# Creative and Progressive Interior Color Design with Eye-tracked User Preference

SHIHUI GUO, YUBIN SHI, PINTONG XIAO, YINAN FU, and JUNCONG LIN,
Xiamen University
WEI ZENG, The Hong Kong University of Science and Technology (Guangzhou) and The Hong Kong University of Science and Technology
TONG-YEE LEE, National Cheng-Kung University

Interior scene colorization is vastly demanded in areas such as personalized architecture design. Existing works either require manual efforts to colorize individual objects or conform to fixed color patterns automatically learned from prior knowledge, whilst neglecting user preference. Quantitatively identifying user preferences is challenging, particularly at the early stage of the design process. The 3D setup also presents new challenges as the inhabitant can observe from any possible viewpoint. We propose a representative view selection method based on visual attention and a progressive preference inference model. We particularly focus on the progressive integration of eye-tracked user preference, which enables the assistance in creativity support and allows the possibility of convergent thinking. A series of user studies have been conducted to validate the effectiveness of the proposed view selection method, preference inference model and the creativity support mechanism.

CCS Concepts: • **Human-centered computing → Interaction design**; **Interactive systems and tools;**

Additional Key Words and Phrases: Interior color design, user preference, creativity support

**ACM Reference format:**
Shihui Guo, Yubin Shi, Pintong Xiao, Yinan Fu, Juncong Lin, Wei Zeng, and Tong-Yee Lee. 2023. Creative and Progressive Interior Color Design with Eye-tracked User Preference. *ACM Trans. Comput.-Hum. Interact.* 30, 1, Article 5 (March 2023), 31 pages.
https://doi.org/10.1145/3542922



## 1 INTRODUCTION

Inspired by natural landscapes with stunning colors, humans actively decorate commercial and residential environments with a wide spectrum of colors. Colorization is a creativity-driven task, given the potentially unlimited diversity of color combinations. Researches in interior color design [9, 33, 71] aim at developing tools that allow users to colorize an indoor scene. This is a

This project is partially supported by the National Natural Science Foundation of China (62077039, 62072383, 61702433), Alibaba Innovative Research Program, Research Project (PZ2020016), the Fundamental Research Funds for the Central Universities China (20720190006) and Ministry of Science and Technology 110-2221-E-006 -135 -MY3 Taiwan.

Authors' addresses: S. Guo, Y. Shi, P. Xiao, Y. Fu, and J. Lin (corresponding author), Xiamen, Fujian, China, 361005; emails: tens444@163.com, xpt_solferino@qq.com, yinan.fu1202@gmail.com, guoshihui@xmu.edu.cn, jclin@xmu.edu.cn; W. Zeng, The Hong Kong University of Science and Technology (Guangzhou) and The Hong Kong University of Science and Technology, Guangzhou, Guangdong, China, 511458; email: weizeng@ust.hk; T.-Y. Lee, National Cheng-Kung University, Tainan, Taiwan, 701; email: tonylee@mail.ncku.edu.tw.



1073-0516/2023/03-ART5 $15.00
https://doi.org/10.1145/3542922

nontrivial task, which requires a careful balance between object-wise color assignment and global compatibility of scene style [34]. For instance, scribbles [18, 31] are commonly used to specify object-wise color but not able to guarantee global compatibility of scene style. On the other hand, various metrics (such as harmony [41], theme [66], and contrast, [28], etc.) can be adopted as constraints when optimizing global compatibility for the whole scene, but the metrics may assign non-optimal colors for individual objects.

Although manual methods like scribbles allow the user to precisely assign the color of a target object, the process is laborious when colorizing a highly complex scene. For 3D case, the situation becomes worse due to the unavoidable frequent camera manipulation [18]. Therefore, recent works (e.g., [21, 69]) have explored a data-driven paradigm that learns object-wise colorization patterns from reference images. The strategy is effective for natural scenes as the color of a natural object is relatively fixed (e.g., sky is usually blue or gray). For human-made scenes, the color of an object can be arbitrary and largely determined by people's preferences, whilst existing methods usually lead to highly homogeneous results [33].

Besides, reasonability is another premise of a colorization task. Reasonable suggestions usually consider general factors such as harmony, style, and contrast. However, these constraints are usually difficult to quantify and do not rigorously guarantee reasonability. Incorporating user preference during the design process is a potential solution. However, accurately capturing users' preferences is an essential yet problematic task: simply asking users what they want is too intrusive and prone to error, while monitoring behavior unobtrusively and then finding meaningful patterns is difficult and usually computationally time-consuming. Eye movements provide clues about the visual attention of a person and have been vastly used in **Human-Computer Interaction** (**HCI**) for evaluation and design of user interface [42, 55, 56, 65]. It is especially suitable for visual content related applications [27, 39], and we show how such a clue can be ingeniously used for omni-directional 3D colorization. Finally, the lack of enough data at the beginning exacerbates the difficulty of capturing user's real preference accurately. Therefore, we need a scheme to incrementally and implicitly capture the user's real preference and a way to progressively incorporate such preference in the colorization process.

Colorization of interior scenes is a highly creative task, especially for non-residential environments like kindergartens, shopping malls, and bars. Although various **creativity support tools** (**CSTs**) have been developed in the HCI community [17] for different application domains, most of them focused on the divergent side of creativity, providing diverse suggestions for inspiration during the design process. However, the convergent process is also necessary to arrive at useful output or production of creativity [14]. Creativity is usually associated with the capacity to produce something new and adequate. Indeed, divergent thinking, referring to the generation of multiple ideas or solutions for a single problem [6], is important for CSTs. However, convergent thinking, associated with finding a single solution to a problem in an analytical and deductive way [6], is also indispensable. While divergent thinking would be needed to generate new ideas, convergent thinking would be helpful in identification and refinement of adequate ideas [64]. The integration of users' preferences acts as a way of convergent thinking to select and refine those best "gems" to converge on a small set. However, careful balance should be made between divergent and convergent thinking on the role they will play in a creativity support system.

This article presents a creativity-support colorization system for 3D interior scene. An original scene (Figure 1(a)) is colorized using an interactive evolutionary algorithm. During this process, user preference is identified by tracking eye movements (Figure 1(b)) and progressively incorporated in the design framework. The colorization results are presented in Figure 1(c, d). User studies show that our method provides higher quality colorization suggestions compared with

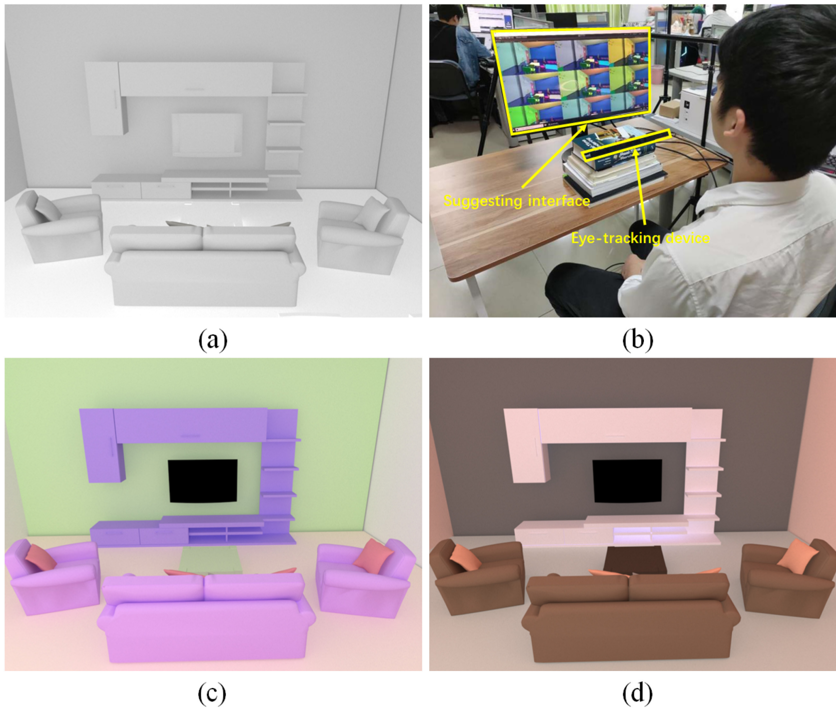



Fig. 1. Creative interior color design with eye-tracked user preference. (a) The original scene; (b) the experiment set-up; (c, d) colorization designs from two participants.

non-personalized cases, and significantly outperforms alternative colorization tools in terms of time cost and user preference. More specifically, our work makes the following contributions:

— We propose a novel approach to incorporate eye-tracking as a non-intrusive way to harness user preference for interior colorization. Our work explores the balance of divergent and convergent thinking in a CST. This is achieved by progressively controlling the importance of such preference in the optimization process. We show this approach ensures the diversity at the initial stage and eventually finalizes user preference.
— We develop an interactive evolution framework to generate diverse colorizations of a scene in an omni-directional manner. The colorization evolution is guided via the captured user preference as mentioned above. We also address the challenge of identifying the representative views in a virtual scene, by considering the visual attention when the user navigates in the 3D scene.
— We empirically evaluate the utility of our approach and the framework through three user studies: (1) the effectiveness of eye-tracking to capture user preference, (2) the support for convergent thinking, and (3) the usability of our system. Both quantitative and qualitative results reveal that our work produces satisfactory colorization results effectively and efficiently.

## 2 RELATED WORK

### 2.1 3D Color Assignment

Colorization of 3D objects is a fundamental function for many applications and, therefore, has been extensively investigated. Many researches explored user-friendly interfaces for manual painting on 3D models. Fu et al. [18] described a novel and efficient WYSIWYG 3D painting interface, which allows users to draw long strokes across different depth layers with occluded regions. Leifman and Tal proposed [30] a scribble-based mesh colorization algorithm, which can reasonably propagate desired colors from scribbles to the whole mesh. Ortega and Vincent [48] presented automated camera control technique to assist direct drawing on 3D shapes. The technique emancipated users from frequent viewpoint manipulation and allowed them to draw around a 3D shape even under occlusion. Manual methods maybe powerful in controlling of the colorization, but they are also very tedious for scenes with tens or hundreds of objects. Chen et al. [9] introduced the magic decorator which automatically generates material suggestions for 3D indoor scenes. Zhu et al. [71] described a similar data-driven approach to colorize 3D furniture models and indoor scenes by leveraging indoor images on the internet as references. However, both works [9, 71] chose to learn from existing samples, which maybe suitable for the colorization of natural scene but not creative tasks such as interior colorization. Similar idea has also been used to generate material suggestions for 3D objects based on the learned relationship between shape and material [23]. A recent work [33] considered more general and view-dependent constraints like harmony and mood in an omni-directional manner, leading to diverse suggestions for inspiration. However, the redundant selection of views led to high computation burden and interference of colorization, and the lack of more specific constraints led to too many possible colorizations yet only few of them fit users' preferences. There is also work [45] investigating the transfer of material style from a guide source to a target 3D scene, while we are targeting on colorization of a blank scene from scratch.

This article introduces two technical improvements on our previous work [33] to address the two critical issues: (1) an attractiveness-based view selection to drastically reduce time cost and improve the effectiveness of color optimization; (2) the integration of eye-tracked user preference to progressively guide the colorization, providing more reasonable yet personalized suggestions.

### 2.2 Creativity Support System

Creativity has been a growing topic in the computing community since the 1990s, especially in HCI research [16]. As declared by Shneiderman [57], the development of **CSTs** has been one of the grand challenges of HCI. A large body of researches was devoted to collaborative creativity, aiming to provides tools for ideation [35, 58], affective communication [1], and content editing [70] Another major topic is to provide diverse suggestions during design process to inspire the user [8, 68]. Finally, evaluation of a CST's effectiveness is also a critical issue and therefore receives intensive attentions. Earlier works adopted qualitative methods, such as observation, think-aloud studies, and interviews. As a common tool for quantitative analysis, survey methods, such as **Task Load Index** (**TLX**), have also been introduced for CSTs evaluation [29]. However, TLX is more appropriate for productivity and other related software. In [7], Carroll et al. proposed a specialized metric, the so-called **Creativity Support Index** (**CSI**), which was then further developed and extensively validated through a rigorous psychometric process [12]. However, most of existing works in this area focused on support of divergent thinking for creativity. A major reason is that most scholars in creativity domain agreed on the importance of divergent thinking while the consensus is not so clear for convergent thinking until recently [14].

As summarized in [16], existing creativity-oriented HCI researches favor the design and development novel CSTs [3, 13, 38], but they may not always be applicable in actual and contemporary forms of professional creative practice. Different from existing works, we choose to explore the balance of divergent and convergent thinking in a CST. Although there are few works [44, 70] with the same concern as us, we uniquely study how to progressively evolve the diverse suggestions to converge by incremental integration of user preferences in a suggestive CST.

### 2.3 Quantifying User Preference Using Eye Tracking Measures

User preference is an important factor in user-oriented situations, such as designing [25, 32, 37, 47, 51], recommendation [40, 59, 62] and other social activities [24, 50]. Quantitative evaluation of personal preferences is critical yet challenging in these applications, mostly for two reasons: on one hand, novel techniques are needed to accurately identify user preference without interfering the ongoing task; on the other hand, in many application scenarios we may not have enough data for inference at the beginning, incremental methods are therefore expected based on a small set of samples. The way to quantize and exploit preference may vary depending on the different applications. There are generally two main categories [36]: (1) analyze each user's attention on different items and then integrate attention weights; (2) adapt the user or item vector according to the most influential users or items. We refer the interested readers to some surveys [15] for more details. Specific to our application, eye tracking appears to be a good candidate for non-intrusive preference capturing as we are mainly dealing with visual media. Eye tracking system is typically used to capture human visual attention characteristics, which find vastly usage in various applications such as saliency detection, object detection, image and video quality assessment, interaction design, and so on. People tend to fixate their attention on the interested objects more frequently [46], many researches thus utilized the amount of fixations to represent the level of attention on an **area of interest** (**AOI**). Fixation points are clustered using algorithms such as mean shift [53], Bayesian online clustering [61], and velocity-based clustering [52] to identify the viewer's AOIs. While the centroid of each cluster is used to represent the AOI, such a simple treatment ignores the inner spatial relationship among fixations and is susceptible to noises. Improvements have been made to address these issues and thus locate the viewer's visual attention more accurately. O. Spakov and D. Miniotas [72] utilized the fixation durations to calculate a weighted centroid while Wang et al. [67] take a densest position-based strategy. X. Chen and Z. Chen [11] proposed a visual attention identification method based on random walks. With the captured visual attention, users' preferences can then be derived [10, 63], especially combining with other data. In [60], H. Song and N. Moon additionally used web social behavior data such as web page visit and search history to infer users' preferences more accurately.

We explore the usage of visual attention, indicated by the eye and head movement, for attractiveness-driven viewpoint selection in 3D scene and robust inference of user preference towards diverse suggestions presented in a CST. While most existing eye-tracked preference models [10, 60, 63] are discrete, we propose a continuous inference model which is more robust to noise. Besides, uncertainty of the captured preference is provided by our model as a cue to guide the integration of preference in colorization.

## 3 DESIGN OF THE SYSTEM

### 3.1 Overview

The goal of our work is to provide users with interior colorization suggestions that are creative and appealing to their preferences. The colorization suggestions shall enable object-wise color assignment that is globally compatible and conforms to the specified style of user's preference. To accomplish the goal, we design a creative colorization framework (Section 3.2) coupled with attractiveness-guided representative view selection (Section 3.3) and eye-tracked user preference modeling (Section 3.4). Figure 2 illustrates the pipeline of our system. Our method builds upon a framework of evolution algorithm, which iteratively produces a new population of colorization suggestions with a dedicated objective function. At each iteration, a set of nine colorization results is presented to users in a $3 \times 3$ grid in the suggestive viewpoint of the colorized scene. This set of

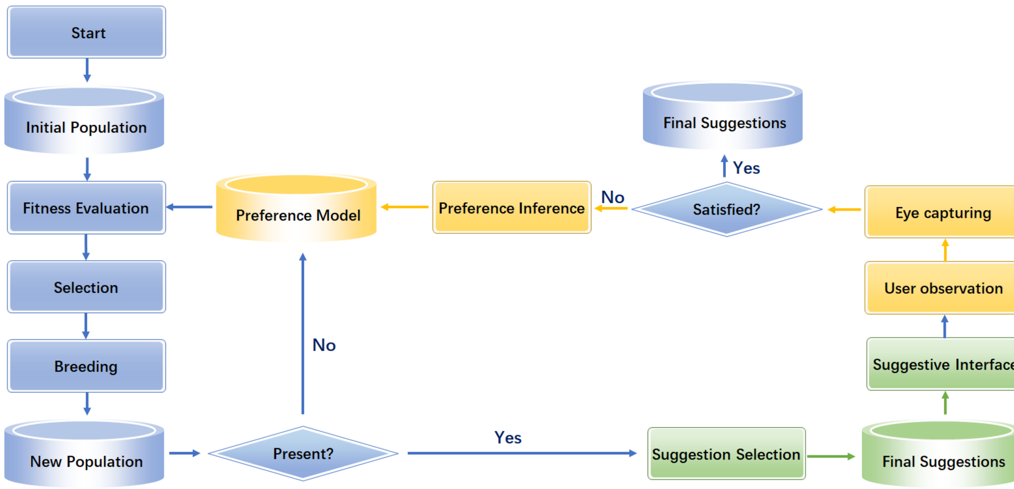



Fig. 2. System pipeline. Blue blocks constitute the main procedure, the interactive evolution framework, which will pause after a certain number of iterations and present colorization suggestions to the user for reference (green blocks). When the user observing these suggestions, the system will infer his preference, which is then incorporated into the evolution framework to guide the colorization of interior scene (yellow blocks). The process will be repeated until the user is satisfied by the suggestions.

suggestions is selected from the evolving population, with a balance between fitness and diversity. We then start from the beginning to select a candidate by checking whether the difference between the candidate and each of the already selected suggestions is greater than a threshold according to Equation (3). The resolution of the grid can also be adjusted, we use a $3 \times 3$ grid by default for several reasons: (1) the number of diverse suggestions presented to the user to be as large as possible under the prerequisite that each suggestion is clear enough for observation (the screen area should not be too small), so that the user can be better inspired; (2) according to Miller's theory of magic number seven [43], most adults can store between five and nine items in their short-term memory; (3) the number of high-quality (high fitness and diverse) suggestions can be provided by the interactive evolution framework in each round is also limited.

When users are observing these suggestions, an eye-tracking device captures their gaze information (position and time), which is used to construct a colorization preference model (see Section 3.4). Users can stop the evolving process if satisfied with the result. Alternatively, the system will incorporate the continuously updated preference model during the fitness evaluation of the new population.

We implemented a prototype system with the Unity Engine using C# script. Tobii 4C Eye Tracker is used for eye tracking, with key parameters listed in Table 1. Detailed evaluations for the device could be found in [19]. Each user needs to go through a calibration process by following the standard routines provided by the manufacturer [22].

## 3.2 Creative Colorization Framework

The core of our colorization framework is an interactive evolution algorithm. The complete representation of color assignment is packed into a long vector $\mathbf{c}$ consisting of all the RGB tuples of object's diffuse component. We take the color vector $\mathbf{c}$ as an individual in the population, and calculate the fitness by objective function defined below. Then we use classic crossover and mutation operations to evolve the population, so as to generate a better color assignment. In each iteration,

Table 1. Specifications of Tobii 4C Eye Tracker

| **Parameters** | **Value** |
|---|---|
| Recommended Max Screen Size | 27 inches with 16:9 Aspect Ratio; |
| | 30 inches with 21:9 Aspect Ratio. |
| Operating Distance | 50–95cm |
| Track Box Dimensions | 40 × 30 cm at 75 cm |
| Image sampling rate | 90 Hz |
| Tracking population | 97% |
| Illuminators | Near Infrared (NIR 850 nm) Only |

we generate $n_s$ samples and reserve $n_e$ elites to ensure the convergence of the algorithm. We set $n_s, n_e$ as 50 and 15 in our current implementation. To avoid the frequent evaluation of objective function, the evolution is accelerated with a surrogate-based kriging model [26].

*3.2.1 Objective Function.* The purpose of the evolution process is to find optimal color assignments (diffuse component) for objects (including light sources) in a 3D scene. The objective function is defined as a weighted sum of several factors including harmony $E_h(\mathbf{c})$, mood $E_m(\mathbf{c})$, contrast $E_c(\mathbf{c})$, anchor constraints $E_a(\mathbf{c})$, and user preference $E_u(\mathbf{c})$:

$$E(\mathbf{c}) = w_h E_h(\mathbf{c}) + w_m E_m(\mathbf{c}) + w_c E_c(\mathbf{c}) + w_a E_a(\mathbf{c}) + w_u E_u(\mathbf{c}). \tag{1}$$

We follow the approach of Lin et al. [33] on definition of the first four terms ($E_h(\mathbf{c})$, $E_m(\mathbf{c})$, $E_c(\mathbf{c})$, and $E_a(\mathbf{c})$) with the same formulas and weight settings ($w_h$, $w_m$, $w_c$, and $w_a$). Ablation studies about the effect of each term could also be found in [33]. Note that the definitions of $E_h(\mathbf{c})$ and $E_m(\mathbf{c})$ are the sums of harmony and mood calculated on the scene images under the selected representative views (Section 3.3), respectively. The definition of preference term $E_u$ will be detailed in Section 3.4, and the corresponding weight $w_u$ for the term is defined as

$$w_u = f(n_{iter}) \cdot \epsilon_u, \tag{2}$$

where $f(n_{iter}) = 0.1 \cdot n_{iter}$ is a monotonically-increasing function of the iterations $n_{iter}$, and $\epsilon_u$ is the uncertainty of user preference (see Section 3.4). In contrast to Lin et al. [33], we add extra $E_u$ to signify user preference and interactively lead the process of the evolution algorithm. We show that the factor of user preference is critical to convergent thinking in the process of creative tasks, providing more satisfactory suggestions as indicated by the user study. Diversity of suggestions presented to the user is also essential for a CST. We design a metric combining color $\mathbf{c}_i$, mood $\psi(V_k)$ and harmony $T(V_k)$ to measure the difference between two suggestions $S_m$ and $S_n$:

$$d(S_m, S_n) = \delta_c \sum_i a_i \parallel \mathbf{c}_i^m - \mathbf{c}_i^n \parallel^2 + \delta_m \sum_k \parallel \psi^m(V_k) - \psi^n(V_k) \parallel^2 + \delta_h \sum_k f(T^m(V_k), T^n(V_k)), \tag{3}$$

where $\delta_c = 1$, $\delta_m = 5$, and $\delta_h = 5$ are weights to balance the three components. We empathize more on the higher level differences (such as mood and harmony) as there are usually many objects in the scene. Different colorizations of objects may lead to similar suggestions by simply exchanging the colors among objects. $f(\cdot)$ is the distance between the two best match harmony templates of $S_m$ and $S_n$ on the current viewpoint $V_k$:

$$f(T^m(V_k), T^n(V_k)) = \begin{cases} 2\pi, & T^m(V_k) \neq T^n(V_k) \\ |\alpha^m(V_k) - \alpha^n(V_k)|, & T^m(V_k) = T^n(V_k), \end{cases}$$

where $\alpha^m(V_k)$ and $\alpha^n(V_k)$ are the angles of the two harmony templates $T^m(V_k)$ and $T^n(V_k)$. Please refer to [33] for details about mood and harmony measurements.

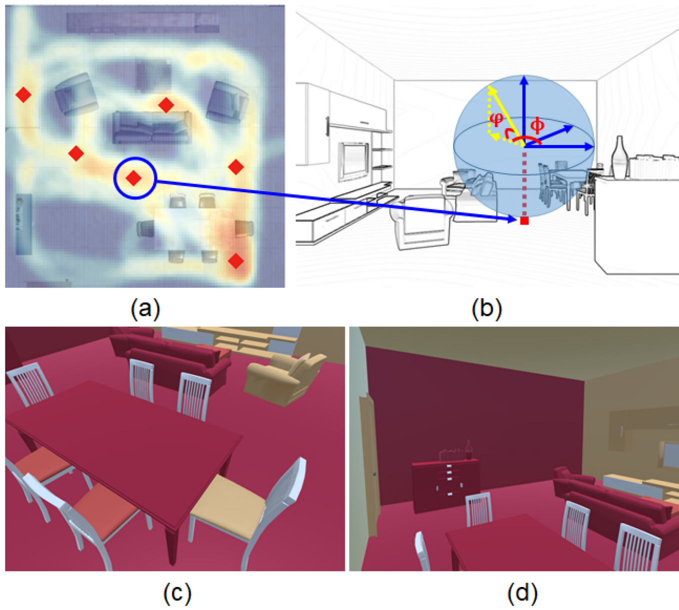



Fig. 3. Viewpoint selection according to attractiveness: (a) action map of the scene; (b) finding optimal viewing directions from a representative observation position; (c, d) observed scene views from representative viewpoints and directions.

A farthest point sampling strategy is adopted to select the most diverse models: first, the point with the maximum distance from the center is selected; we then repeatedly select the point which is the farthest from the selected set until the desired number of diverse models have been selected.

### 3.3 Representative View Selection

A key component in our system is to select representative views, where the objective function for color optimization is defined. Figure 3 illustrates our approach of view selection. Theoretically, there are innumerable views, from which we can observe a virtual 3D scene, making the definition and calculation of the harmony and mood terms impossible in Equation 1. However, as inhabitants, we usually observe a scene from quite a few representative views in real-world. For example, we are usually attracted by the desk in an office other than the ceiling or floor, and seldom climb to a higher position to observe the office. The purpose of this stage is to find a minimum yet representative enough set of views, to make the calculation of the harmony term and mood terms in Equation 1 accurately and quickly.

Given a 3D interior scene, we first infer an action map [54], from which we choose $n_v$ most possible observation positions with the distances between each pair of them are greater than a certain threshold $d_r$. These positions are chosen in a greedy manner: the position $\mathbf{p}_{max}$ with the greatest possibility is chosen and put in a list $\mathbb{P}_r$ and then the next greatest one with its distances to all the positions in $\mathbb{P}_r$ greater than $d_r$ is chosen and included in $\mathbb{P}_r$. This process is repeated until the number of candidates in $\mathbb{P}_r$ reaches the limit $n_v$. We set $n_v = 6$ in our implementation. For each observation position $\mathbf{p} \in \mathbb{P}_r$, we use a stochastic optimization process to find a position $\mathbf{p}_e$ on the unit sphere centering at $\mathbf{p}$. $\mathbf{p}_e$ and $\mathbf{p}$ define an eye ray, which determines a viewpoint $V_e = [\mathbf{p}, \mathbf{p}_e]$ together with other fixed camera parameters. The objective of the stochastic optimization is to find the position $\mathbf{p}_e$ whose $V_e$ is the most attractive.

Table 2. Object Category and Weight

| Category | Examples | Weight |
|---|---|---|
| Decoration | vase, mirror | 0.35 |
| Function | TV, fridge, etc. | 0.25 |
| Storage | bedside table, wardrobe, etc. | 0.2 |
| Furniture | sofa, desk, etc. | 0.15 |
| Frame | wall, ceiling, floor, etc. | 0.05 |

*3.3.1 View Attractiveness Evaluation.* When people enter an interior scene, their attention will be attracted by some objects in the scene while ignoring others. For example, when entering an office, one may note the office table and its surroundings while ignoring ceilings, walls, or floors. For attractiveness of a scene view, we refer to the extent of the inhabitant attention to the scenario of the current view. As such a metric is quite subjective, we therefore take a supervised-learning approach to train a discriminator for evaluation of view attractiveness. There are two main challenges to design such a discriminator: (1) collection of enough training data; and (2) determination and definition of input factors related to view attractiveness.

A naive strategy for the first challenge is to randomly sample scene views and collect user ratings. However, there are no consistent judgment rules to accomplish user ratings, given a large divergence of human aesthetic. Another commonly used solution is a paired comparison, by inviting users to identify a more attractive view from two presented options. We take a heuristic method to address the first challenge and to capture the subconscious opinion of the person. When participants navigate in the scene, we record the duration $t_{mn}$: the $m$th participant stays on the $n$th view and define the attractiveness $A_i$ of each object $O_i$ by

$$A_i = \sum_m \sum_n \frac{t_{mn}}{\sqrt{2\pi}\sigma_a} \exp\left(-\frac{1}{2\sigma_a^2} d(O_i, \mathbf{VC}_{mn})\right), \tag{4}$$

where $d(\cdot)$ measures the distance between $O_i$ and the view center $\mathbf{VC}_{mn}$. By doing so, we build the dataset of object attractiveness. The attractiveness of the view is the summed attractiveness of all objects in the view. For each viewpoint, several most attractive views are selected on each viewpoint by repeatedly selecting the most attractive uncovered object, whose barycentric coordinate defines an eye ray from the viewpoint. Finally, we randomly pick up some additional views.

For the second challenge, we firstly classify objects in the scene into one of five categories (see Table 2) with different weights applied on the related metrics in the feature vector. We then compute the following metrics of each category and pack together to form the input feature vector:

(1) Average of each object's projected barycentric coordinate in the viewing image.
(2) Average of the norm of each object's barycentric coordinate.
(3) Average of each object's distance to the viewpoint in 3D space.
(4) The total number of objects.
(5) Sum of the objects' projection area in the view image.

We then use an LSTM-based Recurrent Neural Network for regression between the feature vector and view attractiveness. The selection of LSTM is based on the sequential characteristic of viewpoint trajectory in the temporal domain. During the experiments, we notice that some users may stare at the empty wall for a while though they are not interested in it at all, leading to quite an amount of abnormal data (about 10% of the total). As L2 loss is very sensitive to abnormality while L1 loss is not so efficient, we adopt Huber loss for a more robust and efficient regression.

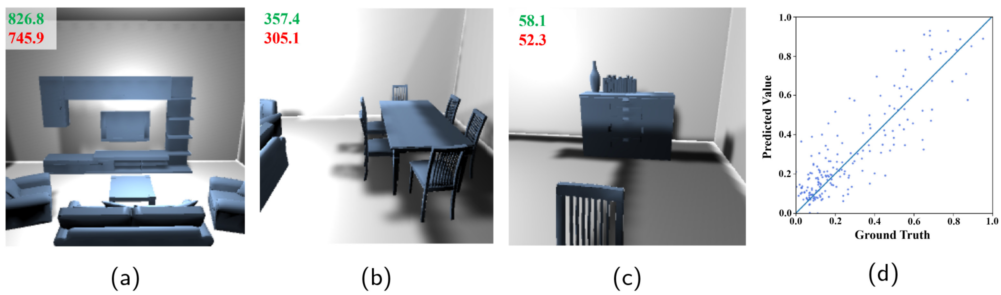



Fig. 4. Attractiveness regression: (a–c) predicted (the red text) and calculated attractiveness (the green text) for given viewpoints; (d) correlation analysis.

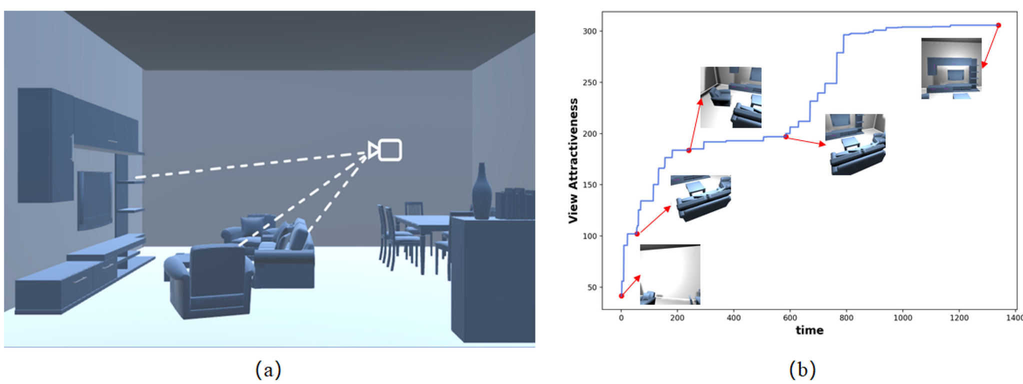



Fig. 5. MCMC-based selection of representative view directions: (a) the 3D scene; (b) the convergent curve.

Figure 4 shows the correlation analysis between the calculated attractiveness and the one predicted by the regressor, with examples of paired comparisons. The comparison achieves 87% consistency between the two. This implies the effectiveness of using our regressor to predict the object attractiveness for viewpoint selection.

*3.3.2 Spherical Domain View Selection.* Following the overall thoughts at the beginning of Section 3.3, we need to find several optimal view directions at each viewpoint. We resort it as a stochastic optimization problem, with the cost of each view evaluated by the discriminator in Equation 4. A **Markov chain Monte Carllo** (**MCMC**) sampler is adopted to explore the function and produce multiple optimized samples [20]. Figure 5 demonstrates the convergence of the MCMC-based selection of representative view directions.

### 3.4 User Preference Model

Figure 6 illustrates the system to capture the user's preference with an eye-tracking device. While users are observing the presented colorization suggestions, the system records their eye focus coordinates $\mathbf{p}_k$ and the corresponding duration $t_k$ on the window region $R$. We assume that the user's attention not strictly concentrates on a single gaze point but declines with the distance away from the gaze point. For a record $e_k(\mathbf{p}_k, t_k)$, we define a distribution function around $\mathbf{p}_k$ measuring the attention intensity within the surrounding region $R$:

$$p(e_k, \mathbf{q}) = \frac{t_k}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma_p^2} \parallel \mathbf{p}_k - \mathbf{q} \parallel\right\}, \tag{5}$$

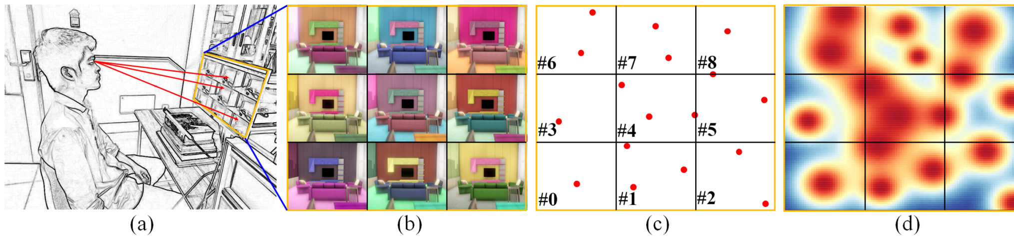



Fig. 6. User preference captured with eye-tracking: (a) system setup; (b) suggestions presented to the user; (c) the user attention visualized on the window; (d) heat map of the user attention.

where $\mathbf{q}$ refers to a point in the surrounding area. The probability distribution $p(\mathbf{q})$ of the user attention on $\mathbf{q}$ can be computed by summing up all the record distribution functions. By this way, we can get a continuous map about user's attention on the screen other than discrete gaze points. And finally, the user preference on a certain suggestion $S_m$ is computed by integrating $p(\mathbf{q})$ over the window region $R_m$ occupied by $S_m$. Such a continuous strategy is more robust to gaze noise.

With samples of user's preference on each presented suggestion, we can then fit a regression model to predict the user's preference on a new colorization. Here, we use the kriging model [26] for regression. An additional advantage of the model is the uncertain estimation for the prediction, which can be used in the optimization process as a weight to control the importance of the user's preference.

## 4 USER STUDY 1: USER PREFERENCE CAPTURE

**Motivation.** One key contribution from our work is to capture the user's preference and progressively guide the scene colorization. To better select the effective mode for capturing user preference, we test three modes for preference capture: mouse-clicking, eye-tracking, and hybrid. In the hybrid mode, the user is allowed to use both mouse-clicking and eye-tracking. Study 1 investigates the differences among these three interactive modes for the effectiveness of preference capture. Besides, we notice that participants may subconsciously pay more attention to certain area (such as the center) of the suggestive view. Such attention bias will inevitably affect the inference of their actual intention from gaze information. Therefore, we divide the user study into two stages, to firstly deal with attention bias and then investigate the effect of different preference capturing modes.

### 4.1 Experiment Setup

A total of 15 different colorizations of the same scene were generated for calibration of attention bias. For the condition of eye-tracking, we used the Tobii eye tracker to monitor the gaze status; for the condition of mouse-clicking, we instructed the user to select the preferred colorization results by clicking the specific grid cells.

### 4.2 Participant

We recruited 20 participants (15 males and five females). The group includes undergraduate, graduate students and staffs with backgrounds of either computer science or digital media technology. Their ages range from 21 to 40. All participants declared normal or corrected vision, with no colorblindness.

### 4.3 Procedure

Before the experiment, all participants were informed of the experiment's purpose and provided their written consent. A video recording was set up throughout the experiment for further post-experiment investigation. The participants were aware of video-recording. This routine was consistently required in the subsequent Study 2 and 3.

**Stage 1: Attention bias calibration.** In this stage, nine colorizations out of 15 were randomly selected and presented as suggestions to the participant, whose gaze information was collected for calibration. The process was repeated for five rounds, with three out of the nine colorizations being replaced in the next round. The new nine colorizations were redistributed in the grid. Each of the 15 colorizations appeared three times in different grid cells. To validate the calibration, we conducted a post-experiment procedure and asked the participants to perform paired comparisons of the 15 colorizations: to select one from two colorizations according to his/her preference. Note the paired comparison is only conducted in this user study. In a real design scenario with the tool, the user only needs to perform the 5-round processes for attention calibration.

**Stage 2: Preference capturing mode investigation.** After bias calibration in the first stage, each participant was then asked to perform interior colorization with our system under three different modes in a random order:

— *mouse-clicking mode.* The participant was instructed to use the mouse only to explicitly specify his/her preferred suggestions on the suggestive interface.
— *eye-tracking mode.* The participant was instructed to use an eye-tracking device to capture his/her preference, while the mouse was turned off.
— *hybrid mode.* The eye-tracking device captured the participant's preference. At the same time, the participant was also allowed to use the mouse to explicitly specify his/her preference.

For both the mouse-clicking and hybrid modes, the user was free to select an arbitrary number of grids based on his/her preference. Each participant was asked to use the three modes in a random order. Note that each participant could take a rest before he/she started performing another task.

### 4.4 Finding

**Stage 1: Attention bias calibration.** Our idea to calibrate attention bias is based on a premise: the user's preference of a certain colorization should be independent of its position in the grid. Let $\mathbf{b} = \{b_1, b_2, \ldots, b_9\}$ be the bias pattern of the participant on the $3 \times 3$ suggestive grid, we can derive the following equation for each of the 15 calibrating colorizations:

$$t_1 - b_i = t_2 - b_j = t_3 - b_k, \tag{6}$$

where $t_1$, $t_2$, and $t_3$ are the three gaze residence times when the colorization appears in grid cell $\#i$, $\#j$, and $\#k$, respectively. This leads to an over-constrained linear system, which is solved in a least square manner to get the bias vector $\mathbf{b}$. From the paired comparison of the 15 colorizations, we inferred an order of them by the participant's preference. The order was then converted as the participant's rating on preference, and served as the benchmark for user preference inferred from gaze residence time.

The bias patterns of eight subjects are visualized in Figure 7. Apparently, the bias patterns for different subjects are different. It is also understandable that the center cell over the grid shows the highest attention bias for most subjects given its centric location. It is interesting that the bottom row shows the highest bias in the case of Subject 8 (the pattern in the bottom right). An investigation with the video recording showed that it was caused by the unintentional gaze when the subject was attracted by the eye-tracking device.

We plotted the rating together with gaze durations in Figure 8. Both terms are normalized to [0, 1] by $\frac{x-x_{min}}{x_{max}-x_{min}}$. In the red circle, we can observe a long stay of the participant's gaze on the #4 grid for different colorizations, which were rated quite different by the participant. While in the blue circle, we can see the participant rated quite high on several colorizations although he/she

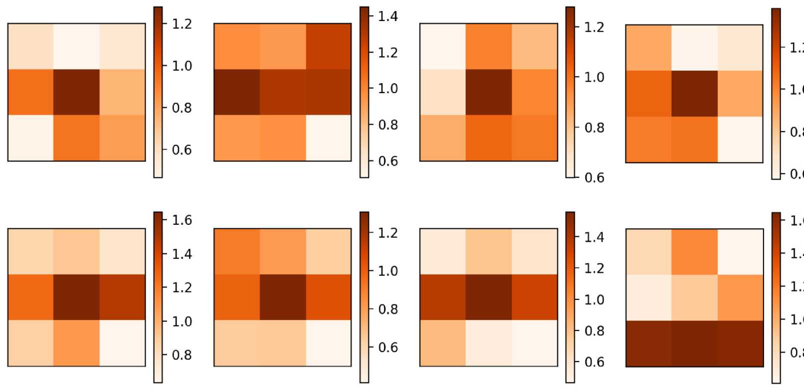


Fig. 7. Bias of gaze duration for eight representative subjects.

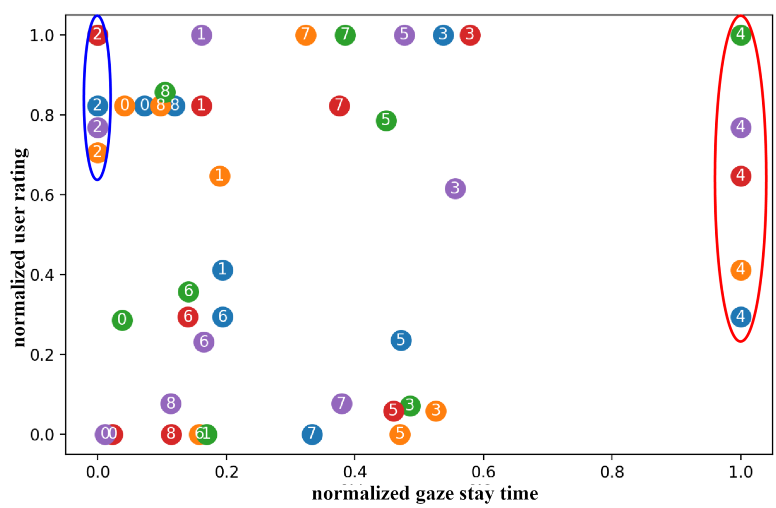



Fig. 8. Scatter plot of a representative participant's rating and gaze duration. Disks with the same color refer to the same interior colorization, while the number in the disk indicates where the colorization appears in the $3 \times 3$ suggestive grid.

only took a glance on the corner grid cell (#2). These observations validate the existence of bias on the participant's attention over the $3 \times 3$ suggestive grid.

Inspired by the bias pattern, we further investigate the accumulation of gaze with respect to the elapsed time. Figure 9 shows the gaze accumulation and change in 1 minute for a participant. The participant conducted a primary observation in 5–10 seconds, found the preferred suggestions in 15–20 seconds and finally fixed up after 30 seconds. This is consistent with the timing pattern revealed in our Stage 2 experiment: when users click the mouse button to explicitly specify their preference, they conducted a first-time click at 12.5±5.6 seconds after the colorization results were presented to them. After 45 seconds, users' preference was explicitly and stably indicated by their gaze. This is consistent with the routine approach for human subjects when making choices from a variety of options. People often narrow down to a small subset and eventually make the final decision.

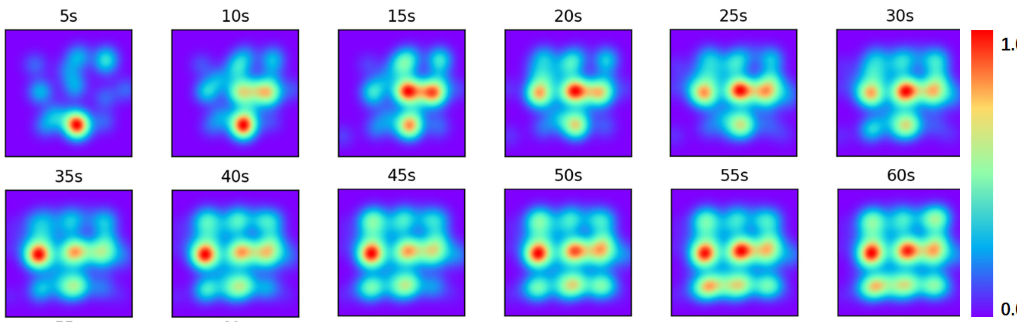



Fig. 9. Gaze accumulation in 1 minute, normalized to the range between 0.0 and 1.0.

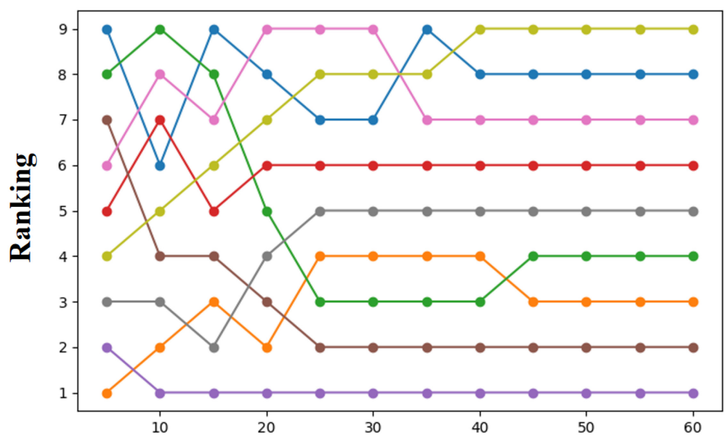



Fig. 10. Change of suggestions' ranking in 1 minute when they are presented to the participant.

The change of suggestions' ranking (Figure 10) validated this assumption. The ranking order is dynamically changing in the first 20 seconds, i.e., the user is actively screening all colorization suggestions. After 20 seconds, he/she identified Cell 4, 6, 9 as a small candidate set and fixed his gaze in this small range. During this period, only the ranking of Cell 4, 6, and 9 changed while others remained relatively stable afterwards. Although a minor fluctuation (Cell 3, 4) arises, such a fluctuation does not interfere with the top-ranking subset and makes little contribution to the final colorization result. It is clear that the gaze residence time on the top candidate (Cell 4) is monotonically increasing. This implies the consistent identification of user preference via the gaze information. In contrast, the low-ranking candidates (Cell 8, 1, 7, 2) either remain low-ranking since the initial experiment or rapidly neglected by the user, as indicated by the gaze residence time. This also shows the consistency and stability of using eye gaze to filter out undesirable results. Without the explicit activities of mouse clicking, our method avoids the repetitive labor and captures the user preference in a non-intrusive fashion. This implies the effectiveness and usability of using eye-tracking to identify user preference of the colorization result.

The bias can be rectified in the following procedure by subtraction of **b** from each new derived eye-tracked preference. The correlation between the rated and eye-tracked preferences is boosted from 0.25 to 0.47, enabled by the introduction of rectification.

**Stage 2: Preference capturing mode investigation.** We here compare the three proposed modes in terms of two aspects: *efficiency* and *consistency*. The former indicates the required time

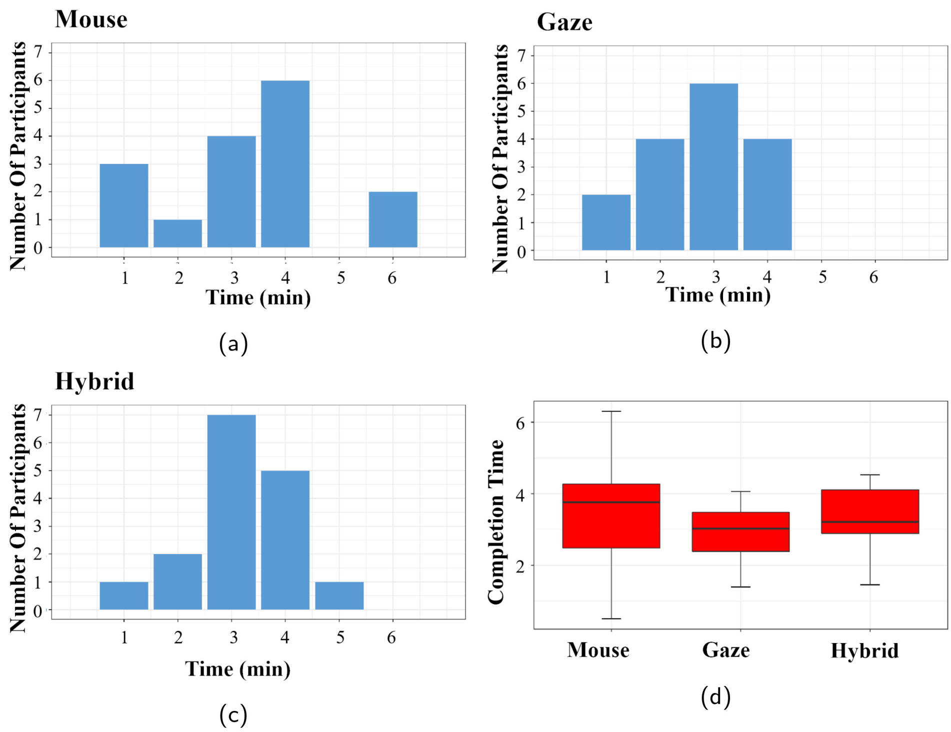



Fig. 11. Different preference capturing modes: (a–c) frequency distribution of task completion time for mouse-only mode, eye-tracking-only mode and hybrid mode; (d) task completion time of the three modes.

to complete the task of interior colorization, while the latter implies the stability of capturing user preference.

Figure 11(a)–(c) shows the frequency distribution of task completion times of participants for the three modes. During the experiment, we found three participants spending excessive time (2–3 times than the averaged cases) in determining the ranking order. This is against the experiment's motivation since the goal of this study is to explore the intuitive and natural preference capture method which requires minimal user cognitive load. The data of these three participants were removed from the following analysis. It is worth pointing out that for these three participants, the eye-tracking mode still outperforms the other two modes in terms of task completion time. We firstly verify the distribution of the timing statistics with Shapiro–Wilk test, in which results indicate an abnormal distribution ($p = 0.0086, 0.0001$, and $0.00001$ for gaze, mouse and hybrid respectively). As the result of Freidman test ($p = 0.02022$) indicates the existence of statistical significance among these three different conditions, we therefore conduct the Wilcoxon Signed-Rank test for paired evaluation, in which results (p = 0.0208, effect size = 1.8333) indicate the significance between gaze and mouse. The effect size is computed by dividing the z value by the square root of the observation number N [49]. Figure 11(d) shows the average task completion time for all subjects. The results show that the mode of mouse-clicking (mean: 3.46, std: 1.55) shows a more diverse distribution of task completion time than eye-tracking (mean: 2.89, std: 0.81) and hybrid (mean: 3.33, std: 0.87) modes. Compared with mouse clicking, eye-tracking reduces the average time by 16%, and the standard deviation by 48%. A main factor is probably the hesitance of selection. The user needs some time to explicitly decide which suggestions on the screen fit

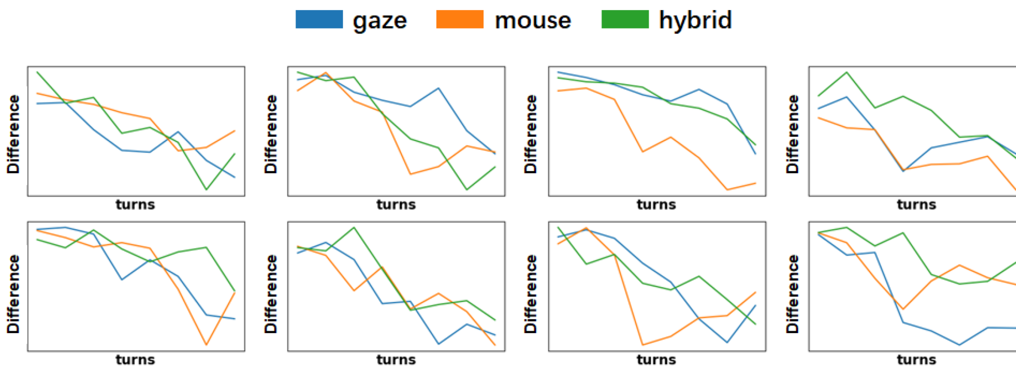



Fig. 12. Consistency of participants' choice between different rounds. The vertical axis denotes the metric of $\varrho$ (Equation (7)), and the horizontal axis is the index of the current colorization turns. Each plot is generated by one participant in one turn.

his/her preference. He/she might struggle on some of the suggestions when himself/herself was not so clear about his/her real preference. Another reason is partially due to the random location of colorization results. The user needs to pick up the mouse and move the cursor on the screen to specify his/her preference. It costs a bit more time if the target position is distant from the current cursor location. Finally, the hybrid use of both modes may potentially introduce additional cognitive load (or interference) as the user needs to decide the mode of selection.

We investigate the consistency of the participant's unconscious intent captured in different modes by comparing the suggestions presented in two neighboring turns $R_k$ and $R_{k+1}$. To be specific, we first sort suggestions in each turn according to the participant's preference, and calculate the difference between the two rounds by

$$\varrho = \sum_{i=1}^{9} d\left(S^{k}_{m(i)}, S^{K+1}_{n(i)}\right). \tag{7}$$

The difference function $d(\ldots)$ follows the same definition in Equation (3). Figure 12 shows cases of different participants. We can find the global decreases of differences for each participant on all the three different modes, indicating more consistent choices of participants holistically. In general, the mode of using gaze information achieves the best performance and results in the smallest difference in terms of colorization. However, fluctuations were observed in almost every curve, reflecting the hesitance of participants in the design process. To better evaluate the extent of hesitance, we calculated two indicators from each curve:

- the number of turning points on the curve, which are effectively local minima of the curve function: $N_{TP} = \varrho_{k-1} < \varrho_k < \varrho_{k+1}$;
- the rate of summed positive and negative gradient (absolute value) for all curve segments: $R_{PN} = \frac{\sum g_p}{\sum g_n}$.

Lower values of $N_{TP}$ and $R_{PN}$ indicate less volatility and higher consistency of user preference identification.

The statistics of the two indicators are plotted in Figure 13, which indicates that the gaze mode achieves the best performance among the three modes. More specifically, the number of turning points for the mouse-clicking, eye-tracking and hybrid modes are 2.81±0.98, 2.31±0.87, 2.69±1.14; the other metric for the three modes are 0.63±0.34, 0.50±0.19, 0.51±0.29. For both metrics, the eye-tracking mode achieves the lowest average value: about 20% lower than the mouse-clicking

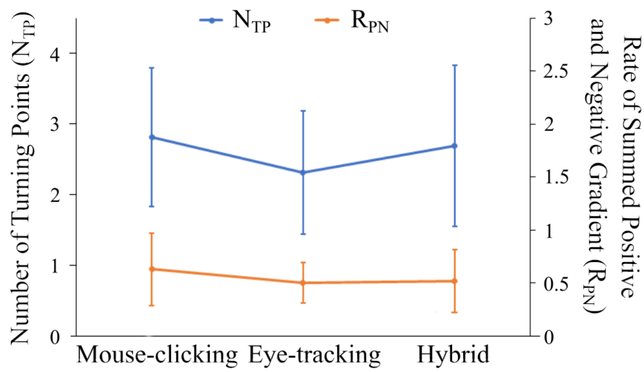



Fig. 13. Statistics of choice consistency.

mode. For the second metric, the standard deviation reduces by about 50%, indicating the high consistency and stability of user preference capture. This advantage benefits from the intuitive and non-intrusive approach of user preference capture with eye-tracking. This fact implies that user preference is a challenge to be precisely identified. Participants may rank two colorizations with the reversed order in two different rounds if the preference of these two is on a comparable level. This introduces confusion and cognitive burden to user due to the demand to explicitly identify and decide the ranking order in the mouse-clicking and hybrid modes.

## 5 USER STUDY 2: CREATIVITY SUPPORT

**Motivation.** User study 2 validates the system's effectiveness on assistance of creative tasks, with an emphasis on the balance between divergent and convergent thinking. We show that our method can progressively guide from divergent thinking to convergent thinking, enabled by the integration of user preference.

### 5.1 Experiment Setup

A set of nine colorization results were concurrently presented to the user in a $3 \times 3$ grid in the suggestive viewpoint of the colorized scene. The colorization results are dynamically generated by the proposed framework in Section 3 under two conditions: with and without user preference.

### 5.2 Participant

We recruited 20 participants (14 males and six females) for this task, 16 of them were from user study 1. The group included undergraduate, graduate students and staffs with backgrounds of either computer science or digital media technology. Their ages range from 21 to 41. All participants declared normal or corrected vision, with no color-blindness.

### 5.3 Procedure

To investigate how the system mitigates from divergence to convergence, we use a metric $\rho$ to evaluate the diversity of suggestions in each turn $k$, which is the sum of the difference between each paired suggestions in the current round:

$$\rho_k = \sum_{m=1}^{8} \sum_{n=m+1}^{9} d\left(S_m^k, S_n^k\right). \tag{8}$$

At the same time, we also let the participant rate the quality of suggestions presented in each round.

Table 3. CSI Dimensions and Statements

| Dimension | Statement |
| --- | --- |
| Enjoyment | I enjoyed using the system or tool. |
| Exploration | The system or tool was helpful in allowing me to track different ideas, outcomes, or possibilities. |
| Expressiveness | I was able to be very creative while doing the activity inside this system or tool. |
| Immersion | I became so absorbed in the activity that I forgot about the system or tool that I was using. |
| **Elitization** | **I think the suggestions are more and more close to my preference.** |
| Satisfaction | I was satisfied with what I got out of the system or tool. |

Table 4. Factor descriptions for Paired-factor Comparison

| Dimension | Description |
| --- | --- |
| Enjoyment | Enjoy using the system or tool. |
| Exploration | Explore many different ideas, outcomes, or possibilities |
| Expressiveness | Be creative and expressive. |
| Immersion | Become immersed in the activity. |
| **Elitization** | **Approach the preferred design gradually.** |
| Satisfaction | Produce results that are worth the effort I put in. |

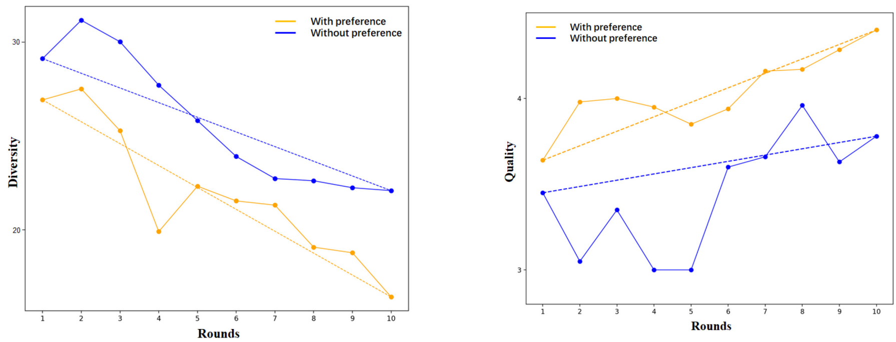
We use the psychometric measurement CSI [7, 12] with necessary modifications to evaluate the effectiveness of the system under two conditions: with and without the integration of user preference. As the system does not support collaborative creativity so far, we modify the metric by replacing the *collaboration* dimension with a new dimension *elitization* for convergent thinking. The 6 dimensions and their statements are listed in Table 3, where the new dimension is highlighted in bold. Each participant was asked to rate on a scale of "Highly Disagree" (1) to "Highly Agree" (10). Paired-factor comparison was also conducted with each factor against every other factor for a total of 15 comparisons. A factor description was selected in response to the statement: "When doing this task, it's the most important that I am able to . . . ", with the factor descriptions listed in Table 4. Finally, the CSI score $S_{csi}$ of a participant toward the system is calculated by

$$S_{csi} = \left( \sum_{i=1}^{6} R_i \cdot C_i \right) / 1.5, \tag{9}$$

where $R_i$ is the participant's rating for dimension $i$, while $C_i$ is the count of the dimension being chosen by the participant in paired-factor comparison. The hyper-parameter (1.5) is used to normalize the score to the range $0 - 100$.

### 5.4 Finding

As we can see from Figure 14, the diversity of suggestions decreases as the experiment proceeds, indicating the convergence of colorization suggestions. This confirms that the effect of user preference is gradually introduced into the evolution framework and guiding the evolving colorization samples. At the same time, participants felt that the quality of suggestions was steadily improved as revealed by their rating.




(a) (b)

Fig. 14. Evolving record of suggestion diversity (left) and quality (right). Dashed color lines refer to the fitting curves for trends. Diversity is calculated according to Equation (8), while quality is rated by users in a 1–5 likert scale.

**Ablation study: with and without user preference.** The condition with preference integration (the curve in yellow) apparently outperforms the opposing condition (the curve in blue). This is verified from multiple aspects:

— *initial values*: the effect of user preference is effective immediately after the first round of evolution. This is visualized as the lower value of diversity (27.91) and higher value (3.64) of quality, compared with the results (diversity: 29.72, quality: 3.45) generated without user preference.
— *final values:* By the end of the experiment, the diversity of colorization generated with preference is lower (19.0) than the case without preference (23.41). Meanwhile, the quality is higher for the case with preference (4.40) than without preference (3.78).
— *curve slope:* the slopes of the fitting lines in the case of diversity are −0.89 (with preference) and −0.63 (without preference). A lower value indicates a faster convergence speed. In the case of quality, the slopes are 0.076 (with preference) and 0.033 (without preference). A higher value indicates a faster improvement of colorization quality. It is also worth noting that without the integration of user preference, the improvement of colorization quality changes with large and nondeterministic variations.

This implies that our method can generate results with higher consistency among the colorization samples, with better quality perceived by users.

As we can see from the radar chart in Figure 15, the integration of the user's preference promotes the system in all aspects of CSI. The averaged CSI scores $S_{csi}$ for the conditions with/without preference are 72.1 and 58.7, with standard deviations to be 11.43 and 15.5, respectively. More specifically, the boost in terms of elicitation (w/o: 5.1, with: 7.5) is higher than the other five metrics (enjoyment: 6.2/6.9, exploration: 6/7.5, expressiveness: 5.2/6.7, immersion: 5.9/6.6, resultsworthefforts: 6.4/7.2). The paired comparison indicates a clear promotion of elicitation when users' preferences are integrated, which validates the effectiveness of our system on supporting convergent thinking.

It is interesting that the aspect of enjoyment was lower in paired comparison and higher in CSI rating. As the total count of winning in paired comparison is fixed to 15, the count for some factor will inevitably decrease when the count for elitization rises. Interviews with the subjects revealed that when using eye-tracking, participants weighed less on the enjoyment, and focused on the

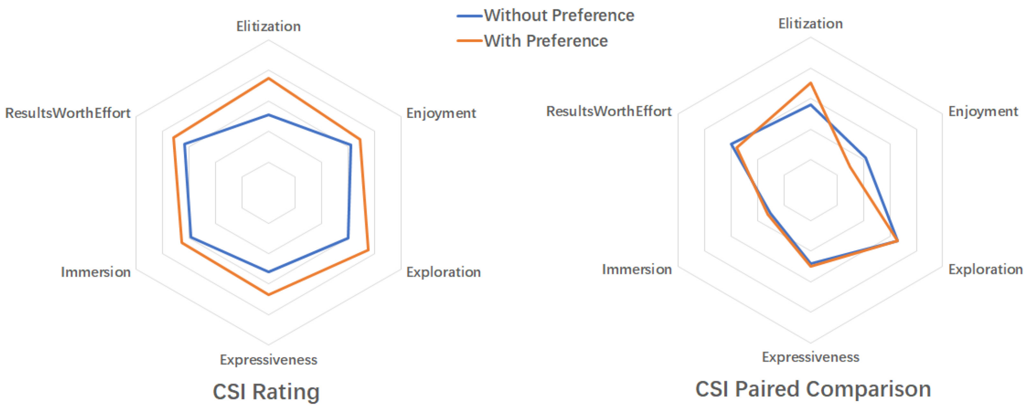



Fig. 15. CSI rating and paired comparison.

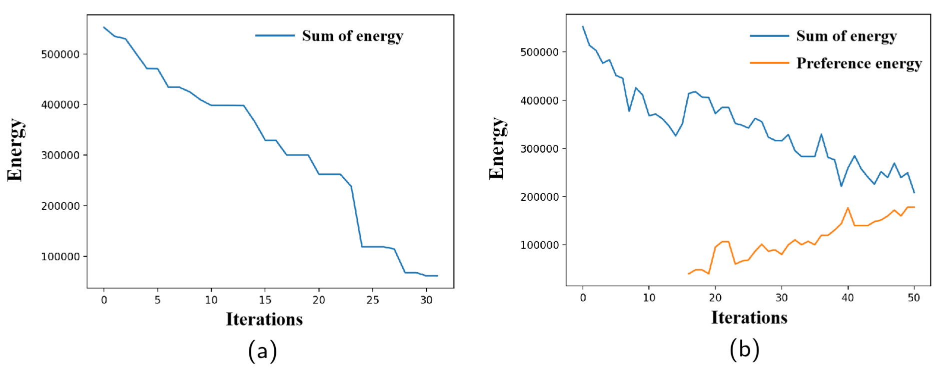



Fig. 16. System convergence: (a) without preference constraint; (b) with varying preference constraint.

output of the creativity task. This explains why the enjoyment in paired comparison is lower for the case with preference.

The convergence of the interactive evolution computing is illustrated in Figure 16. The algorithm converges consistently without preference constraint (Figure 16(a)). When the preference constraint is introduced after 15 iterations and updated every 5 iterations, we can see the regular fluctuation of the convergence curve (Figure 16(b)). This shows that the introduction of user preference is taking effect after the integration.

## 6 USER STUDY 3: SYSTEM USABILITY

**Motivation.** This study conducted an overall evaluation of the whole system. We compared our work with the standard Unity tool. The evaluation is accomplished with **System Usability Scale** (**SUS**) questionnaire [5] and semi-structured interviews.

### 6.1 Experiment Setup

An indoor scene was offered to the participants for colorization. The program of a standard Unity tool was prepared in advance and verified in a previous work [33].

Table 5. SUS Test Questions

| Questions | Description |
|---|---|
| Q1 | I think that I would like to use this system frequently. |
| Q2 | I found the system unnecessarily complex. |
| Q3 | I thought the system was easy to use. |
| Q4 | I think that I would need the support of a technical person to be able to use this system. |
| Q5 | I found the various functions in this system were well integrated. |
| Q6 | I thought there was too much inconsistency in this system. |
| Q7 | I would imagine that most people would learn to use this system very quickly. |
| Q8 | I found the system very cumbersome to use. |
| Q9 | I felt very confident using the system. |
| Q10 | I needed to learn a lot of things before I could get going with this system. |

### 6.2 Participant

We recruited 17 participants (11 males and six females). There was no overlap with participants from the previous two user studies. The group included undergraduate, graduate students and staffs with backgrounds of either computer science or digital media technology. Their ages ranged from 21 to 36. All participants declared normal or corrected vision, with no color-blindness.

### 6.3 Procedure

We compared our system with the standard Unity in terms of efficiency. Figure 17(a) and (b) demonstrates the user interface from our system and the standard Unity coloring tool. Participants were divided into two groups, using the two systems in random orders. When using the standard Unity, participants were instructed to try their best to colorize all the objects in the scene. While using our system, the participants were asked to explore the presented colorization suggestions until they were satisfied. Participants were finally asked to complete the SUS questionnaire and then went through semi-structured interviews. Some questions were raised according to their SUS scoring, while some others were raised randomly. Qualitative data were transcribed for analysis. A real case study was finally performed to further demonstrate the usability of the system.

### 6.4 Finding

From Figure 17(c), we can observe an apparent advantage of our system over the standard Unity tool in terms of efficiency. Users normally can complete the task using our tool within 2 minutes, while it takes around 10 minutes to complete the task manually using the standard Unity tool. We also notice that participants may miss the colorization of some objects, with an averaged missing rate about 8%, when using the standard Unity for colorization. This confirms the advantage and necessity of developing an automatic colorization tool for a 3D indoor scene.

As shown in Figure 18, our system ($M = 77.79$, $SD = 11.10$) achieves better results against Raw Unity ($M = 72.42$, $SD = 18.10$) in almost every aspect of the SUS test, except for Question 4. Participants explained that they need certain help in the use of eye-tracking devices. As the distributions of the averaged score for both systems are normal ($p = 0.9$ and 0.6 for our system and raw unity respectively), one-way ANOVA was then used for analysis, drawing a conclusion that our system is significantly better than the raw unity ($p = 0.034$, $F = 4.961$). We further conducted a post-hoc

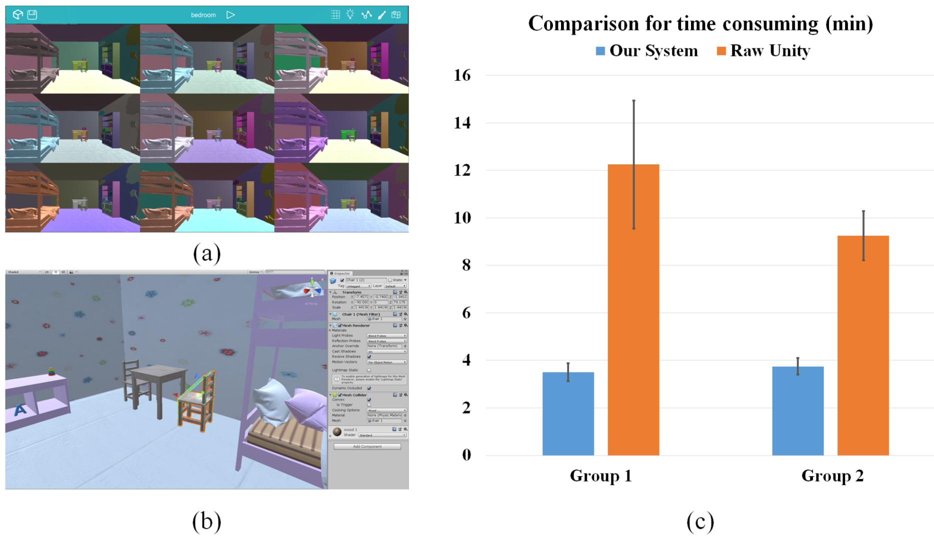



Fig. 17. System usability: (a) our colorization system; (b) the standard unity; (c) performance comparison between our system and the standard unity. The vertical axis is the average time while the horizontal axis indicates the user groups using the two systems in different orders (Group 1 used our system first and Group 2 used the standard unity first).

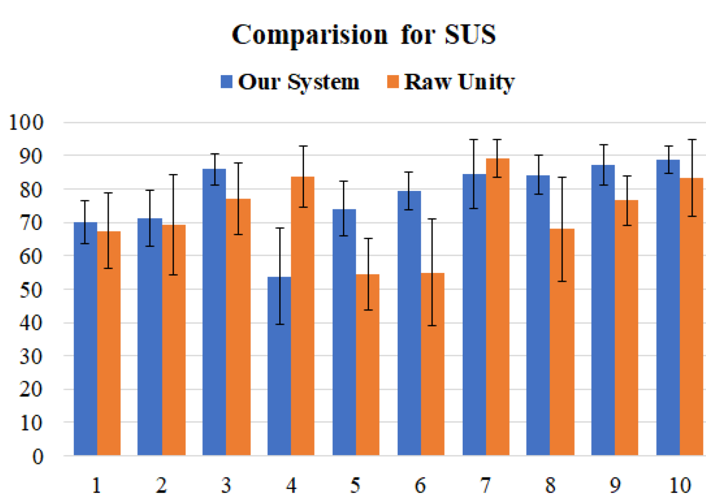



Fig. 18. Results from the SUS evaluation. The numbers in horizontal axis correspond the the number of questions in Table 5, while the vertical axis refers to the score rate by participants in 0–100 scale.

test (t-test), which also indicated a significantly better scoring of our system ($t = 2.592$, $p = 0.015$). According to [2], An SUS score of 70 or above shows that the system is at least passable while systems with scores above 90 are considered highly usable.

Finally, we also compared it with existing data-driven methods to further demonstrate the advantage of our system. Figure 19 shows the colorized scenes with our method and a data-driven system, MagicDecorator [9]. The colorized scenes with our method were generated by different users. The diverse colorization style conforms to the personalized preferences from different

Fig. 19. Subjective comparison: the top row shows suggestions presented by our system while the bottom rows shows those presented by MagicDecorator [9].

Table 6. Statistics of Thematic Analysis Result

| Themes | Total | % |
|---|---|---|
| T1: The proposed system does provide user-preferred high-quality colorizations | 41 | 27.7 |
| T2: The proposed system is more convenient than raw unity | 40 | 27.1 |
| T3: Suggestions provided by the system are creative | 20 | 13.5 |
| T4: The viewpoints provided by the proposed system are good | 15 | 10.1 |
| T5: Manual colorization requires some professional skills | 15 | 10.1 |
| T6: Manual colorization is not WYSWYG | 17 | 11.5 |
| Total | 148 | 100 |

participants. This confirms the effectiveness of our method in satisfying user preference. The visual results show that the incorporation of preference improves the reasonability of suggestions. Meanwhile, the inspiring effect is well maintained. Finally, the user preference towards the system output is enhanced.

### 6.5 Subjective Feedback

Transcripts were thoroughly read and analyzed in line with the inductive Thematic Analysis approach [4] by the second, third, and fifth authors. To be specific, transcripts were firstly coded independently by two coders (second and third authors) with in-vivo codes to obtain an initial code book. Then a group discussion was performed by the three researchers to resolve conflicts and refine the code book. Open and axial coding was used during group discussion. Finally, the third author coded all interviews and the group discussion together. We assigned a total of 148 codes that fell into six main themes. For an overview of relative distribution of themes please see Table 6.

*Theme 1: The proposed system does provide user-preferred high-quality colorizations.* Most participants (13 out of 16, 41 codes) were satisfied with the colorizations generated by our system. They agreed that these colorizations were diverse (involving various styles) and looked good. “I like colorizations provided by the system, they look more comprehensive. Colorizations provided by the system look nice. It will try to generate different colorizations and provide enough suggestions.

These suggestions are useful and the recommendation is successful" (P5). Participants also agreed that the colorizations do gradually approach their preferences in the evolving process. For example, "The suggestions are quite accurate. I love those within the same color series and the system knows that. About 80% of the proposed colorizations fit my preference. It does evolve towards my preference" (P15). Some participants thought the colorizations gradually matched his expectation, even though he was actually not sure about his own preference. "I am actually not so clear about what I prefer, but the system can still show that".

*Theme 2: The proposed system is more convenient than raw unity.* Most participants (13 out of 16, 40 codes) agreed that colorization with the raw unity was complex, tedious, in either choosing colors, moving the camera or the repeated operations. And also it was time consuming. While our system was more simple, convenient and time-saving. Some participants emphasized that it was not easy to figure the whole process with raw unity. Especially, it was tedious to accurately select the object and accurately specified the color for the object. For example, "Repeated mouse operations is tiring. Although you can slowly adjust, it is really time consuming" (P16), "The system is simple and convenient. It has colorized the scene well and thus you can intuitively decide whether the colorization is appropriate or not" (P4). Participants thought the interface was clear and they could easily find the preferred colorizations. Some participants mentioned that "The eye-tracking manner helps to free the hands. And I am also not so hesitate", "I feel everything is under control by generating the whole result immediately. And it is much more conveniently."

*Theme 3: Suggestions provided by the system are creative.* 10 out of the 16 participants (20 codes) pointed out that the system gave them unexpected colorizations and taught them new color combinations. As a contrast, many participants claimed that they chose colors instinctively in manual colorization. And the colorization provided by the system is more audacious while in harmony. "I colorize the objects in a pure subjective way under manual setting. For example, I will colorize a chair in what I think it should be. Many color combinations provided by the system are out of my expectations, but look good in my opinion" (P1), "It can balance different colors well. Some colorizations are comfortable, although you may not think of when coloring the scene by yourself". Participants also mentioned that they were inspired by the suggestions of the system, which let them realize different styles and moods of colorization. "I am not so care about colors in my life before. But now I feel they are so important after using the system. The system lets me pay more attention to colors in my personal life space" (P2).

*Theme 4: The viewpoints provided by the proposed system are good.* About half of the participants (nine out of 16, 15 codes) had comments on view points. Participants thought it is tedious to switch viewpoint in the raw unity. "It is not so convenient to use mouse to switch viewpoints, especially to specify a certain viewpoints. My viewpoint ran out the room frequently, which makes me frustrated" (P5). Most participants thought the viewpoints provided by the system did cover the scene well. "It is glad to be able to switch among several key viewpoints, making everything in the scene quite clear". Some participants mentioned that it would be better if the details of object could also be shown, so that they could be clearer about the colors of every objects. "I think the viewpoints provided by the system cover the whole scene well, which is what you need in the first place. It would be even better if some details can also be shown" (P11).

*Theme 5: Manual colorization requires some professional skills.* Nine out of the 16 participants (15 codes) mentioned that manual colorization required some professional knowledge, either in software usage or in color theory. For examples, some participant pointed out that "it is difficult to find appropriate color when using the raw unity", and "he/she cannot figure out good colorization with the raw unity system due to the lack of aesthetics." P3 said: "The manual colorization does not provide any guidance. And I need to figure out the colorization all by myself. Some knowledge about color theory is needed. It maybe easy for professionals, but troublesome for novices like me,


Fig. 20. House in decoration. The top row shows floor plan and some interior view of the real house, while the bottom row shows corresponding screenshots of the virtual one.

Table 7. Statistics in Real Case Colorization

| Cell | #5 | #1 | #7 | #6+#4 | #2 | #3 |
|---|---|---|---|---|---|---|
| Timing (sec) | 225 | 132 | 98 | 292 | 54 | 115 |
| Rounds | 12 | 10 | 9 | 15 | 2 | 9 |
| Times of manual adjustment | 2 | 1 | 0 | 3 | 0 | 1 |

who are also not good at aesthetics". Some participants thought it was quite complicated to operate with the raw unity in colorization. "There are so many color settings and viewpoint switches. Practices are required on how to adjust the camera and how to color an object" (P3).

*Theme 6: Manual colorization is not WYSWYG.* Half of the participants (eight out of 16, 17 codes) reflected that the selected color was not the same as it is put on the object, probably due to lighting condition, the material of the object or personal sensitivities on color. "I felt the selected color went through some weird changes under lighting. I need to adjust the color again and again, due to such difference" (P12). While with our system, participants felt "you can intuitively see whether the colorization is appropriate or not. And you can easily make comparisons" (P3).

### 6.6 Use Case

A young lady (one of the 17 participants, aged 40), who had just bought a house, was invited to use our system for interior design. Note she does not participant in any other tasks of the user study.

We first built a 3D model of the house according to the provided floor plan. Then a set of furniture was picked up and populated into each room according to the description of the participant. After that, we introduced our system to the participant and showed her how to use the system within 5 minutes in total (the same as for other participants). She was then given 10 minutes (twice the time of other participants) to get familiar with the system after eye-tracking calibration. Finally, the participant started to colorize cells of the house one by one. The color of a primitive shared by several cells were kept unchanged after the first colorization, unless manually altered by the user.

It took the artist about 1 hour and 40 minutes to build the empty house, 1 day to communicate with the participant on placement of furniture for several rounds (about 2 hours and 20 minutes for actual placement in total). Figure 20 shows the real house and its virtual counterpart. The participant was excited at the beginning and felt amazing about the eye-tracker, kept asking "can it really capture my preference?" The participant began colorization from the first bedroom (#5)

Fig. 21. Colorizations of the House. The first row shows cells without colorizations while the second to the fourth rows show different colorizations of each cell.

followed by the study (#1), the second bedroom (#7), the living room (#6) together with the dining room (#4), the kitchen (#2) and finally the toilet (#3). The balcony (#8) was left without colorization. Figure 21 shows the house with furniture populated, and some colorizations generated by the user. Some statistics about the colorization process are listed in Table 7. The participant spent quite a lot of time and rounds in the first try (the first bed room), which is easy to understand. She then proceeded quite quickly for the study and the second bedroom. She also paid quite a lot of time for the living room and dining room due to their complexity. She even flashed across for the kitchen. Surprisingly, she spent quite a long time on the toilet, comparable to the second bedroom. She explained in the final interview that she happened to find satisfactory colorization for the kitchen at the beginning. And for the toilet, she was struggling on 2-3 different suggestions. Generally, the participant was satisfied with the whole process. “I can try various styles. It provides much more styles than traditional decoration manner”. For preference capturing, she felt “the suggestions seem to gradually approach my expectation, although I even do not have any idea at the beginning. It is very nice to freely explore different colorizations in such a convenient manner.” However, she also agreed with some other participants that some of the colorization were too bright and not suitable for some rooms such as the kitchen or the toilet. For these rooms, “It maybe enough to use manual colorization as they are relatively simple, especially if there are some tools to allow easy selection of semantically-related objects which are usually colored together.” Finally, she felt it is also quite tedious to communicate with the artist on placement of furniture in the house. Due to the lack of enough samples for selection, they needed to discuss for several rounds and made some compromise. So it is better for the system to have some modeling functions in the future, or at least have a large database of furniture and related interfaces for exploration.

# 7 DISCUSSION

## 7.1 Summary and Implications

In summary, we received quite positive feedback on the system. Here, we would like to discuss our findings in light of the viewpoints raised in the related work section.

- **The effectiveness of eye tracking in playing its own roles.** Eye tracking serves for two purposes in our research: representative view selection and user preference capturing. The results in Figure 4 illustrate the effectiveness of the 3D attractiveness based view evaluation, as views in line with the expectation of most people (the sofa area in (a) and dining area in

(b) ) are rated much higher than others (the corner area in (c)). The steady convergence of MCMC sampler (Figure 5) demonstrates the feasibility of the new representative view selection scheme. The ablation study in user study 2 (Section 5) clearly shows that the integration of eye-tracked user preference does improve the quality of suggestions. As we can see from Figure 14(b), the integration took effect immediately from the first round and continued to the end, with higher quality rating in the whole evolving process against the case without considering preference.

— **Necessity of convergent thinking in a creativity support system.** The evolving record of suggestion diversity and quality in Figure 14 together imply the existence of convergent thinking after integration of user preference. While the CSI rating and paired comparison in Figure 15 demonstrate the positive effect of such thinking in a creativity support system.
— **Advantage of preference capturing with eye tracking.** From Figure 11, we can see preference capturing with an eye tracker took less time to complete the task by average and varied less among participants according to the standard deviation. While Figure 12 reveals more consistent choices of participants under the mode of eye tracking. Together, the two figures validate the less hesitance with an unconscious intent capturing method such as eye tracking.

We have also learned some lessons, which would be helpful to the design of a general creativity support system with personalized preferences.

— **Gaze and head movement information are very useful in capturing the user's unconscious intent, however, bias needs to be carefully handled.** This is validated by the SUS Test and by the feedback comments given in the interview. Especially, "the user may be attracted due to over-saturated colorizations" as indicated by P1.
— **Duration allowed for the participant to observe suggestions is important.** The user needs some time to do a primary observation before determining the preferred suggestions.
— **The workflow should be semi-automatic rather than fully-manual nor fully-automatic.** Some participants reflected that the colorizations are not totally satisfactory, therefore, manual refinements are expected.
— **Both divergent thinking and convergent thinking are important for a creativity support system;** however, careful balance should be made.

### 7.2 Limitations and Future Works

Some limitations are exposed during the user study:

— Unsatisfactory colorization. Participants have mentioned that some colorizations are over-saturated while some others are not suitable for daily life Besides, colorizations are usually not totally satisfactory.
— Confusing user interaction. Some participants mentioned that instructions to use the system were expected before the start of the experiment.

They shed light on future directions to improve the system:

— Improvements of colorization results. We would further explore the introduction of new constraints to avoid over-saturated colorization and to fit the context of the interior scene.
— Manual refinement. We may provide various manual editing tools such as adjustment of harmony template and object color, so that the user can obtain a final satisfactory result by refining the suggested colorization.
— Guidance. We will provide system instructions during the operation to aid the user.

## 8 CONCLUSION

Colorization of 3D interior scenes is a creative task. This article presents a suggestive interior colorization system with automatic omni-directional color optimization and integration of the user preference in suggestion generation. The user's visual attention inferred from gaze data is exploited to help determine representative views in the 3D scene for color optimization and to capture his/her preference towards presented suggestions in each evolutionary round. Automatic colorization of scene objects and light sources is formulated as an optimization problem with the objective function defined on representative viewing images. The user preference is integrated into the objective function together with other factors (such as harmony, mood, and contrast) to refine and develop most satisfactory colorization. We validate the heuristics of our system through a series of user studies and demonstrate the usefulness and participants' satisfaction of the system. As a result, our proposed colorization method has been experimentally proved to have a potential advantage in designing colorization of 3D interior scenes.

## ACKNOWLEDGMENTS

We thank the anonymous reviewers for their constructive comments.

## REFERENCES

[1] Cecilia R. Aragon and Alison Williams. 2011. Collaborative creativity: A complex systems model with distributed affect. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1875–1884.

[2] Aaron Bangor, Philip T. Kortum, and James T. Miller. 2008. An empirical evaluation of the system usability scale. *International Journal of Human-computer Interaction* 24, 6 (2008), 574–594.

[3] Luca Benedetti, Holger Winnemller, Massimiliano Corsini, and Roberto Scopigno. 2014. Painting with Bob: Assisted creativity for novices. In *Proceedings of the Annual ACM Symposium on User Interface Software and Technology*. 419–428.

[4] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101.

[5] John Brooke. 1996. *SUS: A Quick and Dirty Usability Scale*. CRC Press, Bristol, PA, 189–194.

[6] Dennis R. Brophy. 2001. Comparing the attributes, activities, and performance of divergent, convergent, and combination thinkers. *Creativity Research Journal* 13, 3–4 (2001), 439–455.

[7] Erin A. Carroll, Celine Latulipe, Richard Fung, and Michael A. Terry. 2009. Creativity factor evaluation: Towards a standardized survey metric for creativity support. In *Proceedings of the ACM Conference on Creativity and Cognition*. 127–136.

[8] Siddhartha Chaudhuri and Vladlen Koltun. 2010. Data-driven suggestions for creativity support in 3D modeling. *ACM Transactions on Graphics* 29, 6 (2010), 1–10.

[9] Kang Chen, Kun Xu, Yizhou Yu, Tian-Yi Wang, and Shi-Min Hu. 2015. Magic decorator: Automatic material suggestion for indoor digital scenes. *ACM Transactions on Graphics* 34, 6 (2015), 1–11.

[10] Li Chen and Feng Wang. 2016. An eye-tracking study: Implication to implicit critiquing feedback elicitation in recommender systems. In *Proceedings of the Conference on User Modeling Adaptation and Personalization*. 163–167.

[11] Xiu Chen and Zhenzhong Chen. 2017. Exploring visual attention using random walks based eye tracking protocols. *Journal of Visual Communication and Image Representation* 45 (2017), 147–155.

[12] Erin Cherry and Celine Latulipe. 2014. Quantifying the creativity support of digital tools through the creativity support index. *ACM Transactions on Computer-human Interaction* 21, 4 (2014), 1–25.

[13] Nicholas Davis, Alexander Zook, Brian O'Neill, Brandon Headrick, Mark Riedl, Ashton Grosz, and Michael Nitsche. 2013. Creativity support for novice digital filmmaking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 651–660.

[14] Herie B. de Vries and Todd I. Lubart. 2017. Scientific creativity: Divergent and convergent thinking and the impact of culture. *Journal of Creative Behavior* 53, 2 (2017), 145–155.

[15] E. Frias-Martinez, G. Magoulas, R. Macredie, and S. Chen. 2005. Modeling human behavior in user-adaptive systems: Recent advances using soft computing techniques. *Expert Systems with Applications* 29, 2 (2005), 320–329.

[16] Jonas Frich, Michael Mose Biskjaer, and Peter Dalsgaard. 2018. Why HCI and creativity research must collaborate to develop new creativity support tools. In *Proceedings of the Technology, Mind, and Society*. 1–6.

[17] Jonas Frich, Lindsay MacDonald Vermeulen, Christian Remy, Michael Mose Biskjaer, and Peter Dalsgaard. 2019. Mapping the landscape of creativity support tools in HCI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*.
[18] Chi-Wing Fu, Jiazhi Xia, and Ying He. 2010. LayerPaint: A multi-layer interactive 3D painting interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 811–820.
[19] Agostino Gibaldi, Mauricio Vanegas, Peter J. Bex, and Guido Maiello. 2017. Evaluation of the Tobii EyeX eye tracking controller and Matlab toolkit for research. *Behavior Research Methods* 49 (2017), 923–946.
[20] Walter R. Gilks, Sylvia Richardson, and David Spiegelhalter. 1996. *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC.
[21] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. 2016. Let there be color! Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. *ACM Transactions on Graphics* 35, 4 (2016), 1–11.
[22] Tobii Inc. 2021. Tobii support tutorial. Retrieved Dec 2021 from https://help.tobii.com/hc/en-us/sections/202845209-Tutorials.
[23] Arjun Jain, Thorsten Thormhlen, Tobias Ritschel, and Hans-Peter Seidel. 2012. Material Memex: Automatic material suggestions for 3D objects. *ACM Transactions on Graphics* 31, 6 (2012), 1–8.
[24] Auk Kim and Gahgene Gweon. 2014. Photo sharing of the subject, by the owner, for the viewer: Examining the subject's preference. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 975–978.
[25] Yuki Koyama, Daisuke Sakamoto, and Takeo Igarashi. 2016. SelPh: Progressive learning and support of manual photo color enhancement. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 2520–2532.
[26] D. G. Krige. 1951. A statistical approach to some basic mine valuations problems on the witwatersrand. *Journal of the Chemical, Metallurgical and Mining Engineering Society of South Africa* 52, 6 (1951), 119–139.
[27] Kuno Kurzhals, Fabian Gbel, Katrin Angerbauer, Michael Sedlmair, and Martin Raubal. 2020. A view on the viewer: Gaze-adaptive captions for videos. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–12.
[28] D Lamming. 1991. *Contrast Sensitivity*. Macmillan Press, London.
[29] Celine Latulipe, Ian Bell, Charles L. A. Clarke, and Craig S. Kaplan. 2006. symtone: Two-handed manipulation of tone reproduction curves. In *Proceedings of the Graphics Interface*.
[30] George Leifman and Ayellet Tal. 2012. Mesh colorization. *Computer Graphics Forum* 31, 2 (2012), 421–430.
[31] Anat Levin, Dani Lischinski, and Yair Weiss. 2004. Colorization using optimization. *ACM Transaction on Graphics* 23, 3 (2004), 689–694.
[32] Yuan Liang, Xiting Wang, SongHai Zhang, ShiMin Hu, and Shixia Liu. 2018. PhotoRecomposer: Interactive photo recomposition by cropping. *IEEE Transactions on Visualization and Computer Graphics* 24, 10 (2018), 2728–2742.
[33] Juncong Lin, Pintong Xiao, Yinan Fu, Yubin Shi, Hongran Wang, Shihui Guo, Ying He, and Tong-Yee Lee. 2020. C3Assignment: Camera CubeMaps color assignment for creative interior design. *IEEE Transactions on Visualization and Computer Graphics* (2020).
[34] Sharon Lin, Daniel Ritchie, Matthew Fisher, and Pat Hanrahan. 2013. Probabilistic color-by-numbers: Suggesting pattern colorizations using factor graphs. *ACM Transactions on Graphics* 32, 4 (2013), 1–12.
[35] Yuyu Lin, Jiahao Guo, Yang Chen, Cheng Yao, and Fangtian Ying. 2020. It is your turn: Collaborative ideation with a co-creative robot through sketch. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–14.
[36] Fan Liu, Zhiyong Cheng, Changchang Sun, Yinglong Wang, Liqiang Nie, and Mohan Kankanhalli. 2019. User diverse preference modeling by multimodal attentive metric learning. In *Proceedings of the 27th ACM International Conference on Multimedia*. 1526–1534.
[37] Shuai Ma, Zijun Wei, Feng Tian, Xiangmin Fan, Jianming Zhang, Xiaohui Shen, Zhe Lin, Jin Huang, Radomír Měch, Dimitris Samaras, and Hong'an Wang. 2019. SmartEye: Assisting instant photo taking via integrating user preference with deep view proposal network. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–12.
[38] Neil Maiden, Konstantinos Zachos, Amanda Brown, George Brock, Lars Nyre, Aleksander Nygrd Tonheim, Dimitris Apsotolou, and Jeremy Evans. 2018. Making the news: Digital creativity support for journalists. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*.
[39] Sebastian Marwecki, Andrew David Wilson, Eyal Ofek, Mar Gonzalez-Franco, and Christian Holz. 2019. Mise-Unseen: Using eye tracking to hide virtual reality scene changes in plain sight. In *Proceedings of the Annual ACM Symposium on User Interface Software and Technology*. 777–789.
[40] David Massimo, Mehdi Elahi, and Francesco Ricci. 2017. Learning user preferences by observing user-items interactions in an IoT augmented space. In *Proceedings of the Adjunct Publication of the Conference on User Modeling, Adaptation and Personalization*. 35–40.
[41] Yutaka Matsuda. 1995. Color design. *Asakura Shoten* 2, 4 (1995), 10.

[42] Sven Mayer, Gierad Laput, and Chris Harrison. 2020. Enhancing mobile voice assistants with WorldGaze. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* 1–10.

[43] George A. Miller. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review* 63, 2 (1956), 81–97.

[44] Tricia J. Ngoon, C. Ailie Fraser, Ariel S. Weingarten, Mira Dontcheva, and Scott Klemmer. 2018. Interactive guidance techniques for improving creative feedback. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.*

[45] Chuong H. Nguyen, Tobias Ritschel, Karol Myszkowski, Elmar Eisemann, and Hans-Peter Seidel. 2012. 3D material style transfer. *Computer Graphics Forum* 31, 2 (2012), 431–438.

[46] Jakob Nielsen and Kara Pernice. 2009. *Eyetracking Web Usability.* New Riders.

[47] Peter O'Donovan, Aseem Agarwala, and Aaron Hertzmann. 2015. DesignScape: Design with interactive layout suggestions. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems.* 1221–1224.

[48] Michal Ortega and Thomas Vincent. 2014. Direct drawing on 3D shapes with automated camera control. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 2047–2050.

[49] Julie Pallant. 2011. *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using the SPSS Program.* Allen & Unwin, Crows Nest.

[50] Frank Pallas, Max-Robert Ulbricht, Lorena Jaume-Palasí, and Ulrike Höppner. 2014. Offlinetags: A novel privacy approach to online photo sharing. In *Proceedings of the CHI'14 Extended Abstracts on Human Factors in Computing Systems.* 2179–2184.

[51] Katharina Reinecke, Tom Yeh, Luke W. Miratrix, Rahmatri Mardiko, Yuechen Zhao, Jenny Liu, and Krzysztof Z. Gajos. 2013. Predicting users' first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 2049–2058.

[52] Dario D. Salvucci and Joseph H. Goldberg. 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the Symposium on Eye Tracking Research and Applications.* 71–78.

[53] Anthony Santella and Doug DeCarlo. 2004. Robust clustering of eye movement recordings for quantification of visual interest. In *Proceedings of the 2004 Symposium on Eye Tracking Research and Applications.* 27–34.

[54] Manolis Savva, Angel X. Chang, Pat Hanrahan, Matthew Fisher, and Matthias Niener. 2014. SceneGrok: Inferring action maps in 3D environments. *ACM Transactions on Graphics* 33, 6 (2014), 1–10.

[55] Korok Sengupta, Sabin Bhattarai, Sayan Sarcar, I. Scott MacKenzie, and Steffen Staab. 2020. Leveraging error correction in voice-based text entry by talk-and-gaze. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* 1–11.

[56] Asma Shakil, Christof Lutteroth, and Gerald Weber. 2019. CodeGazer: Making code navigation easy and natural with gaze input. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* 1–12.

[57] Ben Shneiderman. 2007. Creativity support tools: Accelerating discovery and innovation. *Communications of the ACM* 50, 12 (2007), 20–32.

[58] Pao Siangliulue, Joel Chan, Steven Dow, and Krzysztof Z. Gajos. 2016. IdeaHound: Improving large-scale collaborative ideation with crowd-powered real-time semantic modeling. In *Proceedings of the Annual Symposium on User Interface Software and Technology.* 609–624.

[59] Jacob Solomon. 2016. Heterogeneity in customization of recommender systems by users with homogenous preferences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 4166–4170.

[60] Hyejin Song and Nammee Moon. 2019. Eye-tracking and social behavior preference-based recommendation system. *The Journal of Supercomputing* 75 (2019), 1990–2006.

[61] Enkelejda Tafaj, Gjergji Kasneci, Wolfgang Rosenstiel, and Martin Bogdan. 2012. Bayesian online clustering of eye movement data. In *Proceedings of the Symposium on Eye Tracking Research and Applications.* 285–288.

[62] Moshe Unger, Bracha Shapira, Lior Rokach, and ArielBar. 2017. Inferring contextual preferences using deep autoencoding. In *Proceedings of the Conference on User Modeling, Adaptation, and Personalization.* 221–229.

[63] Juan D. Velsquez. 2013. Combining eye-tracking technologies with web usage mining for identifying Website Key objects. *Engineering Applications of Artificial Intelligence* 26, 5–6 (2013), 1469–1478.

[64] E. Villalba. 2017. *Critical Thinking in Relation to Creativity.* Elsevier.

[65] Simon Voelker, Sebastian Hueber, Christian Holz, Christian Remy, and Nicolai Marquardt. 2020. GazeConduits: Calibration-free cross-device collaboration through gaze and touch. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* 1–10.

[66] Baoyuan Wang, Yizhou Yu, Tien-Tsin Wong, Chun Chen, and Ying-Qing Xu. 2010. Data-driven image color theme enhancement. *ACM Transactions on Graphics* 29, 6 (2010), 1–10.

[67] Yingbin Wang, Xiu Chen, and Zhenzhong Chen. 2016. Towards region-of-attention analysis in eye tracking protocols. In *Proc. IS&T Int'l. Symp. on Electronic Imaging: Visual Information Processing and Communication VII.* 1–6.

[68] Kai Xu, Hao Zhang, Daniel Cohen-Or, and Baoquan Chen. 2012. Fit and diverse: set evolution for inspiring 3D shape galleries. *ACM Transactions on Graphics* 31, 4 (2012), 1–10.
[69] L. Yuan, W. Zeng, S. Fu, Z. Zeng, H. Li, C.-W. Fu, and H. Qu. 2021. Deep colormap extraction from visualizations. *IEEE Transactions on Visualization and Computer Graphics* (2021).
[70] Zhenpeng Zhao, Sriram Karthik Badam, Senthil Chandrasegaran, Deokgun Park, Niklas Elmqvist, Lorraine G. Kisselburgh, and Karthik Ramani. 2014. skWiki: A multimedia sketching system for collaborative creativity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1235–1244.
[71] Jie Zhu, Yanwen Guo, and Han Ma. 2018. A data-driven approach for furniture and indoor scene colorization. *IEEE Transactions on Visualization and Computer Graphics* 24, 9 (2018), 2473–2486.
[72] Oleg Åăpakov and Darius Miniotas. 2007. Application of clustering algorithms in eye gaze visualizations. *Information Technology and Control* 36, 2 (2007), 213–216.

Received 3 June 2021; revised 12 April 2022; accepted 11 May 2022