Mood-Driven Colorization of Virtual Indoor Scenes

Michael Solah[*] George Mason University
Haikun Huang[†] George Mason University
Jiachuan Sheng[‡] Tianjin University of Finance & Economics
Tian Feng[§] Zhejiang University
Marc Pomplun[¶] University of Massachusetts Boston
Lap-Fai Yu[‖] George Mason University

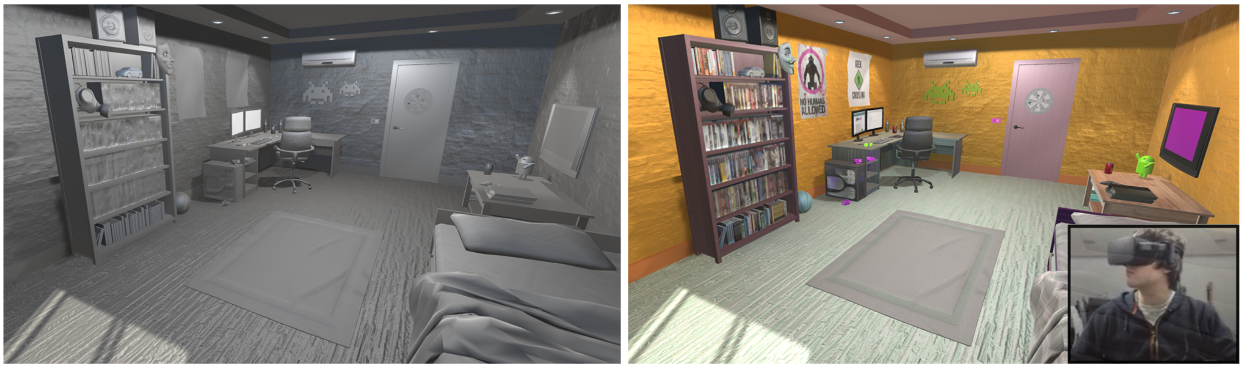


Fig. 1: We present an approach to automatically color the textures of objects in a virtual scene for invoking a desired type of mood. In this example, our approach colorizes a virtual bedroom according to the "cheerful" mood.

Abstract—
One of the challenging tasks in virtual scene design for Virtual Reality (VR) is causing it to invoke a particular mood in viewers. The subjective nature of moods brings uncertainty to the purpose. We propose a novel approach to automatic adjustment of the colors of textures for objects in a virtual indoor scene, enabling it to match a target mood. A dataset of $25,000$ images, including building/home interiors, was used to train a classifier with the features extracted via deep learning. It contributes to an optimization process that colorizes virtual scenes automatically according to the target mood. Our approach was tested on four different indoor scenes, and we conducted a user study demonstrating its efficacy through statistical analysis with the focus on the impact of the scenes experienced with a VR headset.

1 Introduction

Virtual reality (VR) enables a highly immersive experience for digital storytelling, movies, and gaming applications. Among all the factors in virtual scene design, the user's mood caused by such an experience has always been regarded as a crucial one. For example, a fantasy game designer would like to decorate the scene with vibrant colors to invoke a mood of cheerfulness and excitement.

Virtual scene colorization is a tedious process involving many manual operations. It consumes time due to massive trial-and-errors for the mood desired by the designer. The previous successful practices in the tasks relevant to virtual scene design [2, 15, 20, 29, 43, 47] motivated us to exploit optimization techniques against the problem.

In this paper, we propose an approach to automatic adjustment of the colors of textures for objects in a virtual indoor scene so as to match a target mood, facilitating the tasks of virtual scene design (Figure 1). Based on a data-driven optimization process, it edits the input scene through iterations. As the problem requires support in mood recognition, we employ deep learning through a fine-tuned Convolutional Neural Network (CNN) for feature extraction from images for moods. Approaches have focused on color schemes [6, 8, 12], but we provide an approach that works automatically with features from a CNN. The extracted features construct a classifier for mood evaluation of the virtual indoor scene. With our approach, we plan to provide a method to generate colors for a scene that evoke a certain mood without having to depend on pre-conceived color patterns. The contributions of this paper include:

- proposing a novel approach to automatic adjustment of the colors of textures for objects in a virtual indoor scene, enabling the scene to match a target mood;
- a mood classifier trained on a dataset consisting of $25,000$ images for five different moods.
- conducting a user study based on statistical analysis for validating the approach's efficacy (i.e., if the target mood can be identified by users and enhanced in the scenes experienced with a VR headset).

*e-mail: msolah@gmu.edu
†e-mail: hhuang25@gmu.edu
‡e-mail: jiachuansheng@tjufe.edu.cn
§e-mail: t.feng@zju.edu.cn
¶e-mail: marc@cs.umb.edu
‖e-mail: craigyu@gmu.edu

*Michael Solah and Haikun Huang contributed equally to this paper.
‡Jiachuan Sheng is the corresponding author of this paper.

2 Related Work

Our work on mood-driven colorization of virtual indoor scenes relates to the areas of *colors and moods*, *virtual indoor scene design* and *image classification with CNNs*. We provide a brief literature review.

2.1 Colors and Moods

Studies on colors and moods date back to the 19th century (i.e., a survey of past empirical and theoretical work about color and psychological functioning by Elliot [13]). Recently, Pelowski et al. [35] highlighted that colors and moods were significantly interrelated in the study of human's psychological responses to art pieces.

Through a series of user studies, Kurt and Osueke [25] presented that combinations of colors for architectural interiors affected human's mood and perception. Furthermore, Elliot and Maier [14] validated that colors impacted a variety of psychological aspects, not limited to human's performance in tasks. Wardono et al. [44] proposed that the color scheme of an interior design influenced the behaviors and moods of diners.

Karayev et al. [22] classified images based on styles using an extensive and annotated database; They constructed classifiers with a variety of features, including those from Deep Neural Networks (DNNs) and from graph-based visual saliency. This work involved several styles (e.g., romantic, noir, pastel) and the classifiers for predicting the image's style achieved an accuracy of 78%. In comparison, our approach can generate results that match a certain mood, going beyond identification.

Colors and the moods that they may convey have been shown to have universal associations. For example, Adams and Osgood [1] conducted a literature review of over 89 papers finding overwhelmingly that there is shared interpretation in the cross-cultural meaning of color. That being said, there are specific cultural associations of colors that may not be universal, as pointed out by Jonauskaite et al. [21]. They tested emotional associations in a user study with 4,598 participants from 30 nations across 22 languages. They found not only strong universal associations between mood and color, but also strong local associations based on geographical proximity and linguistic similarity.

In our approach, we use an extensive and manually created dataset of images on various rooms to extract features via deep learning. Zhao et al. [49] developed a personality scoring network for judging the aspects/features of a poster determining its personality; Their approach achieved an average accuracy of 80% for classifying the personality labels of images. While we also aim to extract features for describing moods, the purpose of using such features is to optimize the perceived moods of virtual scenes so that they can be experienced by users in VR.

Lukas Brodschelm et al. [6] produced a coloring approach for images based on machine learning for Mood Adaptive Display Coloring. They alter the colors of what is displayed on screen to counteract negative emotions in a user, trying to take advantage of color psychology to alter the user's mood positively so that they can better focus on a task. Cha et al. [7] conducted an experiment to see the relationship between the colors of indoor virtual environments and mood. They found that color can affect a user's heart rate and task performance, further evidencing the value of coloring algorithms. Lin et al. [27] designed an optimization process to automatically suggest color designs for indoor scenes.

2.2 Virtual Indoor Scene Design

Previous attempts on automatic design of indoor scenes have inspired us significantly. For example, Yu et al. [47] proposed an approach for optimizing the positions and orientations of furniture in a 3D room, considering functionality and design parameters. Akazawa et al. [2] presented a method for generating 3D scenes based on contact restraints. Jiang et al. [20] devised a similar method for 3D scene synthesis, with 2D image rendering based on stochastic grammars. Wang et al. [43] explored indoor scenes synthesis with deep learning. Ma et al. [29] introduced a language-driven process for 3D scene synthesis from databases.

Another relevant popular topic is texture assignment for consistent and harmonious scenes. For example, Chen et al. [10] proposed an approach based on simulated annealing that produces realistic textures for objects in virtual rooms. Compared to its capability of producing convincing results, our approach can generate realistic scenes using a realness cost function in the optimization process, and change the hue, saturation, and luminance of textures to fit a specific mood.

Automated color assignment for indoor scenes has been previously explored by Chen et al. [9] using trained Bayesian networks, demonstrating several practical applications. Their approach works by taking colors from some objects and then extrapolating harmonious colors for the rest of the scene. Our approach does not need such reference objects. Liu et al. [28] proposed an approach of automatic wall color re-assignment. Zhu et al. [51] introduced a data-driven approach for indoor furniture colorization. Using a Bayesian network, they can extrapolate color schemes from an indoor scene and generate harmonious results. Our approach can generate realistic results as well as mood specific configurations, as well as mixtures between the two. Fu et al. [16] explored an interactive design method to colorize snapshots of indoor scenes. Our approach distinguishes itself with the focus on moods rather than styles. Specifically, this focus lies in the harmonious color assignment for objects, which invokes a target mood in the viewer. Furthermore, our approach targets the whole 3D scene.

Sangyoon and Choi [17] and Sra et al. [38] proposed methods for automatic creation of textures for 3D scenes by detecting mood from pieces of music. Inspired by their excellent results, our approach allows users to select moods for modifying 3D scenes. Automatic 3D construction of indoor scenes from real world data is also a research area in progress as detailed in a recent survey on the state of the art approaches by Pintore et al. [36]. Our approach could potentially be applied to such reconstructions as well.

2.3 Semi-automatic Photo Recoloring

Another topic that relates to our approach is semi-automatic photo recoloring. Cohen-Or et al. [12] designed an approach based on color histograms in order to select harmonic color schemes for coloring images. They do this by comparing hue histograms to color harmony templates that were obtained from previous experiments. The templates are scored based on how much of the color histogram of the image resides in the template space, and the colors are adjusted by using colors that reside only in the template. They used a Gaussian fall function.

Chang et al. [8] created an interactive palette tool for recoloring images. They adapted previous approaches with a more flexible scheme that can preserve monotonicity, using sophisticated feature preserving strategies to ensure that the color changes in images are natural. Nhao et al. [50] produced an approach that can automatically segment an image and adjust the colors based on an input color palette. The altered image preserved harmony in regards to the image colors. The authors accomplished this by using two networks - one for recoloring the images and another for appropriate segment selection.

2.4 Image Classification with CNNs

CNNs have shown outstanding performances in object detection benchmark datasets and have been effective for image classification. As described in an overview of advancements in CNN models, Sultana et al. [39] demonstrated their improvements over time and roles as the state-of-the-art solution to image classification.

Such considerable success relates closely to the architecture of CNNs. Compared to other DNN models, CNNs involve convolutional layers to process pixel groups in images, and sub-sampling layers to simplify the convolution results, for features extraction; A fully-connected layer is usually in charge of the output. CNNs can detect features invariant to certain transformations (e.g., rotation). Based on such an architecture, researchers have developed a range of different CNN models [39].

3 Methods

As depicted by Fig. 2, our approach takes as input a user-given mood and a virtual indoor scene (i.e., a 3D scene comprising a set of objects). An optimizer refines the colors of object's textures in the input scene to match the mood. During the optimization process, the colors of textures are iteratively adjusted until convergence, in accordance with a database that includes pre-defined textures and colors of real-world objects.

In this paper, we focus on five selected moods: *cheerful*, *melancholy*, *peaceful*, *romantic*, and *scary*. Our selection was based on the intention to test the representative moods sharing semantic similarities, among which users somehow could still differentiate. For example, *melancholy* and *scary* moods are generally described as "negative", but users are expected to distinguish results generated from such similar moods when experiencing them in VR.

Using a total of $25,000$ images, we trained a deep learning classifier for the corresponding moods. In particular, each mood relates to $5,000$

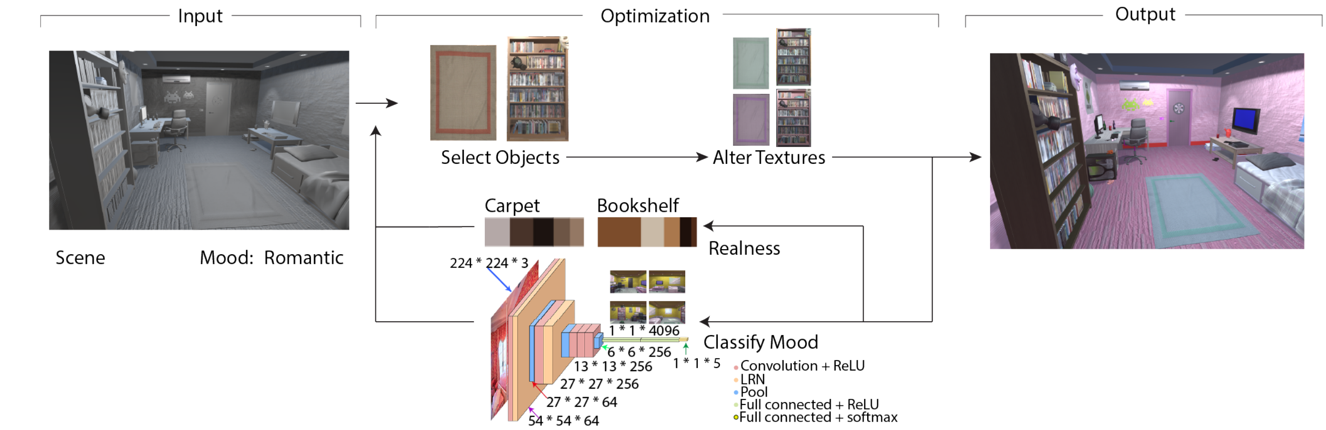


Fig. 2: An overview of our approach. The user first selects a mood type and a virtual room as input. The optimization process then takes place - objects in the scene are selected so their textures can be altered. Images are taken to capture the outlook of the current scene, based on which the current mood cost is calculated using a CNN classifier. A realness cost is also computed to determine how realistic the textures are. The optimization process proceeds iteratively. The output is the scene optimized with respect to the target mood.

images. All images were from indoor scenes, including *bedrooms*, *kitchens*, *living rooms*, *dining rooms*, and *others*.

Virtual scene creation can be time-consuming and challenging. Although researchers have proposed solutions to the relevant tasks and achieved promising results, they are yet to completely overcome an obvious drawback, that is, automation may weaken a designer's freedom. We seek an approach allowing designers to exploit automation, while not significantly altering their creative visions. Since convincing 3D scenes are integral to the creation of immersive VR experiences, we have confidence in our approach regarding its efficacy in ameliorating vagueness of the scenes that invoke a specific mood in the user.

4 Image Mood Classification

Since content-rich indoor scenes require a higher level of representation, hand-crafted features for mood detection would be ineffective and difficult. In comparison, CNNs build maps at both low and high levels to learn a hierarchy of features, which has shown excellent performances in many areas. This fact is rooted in that their architectures are more sophisticated than standard representations, as they comprise several layers of non-linear feature extractors. Therefore, we chose CNNs to solve the problem and believe they can suit the subjective nature of moods.

In this paper, we use the *VGG-F* model pre-trained on the *ImageNet* dataset to extract the features and use a CNN and SVM to classify the indoor images. The *VGG-F* model establishes an intrinsic representation of data because its deep structure is derived by extracting complex structures from a large amount of information; It has been applied successfully in different image classification tasks. For example, Paul et al. [33] predicted survival among patients with lung adenocarcinoma using features from a CNN based on the *VGG-F* model. We believe that the features of the *VGG-F* model pre-trained on ImageNet are sufficient for our goal, given the following pieces of evidence. Firstly, ImageNet is a rich dataset containing more than 10 million natural images. The contents and objects of our mood classification task decoration images are similar to real scenes from daily life so that some of them can be found in the ImageNet dataset. Hence the extracted features can directly or indirectly contain similar features of those decoration images. Secondly, to accommodate the model to the updated mood classification issue, we carry out dataset-specific fine-tuning, which improves the performance. We collect 25,000 indoor decoration images for five moods (5,000 images for each mood), and labeled them by mood. Then we fine-tuned the pre-trained VGG-F using those 25,000 images, where the batch size is 256, the learning rate is 0.0001, the training steps is 2000, and the number of classes is 5. We use 2,000 images to test our fine-tuned model, and the accuracy of classification is 90%, showing that the fine-tuned model is reliable for classifying the mood. The dataset will be released. Note that other datasets of interiors could also be used for the training.

The *VGG-F* model is a feed-forward neural network with five convolutional layers and three fully-connected layers. The input image is converted to have a dimension of 224×224 pixels. The architecture of the fine-tuned CNN model is shown in Fig. 2. The processed images are extracted by the filtering operation of the convolutional layers and the down sampling of the pooling layers. The features of each image are extracted as a 4096-dimensional feature vector, which is used to compute the classification result as the output. The last fully-connected layer has output dimensionality equal to the number of classes, in our case, the number of mood classes is 5. After extracting the features by fine-tuned VGG-F, we use SVM with RBF kernel to classify the mood and achieve 92% accuracy on the 2,000 test dataset. For further comparisons between the results of our CNN features and Five-Dominate Colors features classifiers, refer to our supplementary document. For an input image, a classification result is obtained for each mood type, and the probability of being classified to a certain mood type specified by the designer is regarded as the score of the input image for calculating the mood cost.

The *VGG-F* model has shown usefulness in practical applications. For example, Wozniak et al. [45] have shown that a retrained *VGG-F* model can be used for indoor location classification based on images. The ImageNet dataset [40] continues to be used extensively for the testing of state-of-the-art classification techniques.

Furthermore, Jain et al. used the VGG model to classify emotions from images of cartoons, achieving an accuracy of 95% and an F1 score of 0.85. It outperformed other tested models.

5 Technical Approach

Our approach aims to adjust the colors of textures for objects to match a target mood. To this end, we employ stochastic optimization in our approach with a total cost function, including a *mood* cost formula and a *realness* cost formula that drives textures to be colored similarly to the surfaces of objects in the real world. Due to the complexity of the large search space where the colors of textures are adjusted, we use a Markov Chain Monte Carlo (MCMC) technique with simulated annealing in the optimization process, which will be further discussed in Sect. 5.4.

5.1 Formulation

Let $\Phi = \{\phi_{\mathrm{i}} = (\phi_{\mathrm{i}}^{\mathrm{m}}, \phi_{\mathrm{i}}^{\mathrm{t}}, \phi_{\mathrm{i}}^{\mathrm{c}}, \phi_{\mathrm{i}}^{\mathrm{a}}, \phi_{\mathrm{i}}^{\alpha}) \mid i = 1, ..., p\}$ denote a set of the objects in the scene which comprise the material $\phi_{\mathrm{i}}^{\mathrm{m}}$ (e.g., marble, steel), the type $\phi_{\mathrm{i}}^{\mathrm{t}}$ (e.g., desk, chair.), the five-dominant-colors of the texture $\phi_{\mathrm{i}}^{\mathrm{c}}$, the surface area $\phi_{\mathrm{i}}^{\mathrm{a}}$, and the importance $\phi_{\mathrm{i}}^{\alpha}$ of each object ϕ_{i} in the input scene. While the material $\phi_{\mathrm{i}}^{\mathrm{m}}$, the type $\phi_{\mathrm{i}}^{\mathrm{t}}$, and the surface area $\phi_{\mathrm{i}}^{\mathrm{a}}$ are the inherent attributes of each object, the five-dominant-colors $\phi_{\mathrm{i}}^{\mathrm{c}}$ are extracted from the object's texture which will be described in Sect. 5.3.

Fig. 3: An example of images used to score the dining room during the optimization process. Each image follows the rule of thirds. The placement of cameras and the number of images can be changed by the user if needed.

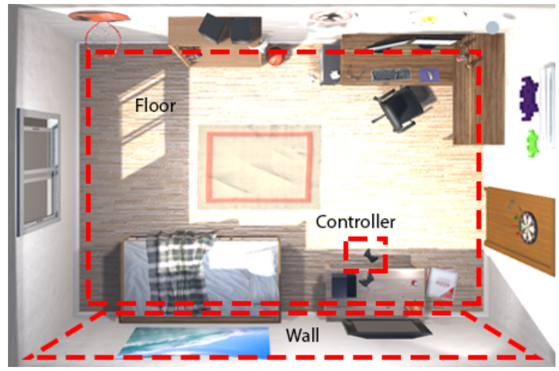


Fig. 4: A comparison of the normalized weights due to the sizes of different objects in the bedroom scene: The controller's weight is 0.003, the wall's weight is 0.081, and the floor's weight is 0.091. All weights are normalized between 0 and 1.

The quality of a scene Φ is evaluated by the total cost function:

$$C_{\text{Total}}(\Phi) = w_{\text{M}} C_{\text{M}} + w_{\text{R}} C_{\text{R}}(\Phi), \tag{1}$$

where C_{M} is the mood cost of the scene, C_{R} is the realness cost encoding how realistic the colors are for objects in the scene, w_{M} and w_{R} are the weights of the mood cost and the realness cost, respectively.

5.2 Mood Cost

Our classifier scores images of the scene based on the desired mood specified by the designer. According to the rule of thirds of photography [23] [24], we divide the camera's screen into three equal parts in the horizontal direction. In this way, there will be two critical focal areas on the camera's screen, in one-third and two-thirds in the horizontal direction. Empirically, if the input image follows the rule of thirds, the predicted result will be more accurate as the input image covers more regions of the scene. So, first, we place the camera in a location that can cover the main object such as a dining table or a bed. Then we aimed a certain focus area of the camera at the room's pillars to make the picture perspective. Most rooms are in a rectangular shape and have four pillars, so $n = 4$ is a common parameter choice. Fig. 3 shows an example of four images used to evaluate the mood of the bedroom scene. Most images used to train the classifier follow the rule of thirds. In cases that they do not, it is not a major issue since we focus primarily on the colors of the image. We define the mood cost as follows:

$$C_{\text{M}} = \frac{1}{n} \sum_{\text{i}=1}^{n} c_{\text{i}}, \tag{2}$$

where n refers to the number of pictures used; $c_{\text{i}} \in [0, 1]$ is the cost of an image i of the scene. We define the cost of an image i as follows:

$$c_{\text{i}} = w_0 m_0 + \sum_{\text{k}=1}^{4} w_{\text{k}} (1 - m_{\text{k}}), \tag{3}$$

where m_0 is the cost of the desired mood, and m_{k} encodes the cost of the moods that are different from the desired mood. w_0 and w_{k} are the weights of the moods accordingly. By default, we set all the moods' weights to 0.2. Note that the value of the terms w_0 through w_{k} are manually set.

Overall, a higher mood cost refers to the fact that the classifier is less confident in the scene belonging to the desired mood.

5.3 Realness Cost

The realness cost evaluates how realistic the colors of the objects' textures would be compared to their counterparts in the real world. We implemented the color comparison method based on the *CIEDE2000* [37] algorithm, since it provides a distance metric that accounts for human perceptibility of color differences. Our approach considers the colors of all objects in the scene during each iteration, with each object's cost weighted by its area relative to the total area of all objects. We define the realness cost as follows:

$$C_{\text{R}}(\Phi) = \sum_{\phi \in \Phi} w_{\phi} \cdot c_{\phi}, \tag{4}$$

where w_{ϕ} is the weight, and c_{ϕ} is the realness cost of the object ϕ in the objects set Φ of the scene. The weight w_{ϕ} of object ϕ is defined as:

$$w_{\phi} = \frac{\phi^{\text{a}} \cdot \phi^{\alpha}}{A}, \tag{5}$$

where ϕ^{a} is the surface area and ϕ^{α} is the importance of the object ϕ. A is the sum of all the objects' weighed areas by the importance. By default, we set ϕ^{α} to 1. Fig. 4 shows an example of the weights of different objects in the bedroom scene. The object's realness cost c_{ϕ} evaluates the realisticity of object ϕ.

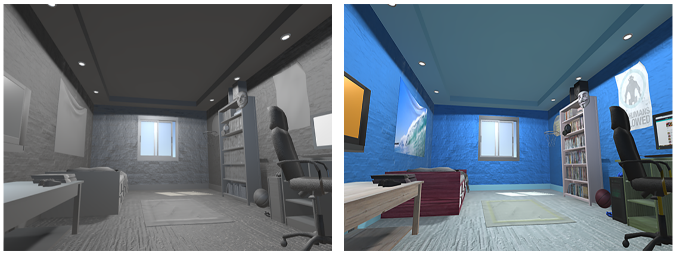
Our color comparison method differentiates each object in the scene with relevant ones in the *Open Surfaces Dataset* [5] including a set of common real-world objects. For example, if a lamp made of plastic is detected, the method compares the five dominant colors of its texture with those of all plastic lamps in the dataset. The Open Surfaces Dataset was created with the aim of segmenting surfaces from real life consumer photographs. The surfaces of the segmented objects are then catalogued by material, texture, and context. In our case, our main interest is the dominant color of these surfaces.

We apply the *K-means* clustering algorithm on the objects in the Open Surface Dataset, and the result stores the color data in RGB space; It helps us to find the five dominant colors for the textures for objects. The RGB color values are firstly converted to CIEL*C*h [48] values so that their distances are more highly related to human perception. This conversion also applies to the textures for objects in the scene. Then, a color distance metric is employed in CIEL*C*h space to evaluate the realisticity of textures, based on the *CIEDE2000* algorithm. We define c_{ϕ} as follows:

$$c_{\phi} = \min\{D(\phi^{\text{c}}, \psi^{\text{c}}) \mid \phi \in \Phi, \psi \in \Psi \text{ and } \phi^{\text{m}} = \psi^{\text{m}} \text{ and } \phi^{\text{t}} = \psi^{\text{t}}\}, \tag{6}$$

where $\Psi = \{\psi_{\text{j}} = (\psi_{\text{j}}^{\text{m}}, \psi_{\text{j}}^{\text{t}}, \psi_{\text{j}}^{\text{c}}) \mid j = 1, ..., q\}$ is a set of the objects' material pairs comprising the material $\psi_{\text{j}}^{\text{m}}$, the type $\psi_{\text{j}}^{\text{t}}$, and the five-dominant-colors $\psi_{\text{j}}^{\text{c}}$ of the texture of each object ψ_{j} in the Open Surface Dataset. $D(.)$ is a metric to compute the color distance between the five-dominant-colors of two objects in CIEL*C*h color space. For a detailed explanation of $D(.)$ please refer to our supplementary document.

K-means has also been applied successfully to differentiate colors. For example, Hassan et al. [18] has successfully shown its application to image segmentation in RGB and HSV color spaces. The CIEDE2000 algorithm has also been shown to statistically improve over past iterations



(a) Initial (b) Result (peaceful)

Fig. 5: Result from running our optimization with a realness cost weight of 0.2 and a mood cost weight of 0.8 for the bedroom scene. The mood, in this case, was peaceful. (a) The input scene. (b) The optimized result.

of algorithms to detect color differentiation as explained by Melgosa et al. [31]. Furthermore, the CIEDE2000 algorithm has shown to be applicable in many domains. For example, Yeerken et al. [46] demonstrated the effectiveness of the algorithm for evaluating masticatory function through color-changeable chewing gum.

5.4 Optimization Process

Our approach modifies the input scene by optimizing it with respect to the total cost function $C_{\text{total}}(\Phi)$. In particular, we implement the optimization with a MCMC technique (i.e., simulated annealing with a Metropolis-Hastings state searching step [11]). To effectively sample solutions from the search space, we define a Boltzmann-like objective function as follows:

$$f(\Phi) = \exp(-\frac{1}{t}C_{\text{Total}}(\Phi)), \quad (7)$$

where t is the temperature parameter of simulated annealing. The temperature t decreases over the iterations. At each iteration of the optimization, our approach applies a move to modify the current scene Φ to create a proposed scene $\Phi^{'}$. Each move randomly selects one or more objects, whose textures will be modified by changing the hue, saturation and lighting of the texture. During each iteration, we set the number of selected objects N to a random integer between 1 and 10. Our optimization is more likely to accept moves that modify more objects early in the process when the temperature t is high. Late in the optimization process when the temperature t is low, it is more likely to only accept moves that modify few objects.

To decide whether to accept the proposed scene $\Phi^{'}$, our approach compares the total cost $C_{\text{Total}}(\Phi^{'})$ of the proposed scene $\Phi^{'}$ with the total cost $C_{\text{Total}}(\Phi)$ of the previous scene Φ. To avoid the result becoming stuck at a local minimum, our approach accepts the proposed scene $\Phi^{'}$ with an acceptance probability based on the Metropolis criterion:

$$\alpha(\Phi^{'}|\Phi) = \min(1, \frac{f(\Phi^{'})}{f(\Phi)}). \quad (8)$$

We apply simulated annealing to efficiently explore the solution space, which is controlled by the temperature parameter t. At the beginning of the optimization, t is set to be high such that the optimizer aggressively explores the solution space to locate a good starting point. Throughout the optimization, the temperature t is lowered gradually, until it reaches a low level which is near zero. By default, we set $t = 1.0$ at the beginning of the optimization and decrease it by multiplying by 0.95 each iteration. The optimization is set to terminate if the change in the total cost is less than 3% over the previous 50 iterations.

Fig. 5 shows an example of the scene optimization process. The total cost decreases over the optimization process. Fig. 7 shows the cost over the optimization iterations. The supplemental video includes an animation of the optimization process.

6 Experiments

We tested the proposed approach on the scenes illustrated in Fig. 6. Sample images from the training dataset are shown in Fig. 9. All images

(a) Bedroom (b) Bathroom

(c) Dining Room (d) Living Room

Fig. 6: The four scenes used for our experiments.

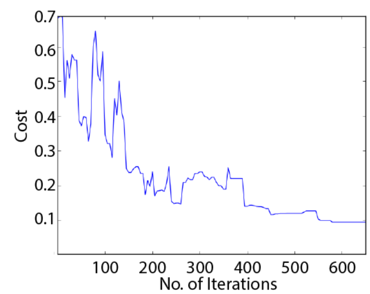


Fig. 7: The plot of our optimization process for the result of Fig. 5

shown comply with copyright rules. Credits to the owners given where required. Other images are free to use. Our supplementary document contains additional examples of training images for each mood.

6.1 Implementation

Our approach was tested using an Alienware R10 desktop with an AMD Ryzen 7 3700X 8-cores 3.6 GHz processor, 64 GB of RAM, and an NVIDIA GeForce GTX 3080 8GB graphics card.

Prior to each experiment, we configured an indoor virtual scene with necessary pre-processing as the input. Specifically, information of each object in the scene was required, including its material and type, for computing the realness cost in the optimization. For each visible object in the scene, we labeled both its material and type. For example, a wooden chair would have a label *wood* for its material and a label *chair* for its type. The surface areas of objects are calculated and their textures with updated colors are generated automatically in pre-processing. Many 3D scenes have label information about materials and types of objects, though manual labeling is required for scenes that do not have this information.

Designers can mark selected objects as *unmodifiable* before starting the optimization process. For example, a designer may fix part of their design and allow the other objects in the scene to be edited by our approach. If required, the designer can also change the weights of certain objects in regards to the realness cost.

It is also possible to color objects by group. For example, the chairs around a table could have the same color. We provide this optional constraint due to the observation that similar or the same color schemes are commonly found for grouped objects in indoor scenes.

After the pre-processing, the scene is to be optimized by our approach, while designers can still determine the contributions of both costs, balancing between its realness and distance from the desired mood. In our experiments, the automatic optimization process took on average about 15 minutes to converge, which improved the time

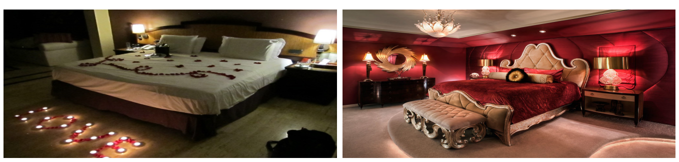
(a) Cheerful Dining Room (b) Scary Bedroom (c) Melancholy Living Room (d) Peaceful Bathroom (e) Romantic Living Room

Fig. 8: Different results from running our optimization process.

(a) Cheerful (right: D Sharon Pruitt)

(b) Melancholy (left: goatling; right: A.Curell)

(c) Peaceful (left: Betta Living; right: US Army)



(d) Romantic (left: daryl_mitchell, right: tec_estromberg)

(e) Scary (left: US Department of Defense)

Fig. 9: Sample room images for training the classifier for each mood.

efficiency compared to manual designs. For example, it took the artists in our user study at least an hour to manually color the rooms in each scene. Artists colored the objects in the scenes using Unity's coloring options as well as other programs like Photoshop.

6.2 Results

Fig. 8 shows the results from the optimization process with a mood weight w_{M} of 0.8 and a realness weight w_{R} of 0.2 (i.e., emphasizing the mood more than realism). For comparison, Fig. 10 shows the results using a mood weight of 0 and realness weight of 1.0 (i.e., considering realism only). Please refer to our supplementary document for additional results. The illustrated scenes were also used in our user study. For results of our optimization for only one scene, please refer to our supplementary document.

In our experiments, a set of CNN and SVM classifiers were trained for mood classification. A total of $25,000$ images, with $5,000$ indoor decoration images for five representative moods (cheerful, scary, melancholy, peaceful and romantic), were used in the training. For more

(a) Realistic Dining Room (b) Realistic Living Room

Fig. 10: Examples of the realistic results from our approach for two different scenes. These results were obtained using a realness weight of 1.0 and a mood weight of 0.

details, please refer to the supplementary document.

The indoor decoration images in our dataset were downloaded from websites (e.g., Flicker and Google) based on mood-related keywords. To finalize the dataset, we conducted a series of selection, analysis and manual labeling on these images, with the assistance of students and professors in arts.

7 User Studies

To test the effectiveness of our approach regarding different aspects, we conducted four separate user studies. Specifically, User Study 1 was accomplished before the ongoing social-distancing efforts due to the pandemic. User Study 2 was done through Amazon MTurk. User Study 3 and 4 were conducted after the pandemic outbreak; a slightly lower number of 13 participants were recruited for User Study 3 due to challenges posed by social distancing.

7.1 Design

User Study 1: Scoring and Classifying Synthesized Results. Synthesized results from our approach were tested in this user study with 40 participants, whose ages are in the range of 18-40 years (mean=26.65, $\pm$5.79). They were compensated for participating. We asked the participants to wear an *Oculus Rift* VR headset and view different scenes optimized by our approach. They completed the following three tasks based on 12 synthesized scenes (shown in our supplementary document).

Task 1: Rating. Participants rated five results generated by our approach using a five-point Likert scale. Ratings reflected how strongly participants believed our results matched the input mood. A score of five refers to *Strongly Agree*, while a score of one means *Strongly Disagree*. For consistency, we used the *Bedroom* scene for all moods throughout the task.

Task 2: Mood Identification. Using the other three scenes besides the bedroom scene, we presented five results and asked participants to select a mood from a list, which they believed to match the scene the best. The task aimed to confirm if participants were capable of matching the target mood in the scene optimization. The following *scene-mood* combinations were provided: Dining Room (Cheerful), Bedroom (Scary), Living Room (Melancholy), Bathroom (Peaceful), and Living Room (Romantic). The mood list included an extra *Others* option in addition to the mentioned moods,which applied to the case that none of five moods was regarded as a match.

Task 3: Realism. Participants were also asked to score 2 scenes based on how realistic they found the colorization, in a similar way

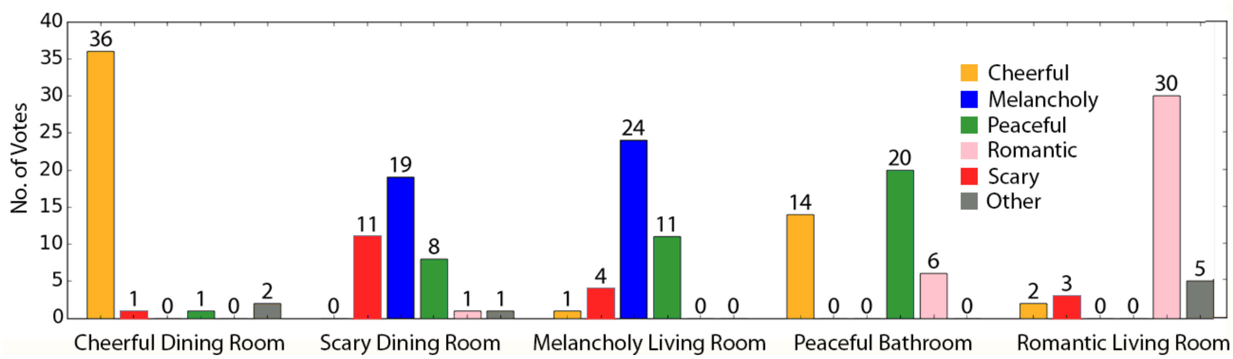


Fig. 11: Bar charts showing the number of votes cast during the user study for the perceived mood of different scenes.

to *Task 1*. Both scenes were synthesized by the optimization process with a realness weight of 1.0 and a mood weight of 0. In this task, the *Dining Room* and *Living Room* scenes were used, and can be seen in Fig. 10.

The scenes used in the above three tasks were viewed in random order. Each participant viewed 12 scenes. This user study started with an instruction page, briefing the participants about the entire process.

User Study 2: Comparing Synthesized Results to Artists' Designs. We conducted an alternative forced choice test. We recruited 35 Turkers on *Amazon* to fill out a form, which compared 20 synthesized results with 20 colored by two 3D artists. A screenshot can be found in our supplementary document.

In particular, both artists were invited to assign colors to the objects in the scenes used in User Study 1, generating a total of 20 scenes (i.e., 2 for each scene-mood combination used in User Study 1). Then, we asked the Turkers to choose one from the two compared scenes (i.e., one synthesized from our approach and one designed by an artist) that better delivered a corresponding mood in each question. The two scenes were displayed randomly regarding their sides (i.e., left or right) and not labeled their origins (i.e., our approach or artists). We compensated the artists and the Turkers for their participation.

User Study 3: Virtual Reality and Desktop Comparison. To test the usefulness of our approach for VR applications, we recruited some participants to view four scenes observing the scene results displayed by a desktop application versus the same ones displayed by a VR application with a headset. The participants were on average 24.77 ± 6.58 years old. 13 of them participated in this study despite challenges posed by social distancing in the current pandemic.

Specifically, the participants rated how well the scenes invoked the target moods, under the conditions of viewing the scenes with a VR application and a desktop application. The ratings were given in a five-point Likert scale, and the scenes were shown in a random order.

After completing this user study, the participants were invited to answer several questions about how they perceived the experience in the VR application compared to the desktop application.

User Study 4: Ablation Comparison. We evaluated and compared the results of the three color modes: realness only, optimal, and mood only. The weights were set to 1.0 (realness) and 0.0 (mood) for realness only; 0.3 and 0.7 for optimal; and 0.0 and 1.0 for mood only. We asked users to select the dominant moods of the result scenes synthesized using the three color modes, as well as to assess whether the color schemes of the result scenes were common in reality. We recruited 39 participants for this user study. They did not participate in other user studies. They were all students and the average age was 24.02 years.

7.2 Results and Analysis

User Study 1: Scoring and Classifying Synthesized Results. The mean score across all scenes from *Task 1* and *Task 3* scenes is 3.918 out of 5.000 (i.e., the participants were on average convinced that the result scenes matched the desired mood). The outlier is the Romantic Bedroom scene receiving a significantly lower score than others. The most common score voted for across all seven scenes was 4. In the next section we further analyze these results.

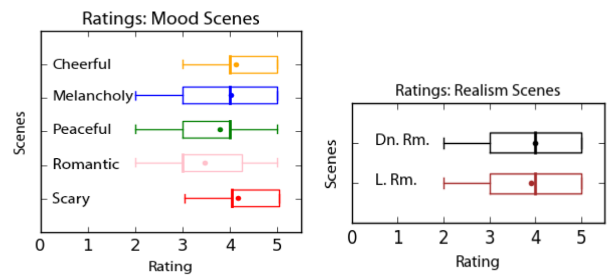


Fig. 12: Box plots showing the participant ratings in the user study. The rectangles denote quartiles, the thick lines denote medians, the dots denote means, and the thin lines denote margins of error. The left bar chart shows the results for *Task 1*, where participants had to rate how well each mood optimized for matched the virtual indoor scene. For consistency, the bedroom scene was used for each mood. The right bar chart shows the results for *Task 3*, where participants had to rate how realistic the results were for the scenes. Dn. Rm. and L. Rm. refer to the dining room and living room, respectively.

Scene	P-value
all scenes	0.020
only mood bedroom scenes	0.007
only positive mood bedroom scenes	0.007

Table 1: P-values of ANOVA tests on quantitative tasks (*Task 1* and *Task 3*) of User Study 1.

Results from *Task 2* are shown in Fig. 11. For four out of the five optimized scenes used in *Task 2*, users selected the mood that was optimized for in the majority.

Furthermore, Fig. 12 shows the box plots of the participant ratings for *Task 1* and *Task 3*. For more information on the results of User Study 1 please refer to our supplementary document.

Analysis. For data from *Task 1* and *3*, we conducted an ANOVA test for the involved scenes, and also between all broadly positive moods (i.e., Cheerful, Peaceful, and Romantic). Table 1 shows the test's result. The null hypothesis is that there is no difference in the ratings of the scenes. This is an important result, as it helps us differentiate between results that were convincing or not. For the broadly negative moods (i.e., Scary, Melancholy), we performed a pairwise t-test. The same was done with the two scenes that were rated based on their realism in *Task 3*.

Our supplementary document contains the results of all post-hoc pairwise t-tests. The reason for performing the analysis was to determine any significant differences between ratings for each scene. If any p-value is less than 0.05, the null hypothesis can be rejected, in which case the compared scenes receive the same average rating from participants. Most p-values for t-tests between two scenes are greater than 0.05, which means no significant difference exists. Since the average rating is 3.918 out of 5.000, it supports that overall participants found our synthesized results convincing.

Scene	Cheerful Dining Room	Melancholy Living Room	Peaceful Bathroom	Scary Bedroom
Rating (VR)	4.31	4.23	4.15	4.00
Rating (Desktop)	2.85	2.77	2.92	2.69
P-value	0.007	0.008	0.029	0.011

Table 2: The average ratings given to the four different scenes used in User Study 3. These ratings indicate how well the synthesized scenes invoke the intended moods under each condition. Ratings are based on a one to five Likert Scale. The scenes can be viewed in Fig. 8. P-values are for t-tests conducted for comparing ratings between the same scenes in the VR app and desktop app.

The t-tests with p-values lower than 0.05 include the following binary sets of scenes: Romantic Bedroom vs. Realistic Dining Room, Cheerful Bedroom vs. Peaceful Bedroom, Cheerful Bedroom vs. Romantic Bedroom, and Scary Bedroom vs. Romantic Bedroom. Most of the null hypothesis rejections were from comparisons with the Romantic Bedroom scene, receiving the lowest score and highlighting the scene as an outlier vs. the others with higher scores.

For the results from *Task 2*, all the p-values of the Chi-Squared test are below 0.05, referring to the rejection of the null hypothesis. In other words, that all categories in each scene would receive the same amount of votes. Fig. 11 shows that the mood that was optimized for received the most votes in each case, except for the Scary Dining room scene. In this scene, the mood Melancholy received the most votes (19). Scary received 11 votes in this case. This is possibly due to the fact that Melancholy and Scary share semantic meaning to a certain degree. In this case, scenes for Melancholy and Scary were optimized with mostly dark colors, so it is somehow reasonable to interpret them interchangeably.

For the Peaceful Bathroom scene, 20 votes were cast for Peaceful while 14 for cheerful and six for romantic; The other two negative moods received no votes. In the Melancholy Living Room scene, while Melancholy was voted for the most (24), Peaceful still received 11 votes. Assuming that Melancholy and Peaceful are based on some common semantic meaning, they can be generally described as relaxed rather than scary. The walls in the scene are dark purple/blue and can conduct low emotional arousal in viewers, according to an early study [41].

For the Romantic Living Room scene, while the majority of participants voted for the romantic mood (30), five votes were cast for *Others* and three for Scary. A possible reason is that many of the furniture objects were in dark colors based on the romantic images in our dataset; There were many dark scenes due to low lit scenes, which included objects with red and pink colors. Research has shown that both black and red relate to the mechanism of human's perception of attractiveness [34]. In the future, investigation can be done to find a more reliable way of deciding in which cases dark colors match a scary mood or a romantic one.

User Study 2: Comparing Synthesized Results to Artists' Designs. The distribution of results in the alternative forced choice experiment is as follows. The synthesized results obtained more votes than the artists' designs in seven out of 20 cases. In four cases, the difference in the number of votes was 1. In the other nine cases, the results from the artists received more votes. Please refer to Fig. 13 for example results from this study. Note that the artists colored the objects in the scene to match each mood. Our synthesized results were obtained with a mood weight of 0.8 and a realness weight of 0.2.

Analysis. In conducting this user study, we hoped that our synthesized results were at least up to par to those hand-crafted by 3D artists. Since our synthesized results received more votes or a difference of only one vote from artists' results in 11 out of 20 comparisons, we could conclude that our approach produces similarly convincing results in our study. For a breakdown of all scenes compared and the number of votes in each case, please refer to our supplementary document.

(a) Synthesis: Cheerful Dining Room (b) Artist's: Cheerful Dining Room

(c) Synthesis: Melancholy Living Room (d) Artist's: Melancholy Living Room

Fig. 13: Examples of scenes used in the MTurk user study. The left column shows our synthesized results. The right column shows artists' creations. (a–b) For the Cheerful Dining Room, participants voted for our result as superior with 19 vs. 16 votes. (c–d) For the Melancholy Living Room, the artist's rendition of the colors for the room received 1 more vote than our result, with 18 vs. 17 votes, respectively.

User Study 3: Virtual Reality and Desktop Comparison. Participants rated the scenes based on how well they invoked the intended moods under the VR app condition and the desktop app condition, using a one to five Likert scale. Table 2 shows the average ratings. Overall, participants rated the effect of experiencing the scenes in VR positively.

Participants rated the experience in VR overall with a rating of 4.17 and the desktop app with a rating of 2.81. We conducted a two-tailed paired t-test comparing the ratings of each scene in VR to its desktop app counterpart. Results for this study can be seen in Table 2. All p-values are below 0.05, indicating a significant difference in the ratings. This indicates a rejection of the null hypothesis that assumes there is no significant difference in the ratings given under the two conditions.

Analysis. As each scene's average rating is higher in the VR version and all p-values obtained from T-tests conducted are less than 0.05, we conclude that the higher rating given to the VR version is significant.

Overall, the participants commented that viewing the scenes in VR invokes the intended mood more strongly, probably due to the greater immersiveness delivered by VR compared to the desktop app. Since VR experiences are growing in popularity both in film and video games, our approach could practically be applied to facilitate designers in prototyping scenes to produce certain moods.

User Study 4: Ablation Comparison. In this user study, we set out to test two questions about our results: Q1) Which mood is most dominant in a scene? The options include the previously mentioned five moods as well as a No Preference option. Q2) Is the color distribution similar to the common color distribution of a scene of that type? The answer is either yes or no.

Table 3 shows the p-values calculated using the Chi-Squared Test for the votes relating to the questions. We include the vote distribution and a screenshot of the user study in the supplementary document.

For Q1, the null hypothesis was that voters would vote the same amount for each mood. For Q2, the null hypothesis was that users would vote the same amount for yes as for no in the binary questions, except for the real scenes, which were scenes where a mood was not optimized for. For these scenes, we assumed that the users would vote overwhelmingly for yes (we assumed 38 votes for yes vs. 1 vote for no, just to avoid divide by 0 errors).

For Q1, we found the p-value to be well below the threshold of 0.05, so our results indicate a difference among voting rates for each mood. For Q2, the results were more varied, meaning either in the cases of the

Scene	Q1 p-value	Q2 p-value
Bathroom Real	<0.01	<0.01
Bedroom Real	<0.01	0.04
Dining Room Real	<0.01	0.04
Living Room Real	<0.01	1.00
Bathroom Peaceful	<0.01	<0.01
Bedroom Scary	<0.01	0.04
Dining Room Cheerful	<0.01	<0.01
Living Room Melancholy	<0.01	<0.01
Living Room Romantic	<0.01	<0.01
Bathroom Peaceful Mood Only	<0.01	<0.01
Bedroom Scary Mood Only	<0.01	0.04
Dining Room Cheerful Mood Only	<0.01	0.02
Living Room Melancholy Mood Only	<0.01	<0.01
Living Room Romantic Mood Only	<0.01	0.8

Table 3: The p-values calculated from the data gathered from our user study 4. The p-values relate to the two questions that were asked.

real scenes that voters did overwhelmingly vote for yes or in all other cases that voters voted somewhat evenly between the NO option and yes option.

For the scenes which were optimized for moods and received a p-value less than 0.01, we found that in five out of six cases voters voted more for yes. In the other cases, the number of votes were more even.

7.3 User Comments

Participants were asked to comment on the results from our approach after accomplishing the tasks in User Study 1. The comments were overall positive, stating that the synthesized scenes enhanced the corresponding target mood. Some of them pointed out that specific minor objects (e.g., a small plate in the *Dining Room* scene) did not match the others regarding colors, while the scene was still convincing.

Most of the negative comments focused on the scene optimized for *Romantic* mood, which correlated with its lower rating scores compared to those of others. Participants were not convinced about the dark ceiling, and the wall in pink that looked unrealistic. Most comments were positive towards the color schemes produced from our approach, and that the results optimized with a realness cost of 1 looked realistic. Participants also expressed their preference for the experience with the VR headset due to the immersiveness.

8 Summary

In this paper, we present a novel approach to automatic color adjustment of the textures for objects in a virtual indoor scene so as to match a mood. Since depicting what a mood *looks like* is subjective, we aimed to ameliorate the problem via CNNs to extract features from a large set of interior room images matching a certain mood. We used simulated annealing in a Metropolis-Hasting algorithm to automatically re-configure colors and textures in an indoor environment so as to enable a closer match to the desired mood.

A statistical analysis of our user study is also provided to indicate that, in general, the participants found the results convincing. They correctly determined the target mood in four out of five scenes, and rated the results from our approach with a score of 3.918 out of 5.000 on average in User Study 1. From User Study 2, we find that our approach can color many scenes to carry moods like 3D artists do. From User Study 3, we find that viewing the synthesized scenes via virtual reality, compared to via a desktop screen, could invoke the target mood more strongly.

8.1 Limitations

Several aspects of our approach are considered worth further investigation. One is about the effective visual partition of an indoor environment based on the designer's focus. For example, the designer would like part of the room to match a mood, or they may expect the walls to have less impact on the mood. Our approach can be extended to address this point by associating the contribution of each object to the mood cost via a corresponding weight that will be manually configurable.

Limitation still exists due to the subjective nature of moods. While we aim to capture the relationships between mood and colors through the use of deep learning, it is still interesting to find a more objective method to measure the emotional responses of participants. While some attempts have been made with EEG devices [26], the problem still confuses people in areas like measuring indicators of positive moods [19]. We expect the data collected from participants with an EEG device to provide informative and useful results.

Another aspect lies in the placement of cameras in the scene. People use cameras to take pictures, within a virtual scene, as images fed to the classifier. In our work, angles were intuitively set when capturing most of the room's interior. Although we regard such a choice as sufficient for the proposed approach, it would be more convenient to place cameras in the scene systematically. For example, cameras could be placed at positions where participants would be looking from because of their nature. Research has been conducted about optimal camera positions in virtual scenes for people to carry out tasks [4], and to get the best visibility [32]. We believe that automation of camera positioning for the most natural view of the scene will advance the research.

8.2 Applications

Multiple areas can benefit from our approach. The most straightforward cases are VR game development and VR film production. If a game developer or a scene designer would like a virtual scene to invoke a specific mood, our approach could act as a tool for fast prototyping facilitating the process. Architects may employ our approach in their duties. For example, if their clients is seeking a certain mood involved in a construction project, they can brainstorm and visualize candidate solutions effortlessly with our approach before the implementation.

Our approach can also be applied to psychological studies. Certain studies involve analyses on human responses to various indoor scenes. For example, Mahmoud [30] investigated how architectural elements affected both human psychology and behavior. Al-Akkam [3] revealed the implementation of the design process for interior architecture considering psychological elements.

8.3 Future Work

Although participants expressed their overall positive opinions on our approach and synthesized results in the user study, we are still interested in the following future explorations that may improve the delivery of the target mood in a scene. For example, objects differ by shapes and thus it is meaningful to examine how the geometric aspect impacts on the perceived mood. This work can be accomplished by the generation of 3D objects to decorate a scene [42] and the delicate alteration of shapes and outlooks for objects, in order to check the influence on the scores output by classifiers. The materials and textures could also have an impact on the perceived mood of a scene, which offers another avenue of exploration.

The approach could also be extended to work with outdoor scenes. One challenge is the more uneven geometry that is found in outdoor scenes, as well as the impact of skyboxes. Depending on the size of the environment, a challenge would be granularity. The optimization could work with images that cover most of the outdoor environment. This may not be enough to cover the mood of the colors in more focused parts of the environment. A mixed approach that works at multiple levels of granularity could solve this issue.

Another approach that could be explored is the possibility of working with an end-to-end approach that can automatically output the appropriate color parameters for a scene. One way of accomplishing this could be through collecting data on many iterations of our approach with appropriate tags for objects. Then, based on average results of our approaches, we could automatically generate a set of parameter keys that could instantly color a scene based on the desired mood.

Acknowledgments

This research is supported by the National Science Foundation under award numbers 1942531 and 2128867.

References

[1] F. M. Adams and C. E. Osgood. A cross-cultural study of the affective meanings of color. *Journal of Cross-Cultural Psychology*, 4(2):135–156, 1973. doi: 10.1177/002202217300400201

[2] Y. Akazawa, Y. Okada, and K. Niijima. Automatic 3d scene generation based on contact constraints. *Intelligenza Artificiale - IA*, 8, 01 2005.

[3] A. Al-Akkam. Toward a psychological design process for interior architecture. *Journal of Architecture and Planning, King Saud University*, 25:21–38, 01 2013.

[4] W. H. Bares, L. S. Zettlemoyer, D. W. Rodriguez, and J. C. Lester. Task-sensitive cinematography interfaces for interactive 3d learning environments. In *Proceedings of the 3rd International Conference on Intelligent User Interfaces*, IUI '98, pp. 81–88. ACM, New York, NY, USA, 1998. doi: 10.1145/268389.268405

[5] S. Bell, P. Upchurch, N. Snavely, and K. Bala. OpenSurfaces: A richly annotated catalog of surface appearance. *ACM Trans. on Graphics (SIGGRAPH)*, 32(4), 2013.

[6] L. Brodschelm, F. Gräber, D. Hieber, and M. Hermann. Mood adaptive display coloring - utilizing modern machine learning techniques and intelligent coloring to influence the mood of pc users. 07 2021.

[7] S. H. Cha, S. Zhang, and T. W. Kim. Effects of interior color schemes on emotion, task performance, and heart rate in immersive virtual environments. *Journal of Interior Design*, 45(4):51–65, 2020. doi: 10.1111/joid.12179

[8] H. Chang, O. Fried, Y. Liu, S. DiVerdi, and A. Finkelstein. Palette-based photo recoloring. *ACM Trans. Graph.*, 34(4), July 2015. doi: 10.1145/2766978

[9] G. Chen, G. Li, Y. Nie, C. Xian, and A. Mao. Stylistic indoor colour design via bayesian network. *Computers & Graphics*, 60, 09 2016. doi: 10.1016/j.cag.2016.08.009

[10] K. Chen, K. Xu, Y. Yu, T.-Y. Wang, and S.-M. Hu. Magic decorator: Automatic material suggestion for indoor digital scenes. *ACM Trans. Graph.*, 34(6):232:1–232:11, Oct. 2015. doi: 10.1145/2816795.2818096

[11] S. Chib and E. Greenberg. Understanding the metropolis-hastings algorithm. *The american statistician*, 49(4):327–335, 1995.

[12] D. Cohen-Or, O. Sorkine, R. Gal, T. Leyvand, and Y.-Q. Xu. Color harmonization. *ACM Trans. Graph.*, 25(3):624–630, July 2006. doi: 10.1145/1141911.1141933

[13] A. J. Elliot. Color and psychological functioning: a review of theoretical and empirical work. *Frontiers in Psychology*, 6:368, 2015. doi: 10.3389/fpsyg.2015.00368

[14] A. J. Elliot and M. A. Maier. Color psychology: Effects of perceiving color on psychological functioning in humans. *Annual Review of Psychology*, 65(1):95–120, 2014. PMID: 23808916. doi: 10.1146/annurev-psych-010213-115035

[15] T. Feng, L.-F. Yu, S.-K. Yeung, K. Yin, and K. Zhou. Crowd-driven mid-scale layout design. *ACM Trans. Graph.*, 35(4):132:1–132:14, July 2016. doi: 10.1145/2897824.2925894

[16] Q. Fu, H. Yan, H. Fu, and X. Li. Interactive design and preview of colored snapshots of indoor scenes. In *Computer Graphics Forum*, vol. 39, pp. 543–552. Wiley Online Library, 2020.

[17] S. Han, A. Bhardwaj, and S. Choi. Automatic transfer of musical mood into virtual environments. *VRST18*, 2018.

[18] M. Hassan, R. Ema, and T. Islam. Color image segmentation using automated k-means clustering with rgb and hsv color spaces. *Global Journal of Computer Science and Technology: Graphics and Vision*, 17:33–41, 01 2017.

[19] X. Hu, J. Yu, M. Song, C. Yu, F. Wang, P. C. Sun, D. Wang, and D. Zhang. Eeg correlates of ten positive emotions. In *Front. Hum. Neurosci.*, 2017.

[20] C. Jiang, S. Qi, Y. Zhu, S. Huang, J. Lin, L.-F. Yu, D. Terzopoulos, and S.-C. Zhu. Configurable 3d scene synthesis and 2d image rendering with per-pixel ground truth using stochastic grammars. *International Journal of Computer Vision*, 126(9):920–941, Sep 2018. doi: 10.1007/s11263-018-1103-5

[21] D. Jonauskaite, A. Abu-Akel, N. Dael, D. Oberfeld, et al. Universal patterns in color-emotion associations are further shaped by linguistic and geographic proximity. *Psychological Science*, 31(10):1245–1260, 2020. PMID: 32900287. doi: 10.1177/0956797620948810

[22] S. Karayev, A. Hertzmann, H. Winnemoeller, A. Agarwala, and T. Darrell. Recognizing image style. *CoRR*, abs/1311.3715, 2013.

[23] S. K. Katukuri. *Viewpoint Recommendation for Aesthetic Photography*. PhD thesis, 2019.

[24] M. Koliska and K. S.-K. Oh. Guided by the grid: Raising attention with the rule of thirds. *Journalism Practice*, 0(0):1–20, 2021. doi: 10.1080/17512786.2021.1916402

[25] S. Kurt and K. K. Osueke. The effects of color on the moods of college students. *SAGE Open*, 4(1):2158244014525423, 2014. doi: 10.1177/2158244014525423

[26] Y.-Y. Lee and S. Hsieh. Classifying different emotional states by means of eeg-based functional connectivity patterns. *PLOS ONE*, 9(4):1–13, 04 2014. doi: 10.1371/journal.pone.0095415

[27] J. Lin, P. Xiao, Y. Fu, Y. Shi, H. Wang, S. Guo, Y. He, and T. Lee. C3 assignment: Camera cubemap color assignment for creative interior design. *IEEE Transactions on Visualization & Computer Graphics*, (01):1–1, dec 5555. doi: 10.1109/TVCG.2020.3041728

[28] T. Liu, Y. Wei, Y. Zhao, S. Liu, and S. Wei. Magic-wall: Visualizing room decoration. MM '17, p. 429–437. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3123266.3123398

[29] R. Ma, A. G. Patil, M. Fisher, M. Li, S. Pirk, B.-S. Hua, S.-K. Yeung, X. Tong, L. Guibas, and H. Zhang. Language-driven synthesis of 3d scenes from scene databases. *ACM Trans. Graph.*, 37(6):212:1–212:16, Dec. 2018. doi: 10.1145/3272127.3275035

[30] H.-T. Mahmoud. Interior architectural elements that affect human psychology and behavior. *The Academic Research Community publication*, 1:10, 09 2017. doi: 10.21625/archive.v1i1.112

[31] M. Melgosa, R. Huertas, and R. Berns. Relative significance of the terms in the ciede2000 and cie94 color-difference formulas. *Journal of the Optical Society of America. A, Optics, image science, and vision*, 21:2269–75, 01 2005. doi: 10.1364/JOSAA.21.002269

[32] S. K. Mohd Yusoff, A. Md Said, and I. Ismail. Optimal camera placement for 3d environment. In J. M. Zain, W. M. b. Wan Mohd, and E. El-Qawasmeh, eds., *Software Engineering and Computer Systems*, pp. 448–459. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[33] R. Paul, S. H. Hawkins, Y. Balagurunathan, M. B. Schabath, R. J. Gillies, L. O. Hall, and D. B. Goldgof. Deep feature transfer learning in combination with traditional features predicts survival among patients with lung adenocarcinoma. *Tomography (Ann Arbor, Mich.)*, 2(4):388—395, December 2016. doi: 10.18383/j.tom.2016.00211

[34] A. D. Pazda, A. J. Elliot, and T. Greitemeyer. Perceived sexual receptivity and fashionableness: Separate paths linking red and black to perceived attractiveness. *Color Research & Application*, 39(2):208–212, 2014. doi: 10.1002/col.21804

[35] M. Pelowski, P. S. Markey, J. O. Lauring, and H. Leder. Visualizing the impact of art: An update and comparison of current psychological models of art experience. *Frontiers in Human Neuroscience*, 10:160, 2016. doi: 10.3389/fnhum.2016.00160

[36] G. Pintore, C. Mura, F. Ganovelli, L. Fuentes-Perez, R. Pajarola, and E. Gobbetti. State-of-the-art in automatic 3d reconstruction of structured indoor environments. *Computer Graphics Forum*, 39(2):667–699, 2020. doi: 10.1111/cgf.14021

[37] G. Sharma, W. Wu, and E. N. Dalal. The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application*, 30(1):21–30, 2005. doi: 10.1002/col.20070

[38] M. Sra, P. Vijayaraghavan, O. O. Rudovic, P. Maes, and D. Roy. Deepspace: Mood-based image texture generation for virtual reality from music. *Conference on Computer Vision and Pattern Recognition Workshops*, 2017.

[39] F. Sultana, A. Sufian, and P. Dutta. Advancements in image classification using convolutional neural network. In *Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*, p. 122–129, 11 2018. doi: 10.1109/ICRCICN.2018.8718718

[40] H. Touvron, A. Vedaldi, M. Douze, and H. Jégou. Fixing the train-test resolution discrepancy: Fixefficientnet, 2020.

[41] P. Valdez and A. Mehrabian. Effects of color on emotions. *Journal of Experimental Psychology: General*, 123(4):394–409, 1994. doi: 10.1037/0096-3445.123.4.394

[42] Y. G. Vasin, M. P. Osipov, S. V. Muntyan, and E. A. Kustov. Procedural modeling and interactive 3d visualization of objects of the internal structure of buildings and facilities. *Pattern Recognition and Image Analysis*, 25(2):278–280, Apr 2015. doi: 10.1134/S105466181502025X

[43] K. Wang, M. Savva, A. X. Chang, and D. Ritchie. Deep convolutional priors for indoor scene synthesis. *ACM Trans. Graph.*, 37(4):70:1–70:14, July 2018. doi: 10.1145/3197517.3201362

[44] P. Wardono, H. Hibino, and S. Koyama. Effects of interior colors, lighting and decors on perceived sociability, emotion and behavior related to social

dining. *Procedia - Social and Behavioral Sciences*, 38:362–372, 12 2012.

[45] P. Wozniak, H. Afrisal, R. Esparza, and B. Kwolek. Scene recognition for indoor localization of mobile robots using deep cnn. In *ICCVG 2018*, 09 2018.

[46] Y. Yeerken, T. Otomaru, M. Said, N. Li, and H. Taniguchi. Applicability of the cielab and ciede2000 formulae for detection of colour changes in colour-changeable chewing gum for evaluating masticatory function. *J Clin Diagn Res*, 11:ZC119–ZC123, 04 2017. doi: 10.7860/JCDR/2017/23950.9754

[47] L.-F. Yu, S. K. Yeung, C.-K. Tang, D. Terzopoulos, T. F. Chan, and S. Osher. Make it home: automatic optimization of furniture arrangement. *ACM Trans. Graph.*, 30(4):86, 2011.

[48] X. Zhang, B. A. Wandell, et al. A spatial extension of cielab for digital color image reproduction. In *SID international symposium digest of technical papers*, vol. 27, pp. 731–734. Citeseer, 1996.

[49] N. Zhao, Y. Cao, and R. W. Lau. What characterizes personalities of graphic designs? *ACM Transactions on Graphics (Proc. of SIGGRAPH 2018)*, 37, 2018.

[50] N. Zhao, Q. Zheng, J. Liao, Y. Cao, H. Pfister, and R. W. H. Lau. Selective region-based photo color adjustment for graphic designs. *ACM Trans. Graph.*, 40(2), Apr. 2021. doi: 10.1145/3447647

[51] J. Zhu, Y. Guo, and H. Ma. A data-driven approach for furniture and indoor scene colorization. *IEEE transactions on visualization and computer graphics*, 24(9):2473–2486, 2017.