complete real-world VR tasks more efficiently and comfortably than existing scene re-

structuring and RDW approaches. Various scenes can be deployed to home-sized space with the compression ratio of 40%, providing an interruption-free real walking experience. This is achieved by preserving both structures and visual similarities from a first-person perspective. Aiming at popularizing the currently space-sensitive VR applications, we contribute the following:

- an end-to-end VR system enabling users to continuously walk through a perceptually similar virtual scene in a small physical space, without interruption or distorted appearance
- an automatic scene restructuring algorithm inspired by biopsychosocial discoveries and tailored for first-person VR; it preserves structural/geometric properties and human spatial scene perception
- systematical validations with computational analysis, psychophysical studies, and real-world usability/stress tests.

To our best knowledge, this is the first scene retargeting method tailored for immersive locomotion and preserving both structural and (first-person) perceptual similarities. Figure 1 shows an example of our system. We will release the source code to inspire follow-up research.

## 2 RELATED WORK

### 2.1 Real Walking in VR

Walking-based navigation is one of the ultimate goals for interactive immersion in VR. However, it has not been fully realized due to the naturally mismatched sizes and shapes between virtual and physical scenes. To address this problem, researchers have deployed two major types of means to redirect VR users’ walking path to save physical space usage.

Rotational and translational gains [Azmandian et al. 2017; Nilsson et al. 2018; Suma et al. 2011], blink [Langbehn et al. 2018], or saccade [Sun et al. 2018] change blindness introduce opportunities of subtly redirecting virtual camera motion without being noticed. They have been broadly adapted in recent VR research. However, to prevent significant sickness, these gains are controlled below the noticeable thresholds. Consequently, the spatial saving ratio remains low. That is, the physical spaces must still be large enough for users to walk through a virtual scene. As a tradeoff, frequent interruptions from resetting or teleportation is introduced, thereby significantly lowering the immersion. Our restructuring approach preserves spatial perception without differing camera changes.

Geometric distortions have been introduced to reduce the interruptions. In [Dong et al. 2017, 2019; Sun et al. 2016], users directly perceive a different virtual scene. The main idea, is to warp the virtual scene floor plan to fit into a given physical space, with as subtle as possible distortion. However, the distortions indiscriminately apply to not only walkable grounds but also regularly-shaped objects. In comparison, our method only restructures the scene without sacrificing objectwise appearance.

Real walking in limited physical space can also be realized by manipulating the scene structure. For instance, Suma et al. [2012] compressed virtual indoor environments into smaller physical spaces by tiling individual rooms unnoticeably. Vasylevska et al. [2013] and Vasylevska and Kaufmann [2017] proposed a

flexible space that dynamically generates indoor scenes and further demonstrated the effectiveness of longer, smoothly curved corridors with more corners. In contrast, our goal is to perceptually preserve visual experience to arbitrary scenes regardless of the original design types.

## 2.2 Retargeting Images and 3D Scenes

Image retargeting is an essential topic in computer graphics and vision [Shamir and Sorkine 2009; Vaquero et al. 2010]. The goal is to preserve the salient objects without visual artifacts. So far, a variety of content-aware methods have been presented with the following: (1) discrete solutions [Avidan and Shamir 2007; Mansfield et al. 2010; Rubinstein et al. 2008, 2009] that remove a seam of minimal importance to preserve image content or (2) continuous approaches [Gal et al. 2006; Jin et al. 2010; Wang and Lai 2009; Wang et al. 2008] that optimize a mapping between the source and target images, driven by saliency-based distortion importance. Rubinstein et al. [2010] presented a comprehensive perceptual study and analysis of the state of the art.

With 3D models, a non-homogeneous resizing method preserving the main characteristic features and structures has been proposed in Kraevoy et al. [2008]. The geometric approach focuses on a single model. Whereas, navigation involves entire 3D scenes including inter-object connectivities. By considering that Lin et al. [2011] and Huang et al. [2016] presented interactive methods that re-create new scenes via preserving structural styles or inserting/removing discrete patterns. These approaches synthesize irregular architectural models by reusing the structural elements from the input. However, we do not seek a global and generic scene restructuring algorithm, but instead a VR-tailored retargeting method for adaptive navigation with varied physical spaces. That is, we introduce perception-aware metrics to maintain the perceptual fidelity of 3D scenes inside HMD’s first-person views. Moreover, existing solutions rely on repetitive structural patterns or uniform layouts [Huang et al. 2016], limiting the applicable scenarios.

Approaches have also been proposed to re-create a plausible layout from discrete elements. With user-specified physical spaces and virtual elemental relationships, Scenograph [Marwecki and Baudisch 2018] adaptively generates corresponding scenes while maintaining narrative structures for walk-through. The method mandatorily depends on a pre-defined narrative storyline and user intervention. Common designer-created 3D scenes, however, may only contain static geometry properties than the required semantic and temporal information. Yu et al. [2011] extract spatial relationships on the furniture placement from indoor scenes samples. Then, the system encodes the knowledge into priors that are later used to synthesize novel scenes. Hartmann et al. [2015] synthesizes new polygonal layouts by employing a deformation to an original domain. Similarly to Huang et al. [2016], the system aims at design applications. Under this scenario, the generated contents are presumed to be visualized with third-person perspectives and minimal occlusions. Immersive medium, in comparison, usually displays content with a user’s eyes as the virtual camera, resulting in significantly different spatial understanding and strong occlusion from local-only views. Existing scene retargeting methods

mainly optimize the spatial positions and scale of objects for different application purposes such as scene design/visualization than immersive walk-through. Thus, we are motivated to propose an automatic scene retargeting approach tailored for VR walk-through with minimal user-specification by tailoring for first-person-view dynamic visibilities and spatial perception.

## 2.3 Scene Perception

Preserving the perception of a virtual scene requires the understanding of how the human understands spaces. Moser et al. [2008] revealed how creatures perceive the space at the cell level. Further, human scene perception has been studied with biomedical data [Epstein and Baker 2019]. Psychologically, our scene perception involves both human visual behaviors and inter-object identification [Henderson and Hollingworth 1999; Konkle et al. 2010]. Our geometric algorithm considers these factors by minimizing perceived changes. These biological and psychophysical understandings motivate the computer applications in scene recognition [Zhou et al. 2014] and modeling [Oliva and Torralba 2001]. However, these human-oriented research has not been well-considered in VR scenarios, where the first-person view users play the main role in the whole experience. Vasylevska and Kaufmann [2017] thus explore means of compressing a space while preserving the perception of spatial arrangement. Its scope lies in regular shapes and indoor scenes that share structures within several categories. Our method applies to generic virtual spaces without limited by pre-defined structures, with the consideration of scale, objectwise relationships, and visibility.

## 3 METHOD

In this section, we first define the problem in Section 3.1, followed by the novel retargeting approach in Section 3.2 and Section 3.3.

### 3.1 Problem Definition and Goal

*Input.* The input consists of a virtual ( $\mathcal{V}$ ) and physical ( $\mathcal{P}$ ) floor plans.  $\mathcal{V}$  contains a set of salient objects,  $\mathcal{S} \triangleq \{s_i\}$ . As a non-trivial condition, we assume the area of  $\mathcal{P}$  is smaller than that of  $\mathcal{V}$ .

*Goals & challenges.* Our goals are twofold. First, users walk freely in  $\mathcal{P}$  as in real life for exploring  $\mathcal{V}$  without collision, camera teleportation, or visual warning. Second, the system shall not introduce the geometric curving of the salient objects.

Inspired by the retargeting techniques [Huang et al. 2016; Shamir and Sorkine 2009; Vaquero et al. 2010], we seek a restructuring mapping  $\Phi$  so that each mapped salient object  $\Phi(s_i)$  lies within the physical floor plan  $\mathcal{P}$ . Then, users walk freely in  $\mathcal{P}$  to navigate through a virtual scene, which contains  $\Phi(s_i), \forall s_i$  and is perceptually similar to the input virtual scene. However, we face two challenges to achieve this goal. First, it is challenging to incorporate first-person-view constraints in computing  $\Phi$  for VR navigation. Second, generic numerical optimizations are unable to satisfy the hard constraint  $\Phi(s_i)$  lies within the physical floor plan  $\mathcal{P}$  for all  $s_i$  without introducing appearance distortion. To this end, we propose a VR-tailored energy function. It seeks  $\Phi$  by rearranging the salient objects of  $\mathcal{V}$  so that users perceive no curved  $s_i$  and are allowed to walk freely in  $\mathcal{P}$ .

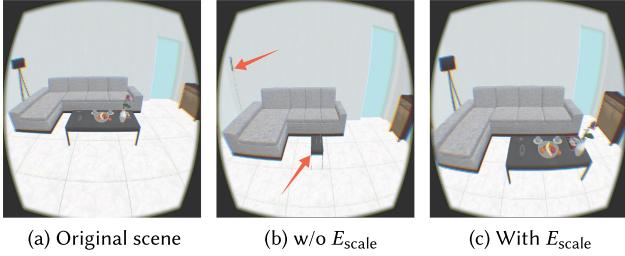


Fig. 2. An ablation study of Equation (2). (a) A sampling view from the original scene. ((b) and (c)) The same viewpoint from the retargeted scene without/with  $E_{\text{scale}}$ . Severe artifacts of shape distortion are created in (b) without the scale-preserving term, such as the almost disappeared floor lamp, and the severely deformed coffee table with no objects on it.

### 3.2 Retargeting

*Representations.* The key idea, of computing  $\Phi$  is to move the salient objects in  $\mathcal{V}$  while preserving visual and perceptual similarities. For convenience, each  $s_i$  is represented as its enclosed bounding box containing three parameters: (i) center  $c_i$ , (ii) length  $l_i$ , and (iii) width  $w_i$  (see the right inset). We minimize the following energy function to determine an optimal placement  $S^* = \{s_i^*\}$ :

$$\{s_i^*\} = \arg \min_{\{s_i\}} E_r(\{s_i\}), \quad (1)$$

where  $E_r$  is the weighted sum of terms described below.

*Preserving scale perception.* Size is an active cue when we perceive a scene [Park et al. 2014]. So, to obtain a similar yet smaller virtual environment, we need to reshape objects through resizing. Hence, we target a non-uniform scaling transformation with a limited range to minimize the visual shape distortion. The scale-preserving term is introduced to avoid generating over- or under-sized objects, as demonstrated in Figure 2. The total scale-preserving energy,  $E_{\text{scale}}$ , is the sum of all the per-object terms:

$$E_{\text{scale}} = \sum_{\{i\}} E_{l,s_i} + E_{w,s_i}, \quad (2)$$

where  $E_{l,s_i}$  and  $E_{w,s_i}$  preserve objectwise length and width, respectively. Denote the scaling factor of length (or width) for  $s_i$  as  $\rho_i = l_i/l_i^0$  (or  $\varrho_i = w_i/w_i^0$ ), where  $l_i^0$  (or  $w_i^0$ ) is the length (or width) of the input  $s_i$ . We encourage  $\rho_i$  and  $\varrho_i$  to be constrained in a range:

$$\begin{aligned} E_{l,s_i} &= \varphi\left(\frac{\rho_i}{\rho_{\min}} - 1\right) + \varphi\left(1 - \frac{\rho_i}{\rho_{\max}}\right), \\ E_{w,s_i} &= \varphi\left(\frac{\varrho_i}{\varrho_{\min}} - 1\right) + \varphi\left(1 - \frac{\varrho_i}{\varrho_{\max}}\right), \end{aligned} \quad (3)$$

where  $\rho_{\min}$ ,  $\rho_{\max}$ ,  $\varrho_{\min}$ , and  $\varrho_{\max}$  are the corresponding minimum and maximum scaling factors for two directions. When scaling factors are constrained, the aspect ratio of the scaling is also theoretically controlled. These parameters are related to the ratio of space compression and can be specified by users according to the experimental requirements. Since the area of  $\mathcal{P}$  is smaller than that of  $\mathcal{V}$ , we set the maximum scaling factor to 1, ensuring scene ob-

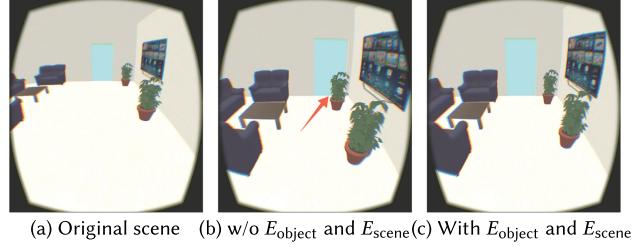


Fig. 3. An ablation study of Equation (4) and Equation (5). (a) A sampling view from the original scene. ((b) and (c)) The same viewpoint from the retargeted scene without/with  $E_{\text{object}}$  and  $E_{\text{scene}}$ . The structure-preserving term preserves the layout. The office door is blocked by the plant in (b), which will affect walking in the scene.

jects not to be enlarged. In our experiment,  $\rho_{\min} = \varrho_{\min} = 0.8$ ,  $\rho_{\max} = \varrho_{\max} = 1.0$ . The barrier penalty function  $\varphi$  is as follows:

$$\varphi(x) = \frac{1}{x + \sqrt{x^2 + \epsilon^2}}.$$

Here  $\epsilon$  is a small increment. The intuition of the barrier penalty is visualized beside the function. It prevents the scaling from exceeding the allowed range. Without it, strong shape distortion is inevitable. For example, the coffee table in Figure 2(b) is over-deformed to support objects on it, and the floor lamp is severely distorted.

*Preserving structural perception.* We perceive a scene starting at the level of objectwise relationships [Zelinsky 2013]. Our next retargeting goal is preserving structures in *object-level* and respecting the original design arrangement/layout in *scene-level*. To compute a new layout that preserves the original structure, we first perform a uniform scaling on  $\mathcal{V}$  so that the scaled  $\mathcal{V}$  (denoted as  $\mathcal{V}_s$ ) is fully inside  $\mathcal{P}$ . The scale is computed as  $\min\{l_p/l_v, w_p/w_v\}$ , where  $l_v$  and  $w_v$  (or  $l_p$  and  $w_p$ ) are the length and width of  $\mathcal{V}$  (or  $\mathcal{P}$ ), respectively. We then utilize  $\mathcal{V}_s$  as a guiding layout. The structure-preserving term includes an object-level loss and a scene-level loss as in Figure 3. One of the plants is incorrectly placed in front of the office door, impeding indoor walking. In the object-level energy, we encourage that similar locating in  $\mathcal{V}_s$  for each salient object:

$$E_{\text{object}} = \sum_{\{i\}} e^{s[\|c_i - c_{i,s}\|_2^2 - R]}, \quad (4)$$

where  $s$  and  $R$  are two constants and  $c_{i,s}$  is the center of the scaled  $s_i$  in  $\mathcal{V}_s$ .

$s$  is a penalty parameter. Since the exponential function in  $E_{\text{object}}$  increases monotonically, a larger  $s$  will lead to a stronger attraction between  $c_i$  and  $c_{i,s}$ . Since  $e^0 = 1$ , when  $\|c_i - c_{i,s}\|_2^2 = R$ ,  $E_{\text{object}} = 1$ . In practice,  $R$  is used to penalize the distance between  $c_i$  and  $c_{i,s}$ . A larger  $R$  indicates less constraint on distance.

We set  $s$  to 1.0 and  $R$  to 0.5 m in our experiments.

At the scene level, we ensure consistent adjacent connectivity among each salient object in  $\mathcal{V}_s$ . That is, the mutual relationship between  $s_i$  should be preserved. We encourage the distance

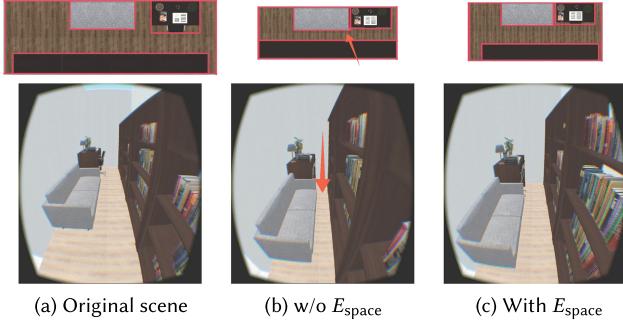


Fig. 4. An ablation study of Equation (6). (a) A sampling view from the original scene. ((b) and (c)) The same viewpoint from the retargeted scene without/with  $E_{\text{space}}$ . Top-down views are also shown in the top row. As shown in (b), the road in the study room is too narrow to walk without the space-preserving term, which make it impossible to walk in the scene.

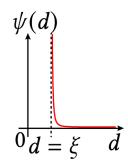
between two objects to be close to the correspondence in  $\mathcal{V}_s$ ,

$$E_{\text{scene}} = \sum_{\{s_i\}} \sum_{s_j \in \Omega(s_i)} \frac{\|\mathbf{c}_i - \mathbf{c}_j\|_2^2}{\|\mathbf{c}_{i,s} - \mathbf{c}_{j,s}\|_2^2} + \frac{\|\mathbf{c}_{i,s} - \mathbf{c}_{j,s}\|_2^2}{\|\mathbf{c}_i - \mathbf{c}_j\|_2^2}. \quad (5)$$

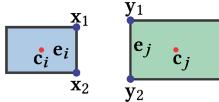
Here,  $\Omega(s_i) = \{s_j | \|\mathbf{c}_{i,s} - \mathbf{c}_{j,s}\|_2 \leq \epsilon_d\}$ , where  $\epsilon_d$  is a distance threshold for defining neighborhoods. We set  $\epsilon_d$  to 3.0 m in our experiments.

*Preserving spatial perception.* Similarly to scales, studies also show the critical role of spatial distance when we perceive a scene [Henderson et al. 2008]. Walking space is a critical factor for both the continuous walking experience and the perceived relationships among objects. A remarkable example is shown in Figure 4. Without the space-preserving term, the main walkable virtual areas (i.e., the ground floor areas excluding objects, or being manually specified by the designer) may be too narrow, as shown in Figure 4(b). Users may have difficulties walking through the study room. Therefore, we constrain the walkable width to be greater than  $d_{\text{width}}$ . That is, the distance between any two objects ( $s_i, s_j$ ) on both sides of the road should not exceed  $d_{\text{width}}$ . Here, the roads are areas that users can walk through in the virtual scene. It can either be manually specified according to designers' intentions or automatically defined as the ground area excluding placed objects. To this end, a barrier function  $\psi$  is introduced:

$$\psi(d) = \begin{cases} \left( \max(0, \frac{\epsilon}{d-\xi} - 1) \right)^2, & \text{if } d > \xi, \\ +\infty, & \text{otherwise.} \end{cases}$$



Here  $\epsilon$  and  $\xi$  are two positive parameters. If  $d - \xi > \epsilon$ , then  $\psi(d) = 0$ . The sides of  $s_i$  and  $s_j$  on the road are denoted as  $\mathbf{e}_i = \overline{x_1 x_2}$  and  $\mathbf{e}_j = \overline{y_1 y_2}$  (see the right inset). We define their distance  $d_{i,j}$  as



$$d_{i,j} = \min\{d_{p2s}(x_1, e_j), d_{p2s}(x_2, e_j), d_{p2s}(y_1, e_i), d_{p2s}(y_2, e_i)\},$$

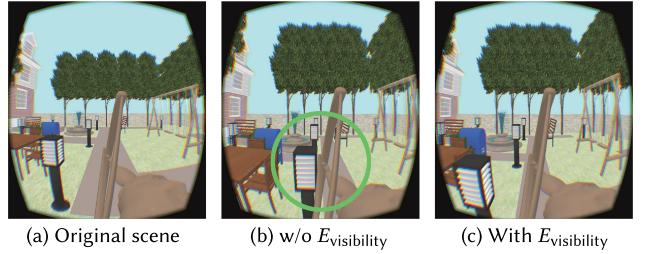


Fig. 5. An ablation study of the visibility-preserving term  $E_{\text{visibility}}$ . (a) A sampling view from the original scene. ((b) and (c)) The same viewpoint from the retargeted scene without/with  $E_{\text{visibility}}$ . In the original scene (a), the arrow of the sculpture is designed to point to the middle tree in a row of trees. This design intent is maintained using the visibility-preserving term, as shown in (c). An unsatisfactory result is shown in (b), the arrow incorrectly points to the street light.

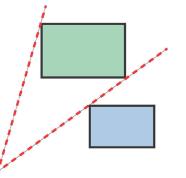
where  $d_{p2s}$  is the point-line distance. The energy for constraining the road width is as follows:

$$E_{\text{space}} = \sum_{(s_i, s_j)} \psi(d_{i,j}), \quad (6)$$

where  $\xi = d_{\text{width}}$  and  $\epsilon = 10$  in our experiments.  $d_{\text{width}}$  can be specified by users, and it is always related to the size of physical space. In our experiment, it is set to 0.45 m.

To circumvent overlaps among elements, we introduce a collision-avoidance term to encourage separation between two adjacent objects. The collision avoidance term has a similar form as Equation (6) with the width set to 0. We continue to utilize the aforementioned guidance layout  $\mathcal{V}_s$ . The roads in  $\mathcal{V}_s$  serve as an initial setting for retargeting. If the initial  $d_{i,j}$  is less than  $d_{\text{width}}$ , then we zoom out those objects while fixing the bounding box of  $\mathcal{V}_s$  until  $d_{i,j} > d_{\text{width}}$ . To guarantee overlap-free, our line search explicitly enforces the satisfaction of road width and collision avoidance constraints during optimization.

*Preserving perceptual visibility.* Visibility is a main component in computer graphics, VR (a first-person view dominated visual media), and film story-telling. Besides these scenarios, the cue of openness also drives the perception of the real-world [Henderson et al. 2011; Hillier and Penn 1991; Peponis et al. 2004]. The role of a physical layout is closely tied to visibility. To avoid missing design intentions, we preserve the visibility of both major areas and objects, as shown in Figure 5. For instance, in Figure 5(b), the arrow incorrectly points to the street light instead of the middle tree in a row of trees. However, a comprehensive visibility analysis of the virtual scene could be impossible due to the infinite number of viewpoints and user-controlled camera orientations. The visibility here is defined based on the line-of-sight between the observer positions and objects. For example, the TV or bookshelf in Figure 1 should be fully visible when the users sit on the sofas while facing them. The observer positions and objects are sampled within  $\mathcal{S}$  based on their semantics (e.g., the frequency of human activity and the importance of object) in reality. We represent the viewshed as a triangle with a vertex located at the observer position (see the



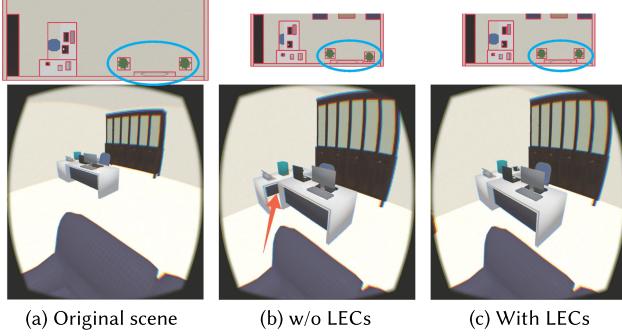


Fig. 6. An ablation study of linear equality constraints LECs. (a) A sampling view from the original scene. ((b) and (c)) The same viewpoint from the retargeted scene without/with LECs. Top-down views are also shown in the top row. Two tables are not aligned in (b) while the alignment is achieved using linear equality constraints in (c).

up inset). Then we adopt the same energy form as the road width term in Equation (6) with a varied threshold  $\epsilon$  as the visibility energy  $E_{\text{visibility}}$ . During retargeting, we preserve the (in)visibility of these objects by keeping them (in)outside the corresponding viewshed frustum. In our experiments, the designer's salient objects and sampling views are specified according to the design intent. Automatically analyzing scenes to weight samples with visibility importance is orthogonal yet compatible with the current system.

**Linear equality constraints.** Finally, we add several **linear equality constraints (LECs)** to visually preserve details, as in Figure 6. We first impose alignment constraints. Physical environments often contain functional object groups or regularly shaped organization, named superstructure. One of the most important superstructures is alignment (see the up inset). It is defined by a sequence of salient objects arranged along a line (e.g., a row of bookshelves against a wall). To build the alignment constraints, we first detect any two or more objects aligned by a line intersecting one of their edges. For simplification, we only consider two axis-aligned directions in this article. The alignment constraints are achieved by enforcing equal coordinates in the aligned direction.

Then, the adjacency constraints are used. Adjacency is another kind of spatial configuration that is frequently used in the physical environment (see the right inset). Some functions of the objects are achieved by building adjacencies (e.g., kitchen cabinetry). It is also worth noting that some physical environments are composed of many components that are visually adjacent but physically disconnected (e.g., a wardrobe against a wall). To build the adjacency constraints, we first detect any pairs of objects whose distance between each other is less than a set threshold. In our experiment, the threshold is set to 0.01 m. Note that we treat each linear wall segment as a wall object in this article. Our goal is to maintain the adjacencies between these pairs of objects during retargeting. The adjacency constraints are satisfied by enforcing equal coordinates of the adjacent edges.

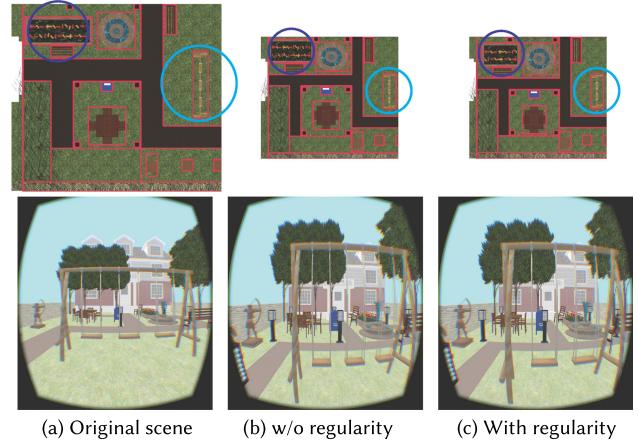


Fig. 7. An ablation study of our regularity optimization. (a) A sampling view from the original scene. ((b) and (c)) The same viewpoint from the retargeted scene without/with regularity optimization. Top-down views are also shown in the top row.

**Optimization problem.** The energy terms are weighted to form the resulting energy as follows:

$$E_r(\{s_i\}) = \omega_{\text{scale}} E_{\text{scale}} + \omega_{\text{object}} E_{\text{object}} + \omega_{\text{scene}} E_{\text{scene}} + \omega_{\text{space}} E_{\text{space}} + \omega_{\text{visibility}} E_{\text{visibility}}, \quad (7)$$

where  $\omega_{\text{scale}}$ ,  $\omega_{\text{object}}$ ,  $\omega_{\text{scene}}$ ,  $\omega_{\text{space}}$ , and  $\omega_{\text{visibility}}$  are the positive weights. The optimization problem becomes

$$\begin{aligned} & \min_{\{s_i\}} E_r(\{s_i\}) \\ \text{s.t. } & AX = b, \end{aligned} \quad (8)$$

where  $X$  denotes the variable vector that is formed by stacking all centers, lengths, and widths in column format, and  $AX = b$  indicates the linear equality constraints.

**Optimization details.** The first and second derivatives of the objective function  $E_r(\{s_i\})$  can be analytically obtained. We adopt Newton's method to solve the optimization problem benefiting from its fast convergence. In each iteration, the line search's step size ensures the constraint satisfaction of the walkable width and collision avoidance.

Our algorithm terminates when the step size is less than  $10^{-6}$  or the maximum iteration number we set is reached. The optimization process is offline and usually takes a few minutes. Please refer to the supplementary video to visualize the process. We will release the implementation of our optimization.

### 3.3 Optimizing Salient Objects

Salient objects play a main role in determining spatial perception. They appear shall be further optimized (including location, size, and components optimization) during retargeting. All salient objects are considered for further optimization. We develop two optimization mechanisms: (1) regularity and (2) hierarchy.

**Regularity.** In a complex scene, repetitive components usually exist in one salient object, such as the four-seat sofa and three rows of bookcases in Figure 1. However, we are not sensitive to the exact number of repetitive objects. Decreasing the number can

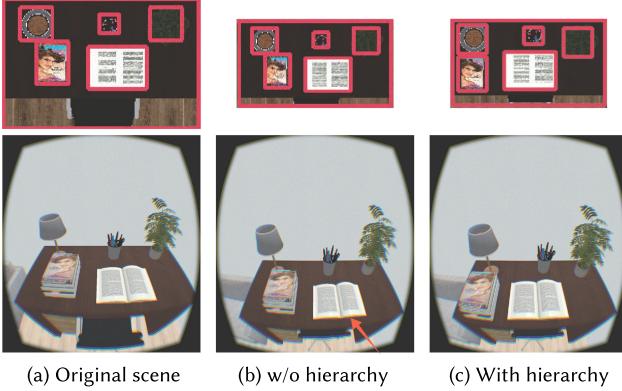


Fig. 8. An ablation study of our hierarchical optimization. (a) A sampling view from the original scene. ((b) and (c)) The same viewpoint from the retargeted scene without/with hierarchical optimization. Top-down views are also shown in the top row. With our hierarchical optimization, the object sizes resemble the input sizes while satisfying other constraints, such as structure-preserving constraints and perceptual visibility constraints. The books become very small in (b) but can almost maintain original size in (c).

reduce visual distortion, as shown in Figure 7: With one repetitive swing removed, the sizes of the others are significantly more preserved. The goal of our regularity optimization is to automatically determine the number of repetitive components.

A posterior method is proposed. We determine whether to delete repetitive components based on the obtained shape distortion after the retargeting optimization. First, we determine objects with repetitive components that are visually intensive and not related to the interaction of applications. Such objects, denote them as  $\{s^R\}$ , can be automatically detected or manually specified. We use a penalty term to minimize the distortion from retargeting:

$$P^R = \sum_{\{i\}} f_p(\rho_i, \rho_{\min}, \rho_{\max}) + f_p(\varrho_i, \varrho_{\min}, \varrho_{\max}), \quad (9)$$

$$f_p(x, x_{\min}, x_{\max}) = \begin{cases} x_{\min} - x, & \text{if } x < x_{\min}, \\ x - x_{\max}, & \text{if } x > x_{\max}, \\ 0, & \text{otherwise.} \end{cases}$$

Then, we detect and delete the repetitive components to lower the energy by traversal. The specific steps are as follows:

- (1) Conduct the retargeting optimization and calculate the penalty term  $P^R$ . Then we build a queue  $Q^R$  and push all objects of  $\{s^R\}$  into  $Q^R$ .
- (2) If  $Q^R \neq \emptyset$ , then pop an object from  $Q^R$  and denote this object as  $s^R$ . Otherwise, stop the update.
- (3) If the number of repetitive components of  $s^R$  is greater than 1, then delete a repetitive component of  $s^R$  and run the retargeting optimization with the new penalty term denoted as  $P_{\text{new}}^R$ . Otherwise, go to Step (2).
- (4) Calculate the reduction between  $P^R$  and  $P_{\text{new}}^R$ . If it exceeds a specified threshold, then accept the deletion, update  $P^R$  to  $P_{\text{new}}^R$ , and go to Step (3). Otherwise, restore the deleted repetitions of  $s^R$  and go to Step (2).

Without the automated optimization, it is difficult for users to manually adjust the number of components, since it is hard for them to predict the shape distortion.

*Hierarchy.* In some situations, the salient objects in the virtual scene can be regarded as a sub-scene, such as the desk in the bedroom, the cabinet in the kitchen, or certain areas in the park. After we finish the global optimization of the whole virtual scene, we use the same retargeting algorithm on components in this sub-scene for further optimization. This hierarchical retargeting step can further lower the distortion of small-scale objects, as shown in Figure 8. Although our retargeting optimization is currently performed in 2D, this hierarchical optimization can be regarded as performing the optimization in 3D.

### 3.4 Discussion

*Choosing salient objects.* For the maximal generality, we consider all objects in the input virtual scene as being salient while constructing  $\mathcal{S}$ . For accelerated computation, we use two positional relationships to group several objects into a set as a single unit: (a) adjacent repetitive relationship and (2) hierarchical relationship. For example, based on the adjacent repetitive relationship, the three rows of bookcases in Figure 1 are treated as a single salient object for the retargeting optimization, and then the regularity optimization is used to determine the final number of the repetitive components. According to the hierarchical relationship, a table and the objects on it are considered as one salient object, and then the hierarchy optimization is applied after the retargeting optimization. However, the subset of  $\mathcal{S}$  for constructing the objectwise visibility term in Section 3.2 was manually indicated by creators. As future work, we foresee semantic analysis may shed light on making the whole system fully automatic.

*Preserving the functionality.* In practice, directly removing the repetitions of the salient objects may potentially affect the functionality and interactivity. For instance, in Figure 1(d), deleting the repetitive part of the bookshelf may change its capacity. The designers have the freedom to identify these objects as non-deletable should the application scenarios require user interaction.

*Complex workspaces.* Our method is adaptive to the physical workspaces that are irregularly-shaped and/or containing interior obstacles. This can be achieved by introducing “ghost object” (with fixed shapes and sizes) in virtual scenes at the corresponding positions. The retargeting procedure remains unchanged. While using the system, users are guided away from the irregular physical boundaries and prevent them from colliding with these virtual obstacles. Figure 9 shows two examples.

## 4 IMPLEMENTATION

*Space and positional tracking.* The size of our physical workspace  $\mathcal{P}$  is  $5.6 \text{ m} \times 5.6 \text{ m}$ , as shown in Figure 10(a). Eight Nokov motion capture cameras locate the users with 3 markers on their heads (Figure 10). There are no physical obstacles inside  $\mathcal{P}$ .

*Rendering system.* An Oculus DK2 HMD connected to a laptop is used to render the binocular views of the retargeted virtual scene

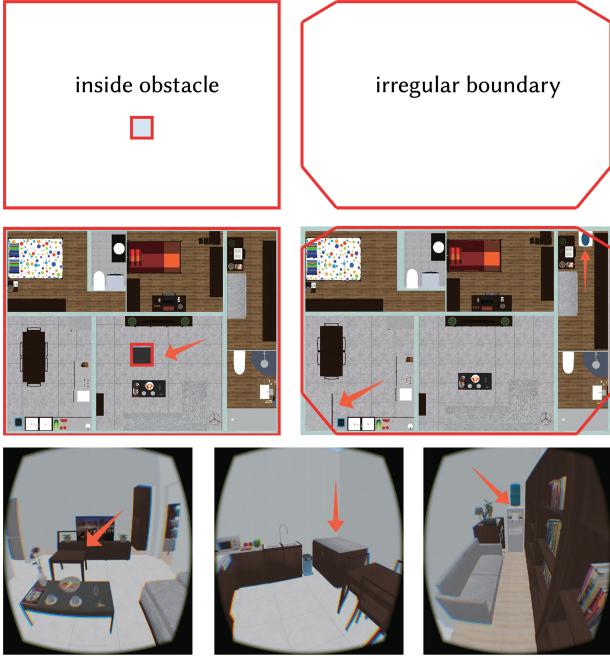


Fig. 9. Using “ghost objects” to adapt to irregularly-shaped real workspaces. Top: Two representative types of real workspaces. The left one contains an interior obstacle. The right one of non-rectangular boundaries. Middle: Corresponding retargeting results. A chair is added in the virtual scene as a ghost object to avoid the obstacle. A cupboard and a water dispenser serve as buffer areas to keep users inside the non-regular boundaries and prevent them from hitting the boundary. Bottom: The ghost objects in HMD views.



Fig. 10. Experimental setup. Eight motion capture cameras are used to track the users’ positions (a). During walking, the laptop and power supply are in the basket carried by users (b).

$\mathcal{M}$ . To ensure high rendering performance, the laptop is equipped with Intel i7-6700HQ CPU, NVIDIA GTX1070 GPU, 16GB RAM, powered by an uninterruptible power supply (Figure 10(b)).

The perceived rendering performance is decided by the tracking rate and the computation. In our implementation, the rendering images have a resolution of  $1,182 \times 1,464$ . Our tracking system provides more than 80 FPS. Overall, we achieve more than 60 FPS.

## 5 EVALUATION

We first conduct psychophysical studies in Section 5.1 to compare our system with three alternative solutions: rescaling (Section 5.1.1), existing retargeting method (Section 5.1.2) and



Fig. 11. The original indoor scene and the retargeted result using our method. Top: Floor plans of the original indoor scene and our result. Middle: several sampling HMD views from the original indoor scene. Bottom: Corresponding views from our result.

redirected walking techniques (Section 5.1.3). The first two are evaluated via **two-alternative-forced-choice (2AFC)** experiment; the third is measured via task efficiency and simulator sickness metrics. Then, we perform usability tests (Section 5.2) and algorithm evaluations (Section 5.3) to validate our capability and extendability under various challenging settings. In the following experiments, we cover various scenarios as much as possible to simulate real-world applications, including indoor/outdoor/human-made/natural spaces. Note that designers created the scenes orthogonal to our system. The elements’ sizes and positions were designed to reflect their physical realities.

### 5.1 Comparisons

**Scene.** For consistency, Section 5.1.1 to Section 5.1.3 share a virtual scene input: As shown in Figure 11, we applied a  $9 \text{ m} \times 12 \text{ m}$  indoor environment. It consists of a living room, a kitchen, a bathroom, a study room, and a master bedroom to simulate an arbitrary virtual home space. We thus added furniture and daily supplies into the scene. All models were properly scaled and placed. The indoor scene is compressed into a  $5.85 \text{ m} \times 7.8 \text{ m}$  space by our retargeting optimization.

#### 5.1.1 Comparisons with Rescaling.

**Rescaling techniques.** To enable exploring the larger  $\mathcal{V}$  via natural walking in  $\mathcal{P}$ , two immediate solutions could be deployed:

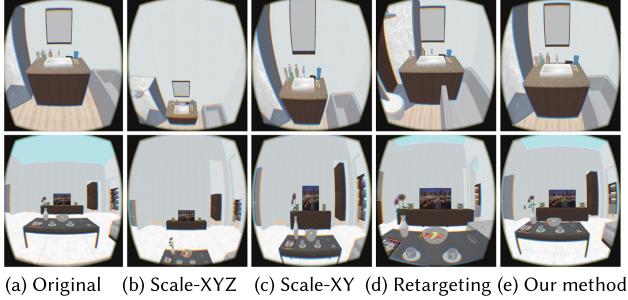


Fig. 12. Comparison with uniformly scaling and existing retargeting method. Five pairs of HMD views sampled from original scene (a), Scale-XYZ (b), Scale-XY (c), the retargeted scene by Huang et al. [2016] (d), and the retargeted scene by our method (e) are shown.

- Uniformly scale the user’s walking speed so that users do not get out of the physical workspace  $\mathcal{P}$ .
- Uniformly rescale  $\mathcal{V}$  (denoted as  $\mathcal{V}_s$ ) so that it fits into  $\mathcal{P}$ . Then users walk freely in  $\mathcal{P}$  to experience  $\mathcal{V}_s$ .

The unnoticeable translational speed gain is limited within a small range (0.86–1.26 according to Azmandian et al. [2016]). It determines the maximum capability of the former approach. We conducted preliminary experiments for validation: Users reported significant sickness immediately and unable to finish the task. Thus, we further compare our method with the latter. To transform the input virtual environment into a small real workspace, a uniformly scaling on the whole scene is a straightforward solution, as demonstrated in Figure 12. As a baseline method for retargeting 3D scenes, we perform a uniformly scaling on the input virtual environment  $\mathcal{V}$  so that the scaled  $\mathcal{V}$  (denoted as  $\mathcal{V}_s$ ) is totally in  $\mathcal{P}$ . The scale is computed as

$$\min \left\{ \frac{l_p}{l_v}, \frac{w_p}{w_v} \right\},$$

where  $l_v$  and  $w_v$  (or  $l_p$  and  $w_p$ ) are the length and width of  $\mathcal{V}$  (or  $\mathcal{P}$ ). There are two options of uniformly scaling for 3D scenes. The first one is to scale only in the length and width directions by preserving the height, denoted as Scale-XY. The second is uniformly scaling the scene in three dimensions at the same time, denoted as Scale-XYZ. In this case, the height of the user’s viewpoint is maintained to be normal. We then conduct a static experiment to determine whether this advantage is visible in practice.

*Participants.* We recruit 20 participants for each experiment (ages 21–28, 6 females). All of them had had the VR experience.

*Stimuli and Tasks.* In this experiment, a virtual camera is first set to automatically roam along the path with specified walking and steering speeds in the indoor scene. The walking path is designed to maximize spatial coverage. Then, the above walking process is divided into 40 segments based on the roaming time, and we randomly sample a snapshot of view for each one. With that, we obtain 40 snapshots of different views.

In each trial, the participant completed a 2AFC task. Specifically, (s)he was first shown a view snapshot from the original scene for 4 s as the pedestal stimulus. It follows by one from the scene generated by **our method (OURS)** and the other by uniformly rescaling

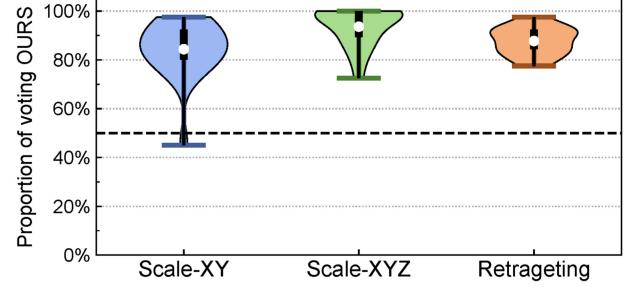


Fig. 13. Statistical violin plot of the 2AFC studies’ results from Section 5.1. The X axis indicates the three comparisons. The Y axis shows the proportional votes of OURS, of the 800 trials (40 trials/subject  $\times$  20 subjects) in each comparison. White dots and the thick black lines indicate means and quartiles respectively. The dashed line on 50% indicates a random guess rate given the 2AFC task. The consistent distribution close to 100% shows that our method significantly outperforms rescaling and Huang et al. [2016] in preserving first-person view perceptual similarity in VR. Individual votes are plotted in Appendix C.

(Scale-XY/XYZ) successively. The order of the two is random. The participants were then instructed to choose the one visually closest to the pedestal. The snapshots were all shown to users in VR. Each subject was shown 40 trials with one from each segment.

*Results and Discussion.* Appendix C plots the individual votes. As summarized in Figure 13, of 40 trials, on average 84.4% (33.74/40, STD 11.9%) snapshots of OURS are chosen by participants when comparing to the uniformly scaling Scale-XY. The value is higher for conditions comparing with the uniformly scaling Scale-XYZ with a score of 93.6% (37.45/40, STD 8.0%). Subjects reported that the uniformly scaling Scale-XY introduces distortion artifacts while the uniformly scaling Scale-XYZ brings a visual experience being exposed to a Lilliputian world for users. In comparison, OURS restructures the virtual scene in the sense of non-homogeneous and can more reliably produce higher-quality scenes. These differences are intuitively demonstrated in Figure 12.

A binomial test on the voting with  $p < 2.2e^{-16}$  for both comparisons indicates that participants statistically rate OURS as being visually more similar to the pedestal stimulus than scene scaling. Naive rescaling produces highly noticeable shape distortion even if the scaling factor is small. Consequently, it introduces an unrealistic and sickening visual experience.

### 5.1.2 Comparisons with Previous Retargeting Method.

*Implementation.* Beyond uniformly rescaling, retargeting virtual scenes to different target scales has been introduced in 3D design research [Huang et al. 2016]. However, instead of keeping design intentions or artistic themes in the space layout, our goal is to reserve the perceived first-person view in VR HMDs. In this experiment, we compare our method with an existing **retargeting method (RETARGETING)** [Huang et al. 2016]. Since OURS is fully automatic while user intervention would affect the results of Huang et al. [2016], we choose to apply the no-interaction version of RETARGETING.

*Participants, Stimuli, and Tasks.* For a fair comparison, this experiment follows the same participant group and tasks as in

Section 5.1.1. As visualized in Figure 12, the stimuli were OURS as sampled in Figure 12(e) and the corresponding views implemented via the approach in Huang et al. [2016] as sampled in Figure 12(d).

**Results and Discussion.** Appendix C plots the individual votes. As summarized in Figure 13, of 40 trials, on average 87.8% (35.1/40, STD 5.5%) snapshots of OURS are chosen by participants when comparing to the restructured scene views generated by RETARGETING. As intuitively illustrated in Figure 12(d), RETARGETING restructures the scene by preserving its global structure for design purposes than local, occluded, and first-person views in a VR setting. Consequently, it may make some walkable areas in the retargeting scene too narrow to walk through, and even introduces unnecessarily perceived objectwise overlaps.

**5.1.3 Comparisons with Redirected Walking.** Redirected walking [Nilsson et al. 2018] utilizes the limited sensitivity of human perception to create a sense of space that is much larger than the physical surroundings. That is, the virtual environment is interactively and imperceptibly rotated when the user navigates a large-scale virtual environment by real walking. In this experiment, we compare our proposed method with a typical redirected walking technique, **Steer-to-Center (S2C)** technique as in Hodgson and Bachmann [2013] and Azmandian et al. [2016].

**Participants.** We recruited 38 participants. Twenty of them had had the VR walking experience. During the experiment, 5 participants experienced general sickness and could not complete the whole task set. We thus dropped these cases in the result analysis, resulting in 33 users (ages between 21 and 30, 12 females) in total.

**Stimuli and Task.** Each participant was instructed to complete the experiment with three steps: a warm-up/reference training and two formal trials. The training process is first provided for all participants to get familiar with the scene and interactive operation. To ensure the validity and fairness of the experiment, the participants were allowed to see the interior of rooms from the doors in the living room. Following a 10-minute break, the subjects conducted the two formal trials with a rest for an hour in between. For each trial, participants were instructed to complete the simulator **sickness questionnaire (SSQ)** [Kennedy et al. 1993] before and after. With the same original scene, each trial is from one of these two conditions: one using OURS and the other uses S2C. The orders were counterbalanced. The tasks of both were the same: collecting the required objects for travel as quickly as possible. Specifically, participants were instructed to find 10 hidden objects in the scene and put them at the target position. Note that they were only allowed to pick items sequentially, but were hinted what and in which area the next item would be. Since the translation gain in the redirected walking techniques changes users' speed thus time consumption, we combine the same translation gain with our method for fair performance comparison with S2C. The size of the walkable area in the retargeted scene is 5.6 m × 7 m. After using the translation gain, users can walk freely to experience the retargeting scene in our 5.6 m × 5.6 m workspace. In the experiments, we used 1.26 as the gain level to ensure task completion with minimal discomfort across participants.

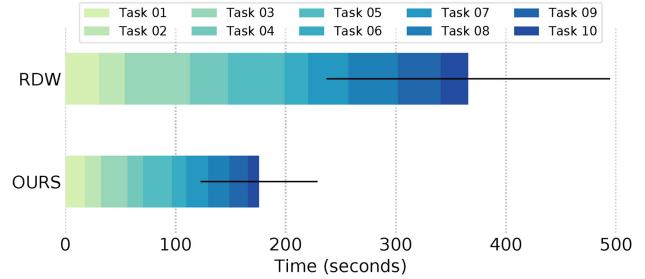


Fig. 14. Time consumption result from Section 5.1.3.

Table 1. Statistics of SSQ Score Gains after Each Trial in Section 5.1.3

Experiment	$\delta_{\text{avg}}^{\text{nau}} / \delta_{\text{dev}}^{\text{nau}}$	$\delta_{\text{avg}}^{\text{ocu}} / \delta_{\text{dev}}^{\text{ocu}}$	$\delta_{\text{avg}}^{\text{diso}} / \delta_{\text{dev}}^{\text{diso}}$	$\delta_{\text{avg}}^{\text{tot}} / \delta_{\text{dev}}^{\text{tot}}$
RDW	30.1/30.3	15.6/20.3	26.6/28.6	12.4/12.2
Ours	6.4/11.6	3.9/10.4	7.6/19.4	3.2/6.4

We report the average and the standard deviation for each metric.

**Metrics.** To evaluate and compare between the two conditions, we design both objective and subjective metrics.

First, task completion efficiencies intuitively reflect users' perceptual familiarity with the original scene as the pedestal. So, we chose them as an objective measurement. With an identical task, the shorter time consumption indicates the validity and quality of the perceived scene. Since our method can effectively compress the virtual scene, the walking distances for picking up items are much shorter than those in the RDW task. For fairness, we rescaled the time consumption of RDW by the space compression ratio of our method during comparisons. Therefore, the performance measurement is independent of the particular translational gain choice.

Beyond efficiency, a well-performed VR system shall not introduce additional simulator sickness, one of the most critical problems among all-immersive platforms. We thus compare the subjective SSQ gain (before/after task) between OURS and S2C RDW. SSQ contains three dimensions, Nausea  $\delta^{\text{nau}}$ , Oculomotor  $\delta^{\text{ocu}}$ , and Disorientation  $\delta^{\text{diso}}$ , with their sum  $\delta^{\text{tot}}$  as an overall measurement.

At last, motivated by Sun et al. [2016], we also subjectively measured loss of locomotion  $\delta^{\text{loc}}$  and visual fidelity  $\delta^{\text{vis}}$  to evaluate the overall experience. Here, the locomotion measures the walking fidelity in direction and distance. The visual fidelity measures the feeling of immersion, the comfort with navigation, and the visual experience from the scene content, design, layout, and structure. After finishing tasks, participants evaluated the above factors via rating. Discrete options (0 for "no loss of fidelity" to 3 for "severe loss") were used to guide the scoring process. The participants were instructed to compare the locomotion fidelity with both the training (pedestal) session and real world, and the visual fidelity with the training session. The participants were then instructed to sort OURS and S2C RDW, based on their interaction, locomotive, and visual experience.

**Results and discussion.** Figure 14 shows the time consumption. Repeated measures ANOVA showed that the differences between the two methods was significant ( $F_{1,32} = 314.8, p < 0.001$ ). In fact, for each task, the differences between the two methods were all

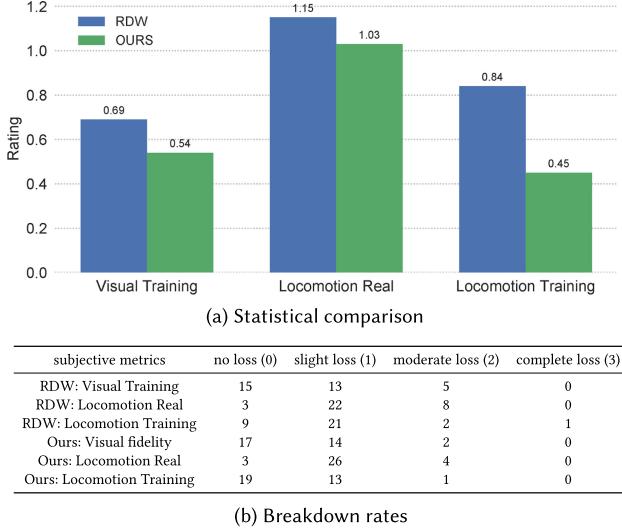


Fig. 15. Subjective fidelity rating. “Visual” for the visual fidelity, “Locomotion” for the locomotion fidelity, “Training”/“Real” for comparing the corresponding walking experience against the reference training/real world. Lower scores indicate better fidelity. (b) The numerical breakdown.

significant (all  $p < 0.001$ ). This indicates that participants completed real-world tasks more efficiently with Ours.

We then investigate the sickness *gains* after each trial: Table 1 shows the statistics. Repeated measures ANOVA showed that the differences between the two methods were significant with Ours being lower (Nausea:  $F_{1,32} = 16.79, p < 0.001$ ; oculomotor:  $F_{1,32} = 9.69, p < 0.005$ ; Disorient:  $F_{1,32} = 11.86, p < 0.005$ ; Total Sickness:  $F_{1,32} = 16.44, p < 0.001$ ). This indicates that Ours introduces significantly lower sickness than RDW in all aspects.

Figure 15(a) shows the users’ visual/locomotive fidelity rating (discrete 0–3, lower means less fidelity loss). As Figure 15(a) shows, comparing to RDW, Ours was rated better in the locomotive fidelity with both real-world (1.03 vs. 1.15,  $t(32) = 1.28, p = 0.1$ ) and VR training sessions (0.45 vs. 0.84,  $t(32) = 3.21, p < 0.05$ ) as pedestal. This trend even holds in visual fidelity (0.54 vs. 0.69). However, we did not observe statistical significance ( $t(32) = 1.09, p = 0.14$ ). The analysis shows that our method offers significantly more faithful locomotive experience than RDW with the original scene as reference. Meanwhile, it does not statistically sacrifice visual fidelity.

Regarding the subjects’ experience rating between RDW and Ours. Around 76% (25 of 33)/82% (27 of 33) users preferred Ours than RDW on walking/interaction experience. When it comes to the visual experience, RDW was more preferable than Ours (11 of 33, 33%). This also agrees with the trend above.

## 5.2 Usability Test

One limitation of VR techniques today is the narrowly applicable scenarios. To demonstrate the extendability of our method in broad real-world applications, we designed an interactive game (we show the inputs and retargeting results in the supplementary materials) with multiple indoor/outdoor scenes, entertaining tasks, and long experimental duration.

Table 2. The Statistical Results from Section 5.2

Metrics	(a) SSQ statistics			
	$\delta_{avg}^{nau}/\delta_{dev}^{nau}$	$\delta_{avg}^{ocu}/\delta_{dev}^{ocu}$	$\delta_{avg}^{diso}/\delta_{dev}^{diso}$	$\delta_{avg}^{tot}/\delta_{dev}^{tot}$
SSQ	8.0/10.7	6.8/11.0	11.7/21.4	4.9/7.8
(b) Fidelity ratings				
Fidelity loss	no (0)	slight (1)	moderate (2)	complete (3)
Locomotion	4	13	2	0
Visual	11	8	0	0

We report average and standard deviation of SSQ scores in (a) and the counts of subjective fidelity rating in (b).

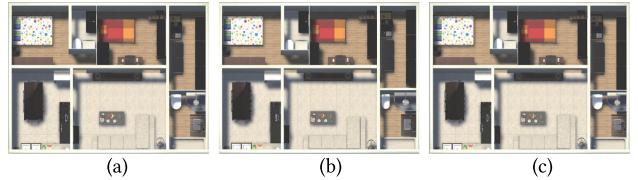


Fig. 16. Different weight sets of the energy function  $E_r(\{s_i\})$ . The virtual scene in Figure 1(a) is retargeted into the same real space, and the ratio of real space is 0.65. The weights for ( $E_{scale}, E_{object}, E_{scene}, E_{space}, E_{visibility}$ ) are set to (0.001, 200, 50, 0.0003, 0.001) (a), (0.0002, 100, 25, 0.0002, 0.002) (b), and (0.0015, 150, 75, 0.0005, 0.003) (c), respectively. The only subtle variances with different parameter setting reveals the robustness of the retargetting method.

**Participants.** Twenty participants were recruited to this experiment. During the experiment, one participant experienced general sickness and cannot complete it. We thus drop this case thus having 19 participants (ages 21–27, 7 females) in the result analysis.

**Stimuli and tasks.** In this experiment, the user plays a detective searching for a treasure with only four clues. Similarly to Section 5.1.3, participants who completed the game early were instructed to freely explore the scene until 8 minutes. These four clues were “The key in the office,” “The villa opposite the zebra crossing,” “The secret in the box,” and “The treasure pointed by the arrow.”

We created four scenes corresponding to individual clues. The game composes of an office, a street with a pedestrian crossing, a basement, and a courtyard garden. Their original and retargeted scales are listed in Appendix B. These physically unrelated scenes were connected via teleportation. The game was deployed at a small physical space (5.6 m × 5.6 m).

Similarly to Section 5.1.3, participants were instructed to complete the SSQ before/after the game, and rate visual/locomotive fidelities between 0 (no loss of fidelity comparing with real-world experience) and 3 (severe loss).

**Results and discussion.** Table 2(a) shows the post-experiment SSQ gain statistics. Comparing with the second row of Table 1, a similar sickness raise level can be observed. Table 2(b) lists the distribution of subjects’ fidelity ratings. The average rate for motion/visual fidelity was 0.89(STD 0.57)/0.42(STD 0.51). This shows that our system preserves similar fidelities after long-term and complicated usage (in comparison with Section 5.1.3 and Figure 15(b)). Comparing with recent literature on measuring immersive simulator sickness [Padmanaban et al. 2018], our method

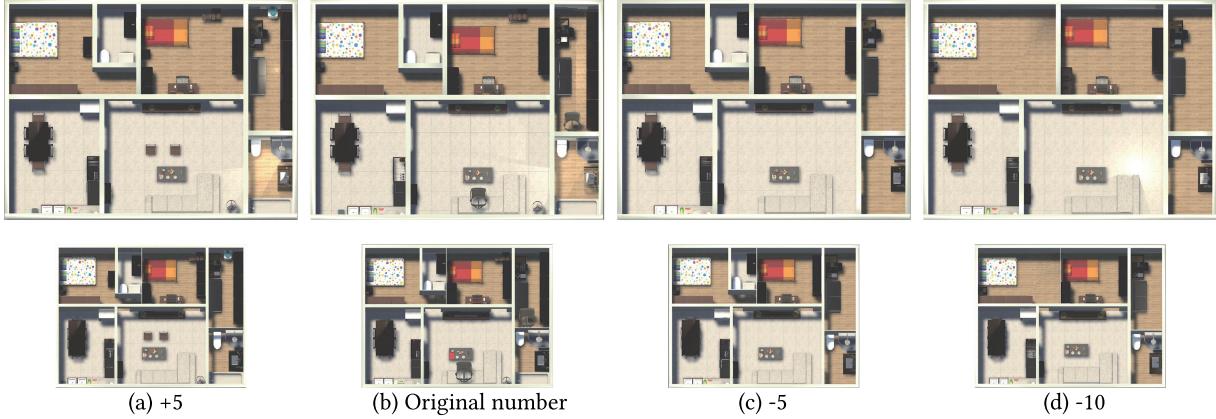


Fig. 17. Various numbers of salient objects. From left to right, the number of salient objects in the virtual scene decreases monotonously.

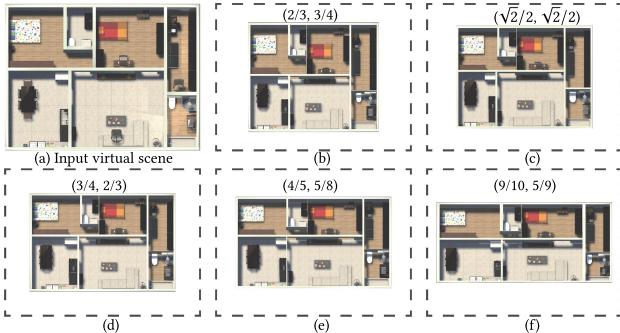


Fig. 18. Different real spaces with the same area. The values above the retargeted scene indicate the ratios between the width and height of the real space and the input virtual scene.

does not raise additional sickness levels. Further sickness reduction as constraints to retargeting optimization [Hu et al. 2019] is a valuable future research direction.

### 5.3 Intra-algorithm Evaluation

To quantify the retargetting quality, we report several objective metrics, including the scaling ratios in length/width, and the ratio between length and width (denoted as  $\text{Scale}_l$ ,  $\text{Scale}_w$ , and  $\text{Scale}_r = \max\{\text{Scale}_l/\text{Scale}_w, \text{Scale}_w/\text{Scale}_l\}$ , respectively). We also show the absolute values of relative displacement of the center of salient objects in the X axis and Y axis after retargeting, and their  $\ell_2$  norm (denoted as  $\text{Dis}_x$ ,  $\text{Dis}_y$ , and  $\text{Dis} = \sqrt{\text{Dis}_x^2 + \text{Dis}_y^2}$ , respectively). For each metric, we report its maximum, minimum, average, and standard deviation. The running time of the retargeting optimization is also recorded. Statistics for these metrics and the timings are reported in the supplementary materials.

*Weights for energy terms.* To test the influence of weights on the results of our method, we use several different sets of weights in retargeting optimization. We show an example with three different sets of weights in Figure 16. The objective metrics of the retargeting results are shown in the supplementary materials. When

the weights are set within specific ranges, the retargeting results are almost the same. Through experiments, we found that weights can be set between  $[0.0005, 50, 10, 0.00005, 0.0002]$  and  $[0.005, 500, 100, 0.0005, 0.002]$ .

*Number of salient objects.* To evaluate the influence of the number of salient objects on our results, we adjusted the number of salient objects in the virtual scene in Figure 1(a) to generate three new virtual scenes. Compared with the original scene, five salient objects are added, five salient objects are reduced, and ten salient objects are reduced, respectively. We retarget these four virtual scenes into the same real space and the ratio of space is set to 0.65, as shown in Figure 17. From the visual results in Figure 18 and the objective metrics in the supplementary materials, the greater the number of salient objects increases, the larger the visual distortion is. The reason is that more objects in the scene will reduce the space available for optimization, and we have to introduce more distortions for these objects to satisfy hard constraints. Besides, as the feasible space of the optimization problem becomes smaller, more optimization time is needed to find a feasible solution.

*Shape of real space.* To evaluate the influence of the real space shape on our results, we retarget the virtual scene in Figure 1(a) into five different-shaped real spaces with the same area. The results are shown in Figure 18. Based on the visual results in Figure 18 and objective metrics in the supplementary materials, we found that the visual distortion of salient objects in the scene increases accordingly, as the shape difference between the real space and the virtual scene becomes more severe.

*Stress Test.* To evaluate the performance of our method at varied scales in real-life usage, we conducted a stress test. In this experiment, the virtual scene in Section 5.1.3 (i.e., Figure 1(a)) was retargeted into different sizes. The ratio of space compression ranges from 0.75 to 0.45 at 0.05 intervals, as shown in Figure 19. We were always able to obtain a retargeted scene unless the compression ratio is too small to meet spatial constraints. Note that with the compression ratio is getting smaller, the visual distortion of salient objects in the scene increased accordingly. We recorded the rescale ratio of objects after retargeting and the result is shown in



Fig. 19. Retargeting results of different sizes from 0.75 to 0.45. The first one is shown as the origin scene. Corresponding HMD views at the same position are also shown for each scene. As the retargeting size decreases, the visual distortion of the scene becomes more and more pronounced. The object-level distortion along with the compression stress level is plotted in Figure 20. The objective metrics of the retargeting results are shown in then supplementary materials.

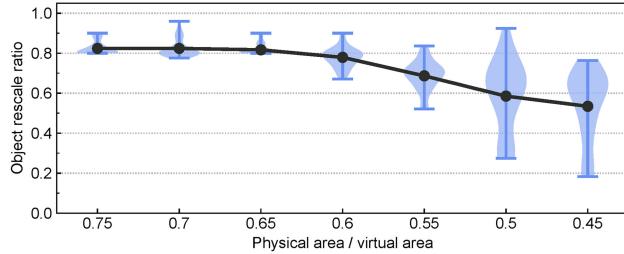


Fig. 20. Statistical plot of Section 5.3. The X axis indicates the ratio of space compression. The Y axis shows the rescale ratio of salient objects in the scene after retargeting. This visualizes the trend of salient objects' distortion level changes along with the scene compression scales.

Figure 20. When the ratio of space compression is close to 0.5, the shapes of some salient objects will be distorted obviously after retargeting, which may lead to unrealistic user experience. However, our system can still generate visually natural results that appear similar to the original with less extreme compression demands.

*Various virtual scenes.* In addition to five virtual scenes used in the walking experiments, we also designed three different types of virtual scenes to verify the versatility of our method. We show the inputs and retargeting results in then supplementary materials.

## 6 CONCLUSION

In this article, we present a novel approach that restructures a given virtual scene into limited physical spaces to enable a natural walking. Beyond geometries and topologies, we also emphasize the VR-oriented spatial perception with temporal samples. We tackled the traditionally open problems across applicability over non-occluded scenes, abrupt teleportation experience, and limited spatial saving ratio. This is achieved via spatially and perceptually aware scene restructuring, and temporal first-person immersive view sampling.

*Large distortion.* Benefiting from preventing interrupting teleportations, our system introduces less sickness than gain-based RDW approaches as in Section 5.1.3. However, the noticeable structural distortion would be introduced when we compress a scene into a much smaller size, as demonstrated in the stress test Section 5.3. This trend becomes more visible when the virtual scene is outdoor or widely open, as shown in Figure 21. It is due to the more commonly larger salient objects and wider walkable areas in outdoor spaces: in such scenes, the visual distortion becomes more sensitive to the compression ratio thus more easily to be perceived as the restructuring result becomes crowded. The other challenging case is to retarget extremely cluttered environment. In such condition, the original scenes lack sufficient space to have the

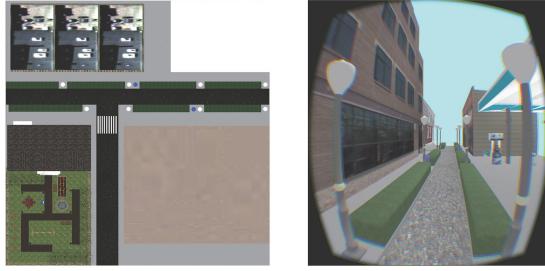


Fig. 21. Failure case with large-scale open scenes. The main road becomes extremely narrow and the whole scene seems to be crowded.

optimization achieve low loss. As a result, the objects may still show strong distortion or artifacts.

*Typical sickness.* Moreover, after long-term scenario usage, the typical sickness, especially fatigue is also raised to a higher level (Section 5.2). To address this limitation, we plan to investigate a proper modeling of sickness to our structuring solution.

*Combined with other real walking techniques.* Our method can be automatically adapted to other real walking methods, which can significantly expand the capability of these methods. Specifically, we can first reduce the size of the input virtual scenes using our method, and then apply other real walking techniques on the above-retargeted scenes. Therefore, adding new constraints and optimization goals to our method so that the needs of other real walking methods can be better satisfied is a promising research direction in the future.

*Axis-aligned bounding box simplification.* Our simplification of the axis-aligned bounding box adapts to most of the scenes without introducing more complex descriptions and parameters. However, the objects may not be well approximated in some cases, thereby resulting in nonintuitive and unnatural scene layouts. To resolve this issue, one straightforward and available approach is to use a better simplification method. For example, we can use oriented bounding boxes or K-discrete oriented polytope. However, our descriptions and objective function are required to be modified into more complex fashions, resulting in additional optimization difficulty and complexity. We foresee it as interesting and challenging future work.

*Parameters.* In the previous section, we have studied the weights for energy terms. However, there are many other parameters in our experiment. In general, they are manually specified, such as the threshold of scaling factors and the road width. Automatic selection of parameters is also worth exploring.

Currently, the salient objects share a constant importance value. Adaptive importance based on each object’s appearance and shape may effectively improve the restructuring quality and efficiency. The current deep learning-based approaches may shed light on automatically assigning the values.

## APPENDICES

### A PARTICIPANT STATISTICS

Table 3. Demographical Statistics of Participants

Experiment	#part	Age Range	$a_{\text{avg}}/a_{\text{dev}}$	Male, Female	#used
Scale-XY	20	[21, 28]	24.65/1.57	14, 6	20
Scale-XYZ	20	[21, 28]	24.65/1.57	14, 6	20
[Huang et al. 2016]	20	[21, 28]	24.65/1.57	14, 6	20
Redirected walking	33	[21, 30]	24.70/2.13	21, 12	20
Treasure hunting	19	[21, 27]	24.47/1.31	12, 7	18

### B USABILITY TEST SCENE DETAILS

Table 4. Scene Scale Details in the Usability Test (Section 5.2)

	Original	Optimized	Walkable Area
Office	10 m × 7 m	6.5 m × 4.55 m	6.5 m × 4.55 m
Street	48 m × 48 m	33.6 m × 33.6 m	7.2 m × 1.8 m
Basement	12 m × 12 m	7.8 m × 7.8 m	6.6 m × 7 m
Garden	13 m × 15 m	8.45 m × 9.75 m	6 m × 7 m

### C INDIVIDUAL VOTES

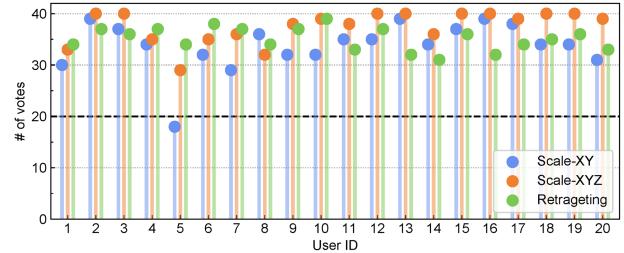


Fig. 22. Individual vote results from Section 5.1.1 and Section 5.1.2. The X axis is the user ID while the Y axis is the number of votes to OURS. The total number of trials was 40, thus the dashed line on 20 indicates a random guess rate. Our method significantly outperforms rescaling and Huang et al. [2016].

### ACKNOWLEDGMENTS

We thank user study participants for evaluating our system and the anonymous reviewers for their constructive suggestions and comments.

### REFERENCES

- Shai Avidan and Ariel Shamir. 2007. Seam carving for content-aware image resizing. *ACM Trans. Graph.* 26, 3 (2007), 10.
- Mahdi Azmandian, Timofey Grechkin, Mark Bolas, and Evan Suma. 2016. The redirected walking toolkit: A unified development platform for exploring large virtual environments. In *Proceedings of the IEEE 2nd Workshop on Everyday Virtual Reality (WEVR’16)*. 9–14.

- Mahdi Azmandian, Timofey Grechkin, and Evan Suma Rosenberg. 2017. An evaluation of strategies for two-user redirected walking in shared physical spaces. In *Proceedings of the IEEE Conference on Virtual Reality (VR'17)*. 91–98.
- Zhi-Chao Dong, Xiao-Ming Fu, Zeshi Yang, and Ligang Liu. 2019. Redirected smooth mappings for multiuser real walking in virtual reality. *ACM Trans. Graph.* 38, 5 (2019), 149:1–149:17.
- Zhi-Chao Dong, Xiao-Ming Fu, Chi Zhang, Kang Wu, and Ligang Liu. 2017. Smooth assembled mappings for large-scale real walking. *ACM Trans. Graph. (SIGGRAPH ASIA)* 36, 6 (2017).
- Russell A. Epstein and Chris I. Baker. 2019. Scene perception in the human brain. *Ann. Rev. Vis. Sci.* 5, 1 (2019), 373–397. <https://doi.org/10.1146/annurev-vision-091718-014809>
- Ran Gal, Olga Sorkine, and Daniel Cohen-Or. 2006. Feature-aware texturing. In *Eurographics Symposium on Rendering Techniques*. 297–303.
- Stefan Hartmann, Björn Krüger, and Reinhard Klein. 2015. Content-aware retargeting of discrete element layouts. In *Proceedings of the International Conference on Computer Graphics, Visualization and Computer Vision*, Vol. 23. 173–182.
- John Henderson, David Zhu, and Christine Larson. 2011. Functions of parahippocampal place area and retrosplenial cortex in real-world scene analysis: An fMRI study. *Vis. Cogn.* 19, 7 (08 2011), 910–927. <https://doi.org/10.1080/13506285.2011.596852>
- John M. Henderson and Andrew Hollingsworth. 1999. High-level scene perception. *Ann. Rev. Psychol.* 50, 1 (1999), 243–271.
- John M. Henderson, Christine L. Larson, and David C. Zhu. 2008. Full scenes produce more activation than close-up scenes and scene-diagnostic objects in parahippocampal and retrosplenial cortex: An fMRI study. *Brain Cogn.* 66, 1 (2008), 40–49. <https://doi.org/10.1016/j.bandc.2007.05.001>
- Bill Hillier and Alan Penn. 1991. Visible colleges: Structure and randomness in the place of discovery. *Sci. Context* 4, 1 (1991), 23–50.
- Eric Hodgson and Eric Bachmann. 2013. Comparing four approaches to generalized redirected walking: Simulation and live user data. *IEEE T. Vis. Comput. Gr.* 19, 4 (2013), 634–643.
- Ping Hu, Qi Sun, Piotr Didyk, Li-Yi Wei, and Arie E. Kaufman. 2019. Reducing simulator sickness with perceptual camera control. *ACM Trans. Graph.* 38, 6, Article 210 (Nov. 2019), 12 pages. <https://doi.org/10.1145/3355089.3356490>
- Chun-Kai Huang, Yi-Ling Chen, I-Chao Shen, and Bing-Yu Chen. 2016. Retargeting 3D objects and scenes with a general framework. *Comput. Graph. Forum* 35, 7 (2016), 33–42.
- Yong Jin, Ligang Liu, and Qingbiao Wu. 2010. Nonhomogeneous scaling optimization for realtime image resizing. *Vis. Comput.* 26, 6–8 (2010), 769–778.
- Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *Int. J. Aviat. Psychol.* 3, 3 (1993), 203–220.
- Talia Konkle, Timothy F. Brady, George A. Alvarez, and Aude Oliva. 2010. Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychol. Sci.* 21, 11 (2010), 1551–1556.
- Vladislav Kraevoy, Alla Sheffer, Ariel Shamir, and Daniel Cohen-Or. 2008. Non-homogeneous resizing of complex models. *ACM Trans. Graph.* 27, 5 (2008), 111.
- Eike Langbehn and Frank Steinicke. 2018. *Redirected Walking in Virtual Reality*. Springer Encyclopedia of Computer Graphics and Games, 1–11.
- Eike Langbehn, Frank Steinicke, Markus Lappe, Gregory F. Welch, and Gerd Bruder. 2018. In the blink of an eye: Leveraging blink-induced suppression for imperceptible position and orientation redirection in virtual reality. *ACM Trans. Graph.* 37, 4 (2018), 66:1–66:11.
- Jinjie Lin, Daniel Cohen-Or, Hao Zhang, Cheng Liang, Andrei Sharf, Oliver Deussen, and Baoquan Chen. 2011. Structure-preserving retargeting of irregular 3D architecture. *ACM Trans. Graph.* 30, 6 (2011), 183:1–183:10.
- Alex Mansfield, Peter Gehler, Luc Van Gool, and Carsten Rother. 2010. Scene carving: Scene consistent image retargeting. In *Proceedings of the European Conference on Computer Vision*. 143–156.
- Sebastian Marwecki and Patrick Baudisch. 2018. Scenograph: Fitting real-walking vr experiences into various tracking volumes. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology (UIST'18)*. Association for Computing Machinery, New York, NY, 511–520. <https://doi.org/10.1145/3242587.3242648>
- Edvard I. Moser, Emilio Kropff, and May-Britt Moser. 2008. Place cells, grid cells, and the brain's spatial representation system. *Annu. Rev. Neurosci.* 31, 1 (2008), 69–89.
- Niels Christian Nilsson, Tabitha Peck, Gerd Bruder, Eri Hodgson, Stefania Serafin, Mary Whitton, Frank Steinicke, and Evan Suma Rosenberg. 2018. 15 years of research on redirected walking in immersive virtual environments. *IEEE Comput. Graph. Appl.* 38, 2 (2018), 44–56.
- Aude Oliva and Antonio Torralba. 2001. Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int. J. Comput. Vis.* 42, 3 (2001), 145–175.
- Nitish Padmanaban, Timon Ruban, Vincent Sitzmann, Anthony M. Norcia, and Gordon Wetzstein. 2018. Towards a machine-learning approach for sickness prediction in 360° 3D stereoscopic videos. *IEEE Trans. Visual. Comput. Graph.* 24, 4 (2018), 1594–1603.
- Soojin Park, Talia Konkle, and Aude Oliva. 2014. Parametric coding of the size and clutter of natural scenes in the human brain. *Cerebr. Cortex* 25, 7 (01 2014), 1792–1805. <https://doi.org/10.1093/cercor/bht418> arXiv:<https://academicoup.com/cercor/article-pdf/25/7/1792/14102860/bht418.pdf>.
- John Peponis, Ruth Conroy Dalton, Jean Wineman, and Nick Dalton. 2004. Measuring the effects of layout upon visitors' spatial behaviors in open plan exhibition settings. *Environ. Plan. B: Plan. Des.* 31, 3 (2004), 453–473.
- Michael Rubinstein, Diego Gutierrez, Olga Sorkine, and Ariel Shamir. 2010. A comparative study of image retargeting. *ACM Trans. Graph.* 29, 6 (2010), 160.
- Michael Rubinstein, Ariel Shamir, and Shai Avidan. 2008. Improved seam carving for video retargeting. *ACM Trans. Graph.* 27, 3 (2008), 16.
- Michael Rubinstein, Ariel Shamir, and Shai Avidan. 2009. Multi-operator media retargeting. *ACM Trans. Graph.* 28, 3 (2009), 23.
- Ariel Shamir and Olga Sorkine. 2009. Visual media retargeting. In *ACM SIGGRAPH ASIA 2009 Courses*. 11.
- Evan A. Suma, Seth Clark, David Krum, Samantha Finkelstein, Mark Bolas, and Zachary Warte. 2011. Leveraging change blindness for redirection in virtual environments. In *Proceedings of the IEEE Conference on Virtual Reality (VR'11)*. 159–166.
- Evan A. Suma, Zachary Lipps, Samantha Finkelstein, David M. Krum, and Mark Bolas. 2012. Impossible spaces: Maximizing natural walking in virtual environments with self-overlapping architecture. *IEEE T. Vis. Comput. Gr.* 18, 4 (2012), 555–564.
- Qi Sun, Anjul Patney, Li-Yi Wei, Omer Shapira, Jingwan Lu, Paul Asente, Suwen Zhu, Morgan McGuire, David Luebke, and Arie Kaufman. 2018. Towards virtual reality infinite walking: dynamic saccadic redirection. *ACM Trans. Graph.* 37, 4 (2018), 67:1–67:13.
- Qi Sun, Li-Yi Wei, and Arie Kaufman. 2016. Mapping virtual and physical reality. *ACM Trans. Graph. (SIGGRAPH)* 35, 4 (2016), 64:1–64:12.
- Daniel Vaquero, Matthew Turk, Kari Pulli, Marius Tico, and Natasha Gelfand. 2010. A survey of image retargeting techniques. In *Applications of Digital Image Processing XXXIII*, Vol. 7798. 328–342.
- K. Vasylevska and H. Kaufmann. 2017. Towards efficient spatial compression in self-overlapping virtual environments. In *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI'17)*. 12–21. <https://doi.org/10.1109/3DUI.2017.7893312>
- Khrystyna Vasylevska, Hannes Kaufmann, Mark Bolas, and Evan A. Suma. 2013. Flexible spaces: Dynamic layout generation for infinite walking in virtual environments. In *Proceedings of the 2013 IEEE Symposium on 3D User Interfaces (3DUI'13)*. 39–42.
- Shu-Fan Wang and Shang-Hong Lai. 2009. Fast structure-preserving image retargeting. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. 1049–1052.
- Yu-Shuen Wang, Chiew-Lan Tai, Olga Sorkine, and Tong-Yee Lee. 2008. Optimized scale-and-stretch for image resizing. *ACM Trans. Graph.* 27, 5 (2008), 118.
- Lap-Fai Yu, Sai-Kit Yeung, Chi-Keung Tang, Demetri Terzopoulos, Tony F. Chan, and Stanley J. Osher. 2011. Make it home: Automatic optimization of furniture arrangement. *ACM Trans. Graph.* 30, 4, Article 86 (Jul. 2011), 12 pages. <https://doi.org/10.1145/2010324.1964981>
- Gregory Zelinsky. 2013. Understanding scene understanding. *Front. Psychol.* 4 (2013), 954. <https://doi.org/10.3389/fpsyg.2013.00954>
- Bolei Zhou, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. 2014. Learning deep features for scene recognition using places database. In *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (Eds.), 487–495.

Received September 2021; revised April 2021; accepted June 2021