Magic Decorator: Automatic Material Suggestion for Indoor Digital Scenes

Kang Chen[1] Kun Xu[1*] Yizhou Yu[2] Tian-Yi Wang[1] Shi-Min Hu[1]
[1]TNList, Tsinghua University, Beijing [2]The University of Hong Kong

Figure 1: *Input scene (first column) and three material suggestions automatically generated by our system.*

Abstract

Assigning textures and materials within 3D scenes is a tedious and labor-intensive task. In this paper, we present *Magic Decorator*, a system that automatically generates material suggestions for 3D indoor scenes. To achieve this goal, we introduce *local material rules*, which describe typical material patterns for a small group of objects or parts, and *global aesthetic rules*, which account for the harmony among the entire set of colors in a specific scene. Both rules are obtained from collections of indoor scene images. We cast the problem of material suggestion as a combinatorial optimization considering both local material and global aesthetic rules. We have tested our system on various complex indoor scenes. A user study indicates that our system can automatically and efficiently produce a series of visually plausible material suggestions which are comparable to those produced by artists.

CR Categories: I.3.7 [Computing Methodologies]: Computer Graphics—Three-Dimensional Graphics and Realism; I.3.8 [Computing Methodologies]: Computer Graphics—Applications;

Keywords: Material suggestion, Indoor scene, Data-Driven content creation, Computer-aided aesthetic design

Links:

1 Introduction

Despite the growing number of 3D models available on the Internet, 3D modeling remains a difficult job that requires specialist skills. It is hard for ordinary users to master modern modeling software such as *Autodesk 3DS Max* and *Trimble SketchUp*. This is the main reason why the number of 3D digital models is far less than the number of 2D digital images. Helping average users to create 3D models and scenes is thus a topic of current interest in computer graphics [Xu et al. 2011; Chen et al. 2014].

*Corresponding author, xukun@tsinghua.edu.cn

Building correct geometric shapes is just the first step of 3D modeling. To complete designs, one also needs to specify material properties and textures for each 3D part in the model. For 3D models such as cars, this is manageable as there are not many alternative materials for each part (e.g. the surface of a car is typically made of shiny painted steel, windows and tyres are made of glass and rubber respectively). However, when it comes to 3D indoor scenes, the problem becomes much more complex as a scene may contain hundreds of semantic parts [Jain et al. 2012] and each part may have a variety of material choices. Specifying material properties for the entire scene is a tedious and labor-intensive task. Even worse, one also needs to consider the aesthetic appearance and visual consistency of the scene as a whole. This requires specialized aesthetic knowledge, which most ordinary users may lack. Thus, our goal is to develop a method that can perform automatic material suggestion for indoor digital scenes.

Although very desirable, performing automatic material suggestion is challenging for several reasons. Firstly, indoor scenes primarily contain man-made furniture and objects. There exist strong correlations between object categories and the type of materials they are made of. In addition, aesthetic rules exist regarding material choices for groups of functionally related objects as well as for entire scenes. Acquiring knowledge of such relationships and rules from real-world data is crucial to the success of material suggestion. The questions are how to acquire such knowledge from real data, and which data sources should be used. Secondly, it is nontrivial to represent the acquired knowledge in a generalizable form that can be applied to novel scenes to obtain high-quality material assignment results. Thirdly, interior design has many different styles, and there is not a unique correct solution to material selection for a digital scene. Different users may have different preferences. A further challenge is how to develop mechanisms for incorporating user-provided constraints.

In this work, we present a novel system that automatically generates material suggestions for every object part in a 3D digital scene. Since at present the number of available digital scenes with both high-quality geometry and material properties is very limited, our system acquires knowledge from annotated images of objects and scenes. For this purpose, we built an image database with thousands of indoor objects and interior design cases. Our system learns *local material rules* from this image database at both the part level (feasible materials for each part) and the scene level (potential combinations of materials in a scene). *Global aesthetic rules*, which account for the harmony among the entire set of colors in a scene, are also extracted from the image database.

Given a hierarchically organized indoor scene with a semantic tag for every object and part, our system can automatically generate a series of visually plausible material configurations for the scene. The whole process is guided by learned knowledge. Specifically, we cast the solution process as a combinatorial optimization problem, with multiple energy terms encoding local material and global aesthetic rules, which is solved by simulated annealing. As different users may have different preferences, our system can narrow down the range of plausible results by incorporating various forms of user-provided constraints. A user study indicates that our generated material configurations are comparable to those manually chosen by artists.

In summary, the contributions of this paper are as follows.

• We have built a database of annotated images of objects and indoor scenes for complex scene-level material suggestion, and extracted knowledge from these annotated real-world images instead of 3D digital models. Statistics collected from our image database have been used for constructing local material rules and global aesthetic rules.

• We introduce a new data-driven technique for scene-level material suggestion. The problem of material suggestion for a 3D scene is cast as a combinatorial optimization problem that searches for feasible material configurations under the guidance of the learned material distributions. This optimization problem is solved by simulated annealing.

• We develop mechanisms to incorporate various forms of user-provided constraints.

2 Related Work

Image Colorization Image colorization has been an important topic in graphics and vision. Image colorization methods can be split into three categories, interactive colorization techniques [Levin et al. 2004; An and Pellacini 2008; Xu et al. 2009], color transfer techniques [Welsh et al. 2002; Faridul et al. 2014], and data-driven colorization methods [Chia et al. 2011; Lin et al. 2013; Huang et al. 2014]. The data-driven methods are the ones most related to our work. Chia et al. [2011] gave a colorization method based on Internet images. Suitable reference images are found by filtering images from the Internet, and final colorized images are obtained by performing color transfer from such reference images. Lin et al. [2013] introduced a colorization method for 2D patterns using a probabilistic factor graph. Their model considers both color compatibility and the spatial arrangement of pattern segments, and is trained using example patterns. Colorized results are obtained with Markov Chain Monte Carlo model sampling. Their method can impressively colorize patterns. However, since only low-level features are considered, it would be hard for their model to deal with natural images or 3D scenes, which need high-level semantic information.

Appearance Assignment Chajdas et al. [2010] propose a method which assists users to assign surfaces textures to large scenes. User assigned textures are automatically propagated to other surfaces based on a similarity measure. Leifman and Tal [2012] propose a mesh colorization method. Users are required to scribble initial colors in certain regions, and these colors are propagated to other regions under the assumption that adjacent pixels having similar intensities should have similar colors. Nguyen et al. [2012] propose a technique for transferring material styles from source images to a target 3D scene. They first heuristically estimate material parameters from a single image and then find an optimal material assignment for the target 3D scene which minimizes a visual closeness cost. Yu et al. [2012] proposed a data-driven approach for automatic dress outfit synthesis. They train a Bayesian network on real outfit examples, and obtain suitable outfits by optimizing a cost function which considers both color compatibility and dress code suitability. Recently, Bell et al. [2013] presented *OpenSurfaces*, which is a large database of image segments, each of which covers a partial object surface with the same material. Each image segment is annotated with an object tag, a material type and reflectance parameters. It is used as the material database in our system.

However, little work has been conducted on automatic assignment of true material properties, including both reflectance models and texture, for complex indoor digital scenes. Jain et al. [2012] discuss *Material Memex*, which automatically assigns materials to parts in a 3D model. It trains a factor graph which accounts for contextual information between materials and shapes using a database of 3D models with predefined material properties. While their system facilitates material assignment, it requires a database with a moderate number of high-quality 3D models as training data. This makes it hard to adapt their system to cope with 3D indoor scenes—we can only find a very limited number of digital scenes with both high-quality geometry and material properties. To overcome this difficulty, our system collects training data from an annotated *image* database of indoor objects and scenes.

Color Compatibility Various methods [Cohen-Or et al. 2006; O'Donovan et al. 2011] have also been presented to assess the harmony or compatibility of multiple colors. Cohen-Or et al. [2006] improve color harmony in an image by making the colors fit a harmonious template. O'Donovan et al. [2011] further developed a data-driven model to predict the rating of a 5-color palette. In our work, we also take into account color harmony when computing a scene-level aesthetic result.

Data-Driven Content Creation With the availability of large datasets of images, 3D models and scenes (e.g. Trimble 3D Warehouse), data-driven content creation has attracted much interest in recent years. Different data-driven techniques have been proposed for various purposes, such as context based model search [Fisher and Hanrahan 2010], furniture arrangement [Yu et al. 2011], color theme enhancement [Wang et al. 2010], scene arrangement [Fisher et al. 2012], sketch based scene retrieval [Xu et al. 2013], and scene modeling [Chen et al. 2014]. Our work shares the same spirit of the above methods, but it is targeted at a different topic, which is complex scene-level material assignment.

3 Overview

Our system takes a 3D indoor scene without material properties as input. We assume that the scene is organized as a 3-level scene-object-part hierarchy, and all objects and parts therein have already been given text tags, such as 'chairs' and 'table legs'. Our goal in this paper is to automatically assign suitable, aesthetic material properties to all objects parts in the scene.

To achieve this goal, the core idea is to determine whether a specific material configuration for the input scene follows real-world knowledge and aesthetic standards. This is achieved using knowledge extracted from an image database of real-world objects and scenes. Inspired by recent data-driven techniques [Xu et al. 2013; Chen et al. 2014], our method builds a database of images (Sec. 4.1) and extracts reliable *local material rules* from the image database

(Sec. 4.2). A local material rule describes usual material patterns for a small group of objects or parts. Figure 2 illustrates such patterns. A night table is usually made of wood, a dining chair often comprises a wooden frame and fabric cushions, and a bed-frame, night table and dresser in the same bedroom typically share similar or the same material. In addition, since an aesthetically pleasing set of colors usually agree with some harmonious template [Cohen-Or et al. 2006; O'Donovan et al. 2011], we also introduce *global aesthetic rules* to account for harmony among the whole set of colors in a specific scene (Sec. 4.3).

The overall score of a specific material combination is determined by how well the local material rules and global aesthetic rules are satisfied (Sec. 5.1). The materials of all objects in the scene are chosen by minimizing the overall score using simulated annealing (Sec. 5.2).

4 Knowledge Building

We represent a scene with a scene-object-part 3-level hierarchy. As shown in Figure 3, the top-level has a single node, which represents the scene. The nodes at the second level represent objects in the scenes (e.g. tables, chairs), and the leaf nodes at the third level represent object parts (e.g. table top, table leg). Object parts are the basic units in our material assignment tasks.

4.1 Database

With the prevalence of digital cameras, we have witnessed an explosion of digital photos on the Internet in the past decades. In contrast, the growth of free 3D digital models has been relatively slow. Many techniques have been proposed to enrich the set of 3D models with the aid of various types of 2D images [Xu et al. 2011; Chen et al. 2013; Huang et al. 2015; Miao et al. 2015]. Inspired by these works, we propose to extract material information from a large collection of images of indoor objects and scenes, and take advantage of the extracted information to aid material and texture assignment during 3D modeling.

Figure 2: *Local material/color patterns. A night table is usually made of wood, a dining chair often comprises a wooden frame and fabric cushions, and a bed-frame, night table and dresser in the same bedroom typically share similar or the same material.*

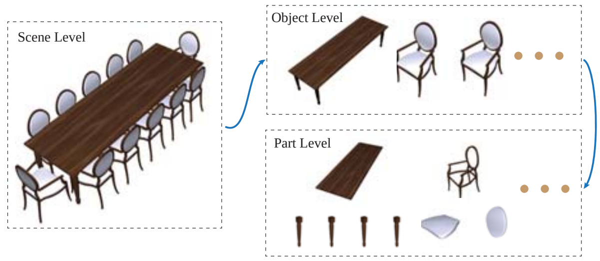

Figure 3: *3-Level hierarchy for a scene.*

Our Database We built our own image database by searching Internet images using the names of indoor objects (e.g. bed, night table, dining chair) and indoor scenes (e.g. bedroom, living room) as keywords. For each keyword, the top 800 returned images were retained. In total, we used 56 keywords related to indoor objects and 10 keywords related to indoor scenes. In order to obtain aesthetic images of indoor scenes, the keywords used during search include modifiers such as 'beautiful bedroom' and 'bedroom design'.

Next, we used the GIST scene descriptor [Oliva and Torralba 2001] to detect and remove duplicated images, and we also manually removed irrelevant images. At the end, we collected around 6,500 images. Within this collection, about 4,000 images depict specific indoor objects and about 2,500 images depict indoor scenes containing multiple objects. All the 2,500 scene images are used for extracting indoor color themes which we will later introduce in section 4.3. However, not all scene images are suitable for region labeling because object regions in many images are too small. After further selection, around 1,000 of them were chosen for region labeling.

Figure 4: *Examples of labeled representative regions in the image database.*

Representative Region Labeling After building the image database, we next needed to annotate associations between object categories and materials for each database image. This was done through manual labeling. We manually added such tags to all object parts (e.g. chair legs, table frame) existing in each image. Since we did not need the precise shape of each object part, to simplify the labeling process, we only needed to specify a rectangular or elliptical region, which shows the representative appearance of the material of that object part. Figure 4 shows examples of labeled images (two object images and one scene image) from our database. If multiple object parts in the same image exhibit the same material, we associate the same representative region with all object parts.

For example, in Figure 4, we assign the dining table, dining chair frame and bookshelf to the same box region on the table surface, indicating that all these objects in this image share the same material, whose representative appearance can be found in the labeled region.

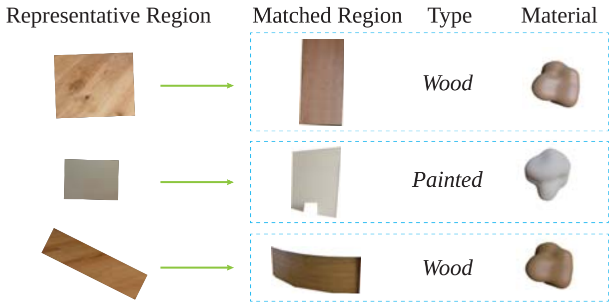

Figure 5: *Material annotation examples.*

Figure 6: *Examples of texture maps.*

Material Annotation Next, we need to annotate material information for each representative region. However, automatic material recovery from a single image is a long-standing open research problem in the graphics and vision communities. Fortunately, the recent work of *OpenSurfaces* [Bell et al. 2013] provides a large rich annotated image database for indoor scenes (objects). Specifically, for each labeled region, *OpenSurfaces* provides a text tag describing its object category, a text tag describing the material type and material parameters (i.e. diffuse, specular and roughness parameters). We can obtain material information for each representative region in our database by matching it to labeled regions in *OpenSurfaces*. To do so, we define an appearance distance between two regions as follows:

$$D(r_i, r_j) = s(r_i, r_j)(D_c(r_i, r_j) + \lambda_a D_t(r_i, r_j)), \qquad (1)$$

where r_i and r_j are two regions, one from our database, and the other from *OpenSurfaces*. $s(r_i, r_j)$ is set to one if the labeled object tags of the two regions are consistent, and is set to infinity otherwise. $D_c(r_i, r_j)$ accounts for color difference and is defined as the χ-squared distance [Asha et al. 2011] between the HSV color histograms of the two regions. In our experiments, the hue and saturation channels are quantified into 10 bins, while the value channel has 3 bins to reduce the influence of different illumination conditions. $D_t(r_i, r_j)$ accounts for the texture difference and is defined as the distance between the GLAM (Gray Level Aura Matrix) using the BGLAM (Basic Gray Level Aura Matrix) measure [Qin and Yang 2005]. In our experiment, GLAM is computed in a four-nearest-neighbor neighborhood system with 16 gray scale levels. D_c and D_t are both normalized to [0,1] by dividing by their respective maximal values. λ_a is a weight parameter to control the relative contributions of the two terms.

Our material annotation process has two passes. Specifically, for each representative region in our database, we first obtain the nearest 200 regions in *OpenSurfaces* using the appearance distance metric in Equation 1 with $\lambda_a = 0$. The purpose of the first pass is to quickly exclude regions with large color differences. In the second pass, we re-rank the obtained 200 regions using the appearance metric but with $\lambda_a = 1$ and retain the top 50 regions. To avoid noises, we further cluster the top 50 regions according to their material types. The region with the smallest appearance distance in the largest cluster is taken as the final matched region. Its material information is then transferred to the corresponding representative region in our database.

We have also evaluated other choices of feature combinations in defining the appearance distance in Equation 1. For this purpose, we randomly chose 500 representative regions in our database to form a testing dataset. For every choice of feature combination, we used the corresponding appearance distance to obtain the best matching region in *OpenSurfaces* for every region in the above testing dataset, and manually verified its correctness. In this way, we measured the precision of the above material annotation process over the testing dataset for different feature combinations, as shown in Table 1. Since it is hard to verify the correctness of material parameters, the precision is based on material types. From the results, we can see that our feature combination (HSV + BGLAMs + object tags) achieves a high precision. In addition, with the help of object tag information, the annotation precision is significantly improved. This again demonstrates the value of richly annotated databases.

Figure 5 gives an example of annotated results. Figure 5 (left) shows three regions labeled in Figure 4 (right), and Figure 5 (right) shows the matched shapes in *OpenSurfaces* along with the manual annotations of their material types and blobs rendered with the labeled parameters.

In addition to the above automatic annotation process, we manually annotate two material types, *mirror* and *glass*, and their associated parameters because they are highly specular and their reflection/transmission parameters cannot be reliably recovered from a single image.

Texture Maps Due to occlusion, incompatible illumination and projective transformation, cropped regions in both *OpenSurfaces* and our database are not ideal choices for texture maps. Meanwhile, texture maps are of vital importance for the appearance of natural materials such as wood and marble. In order to produce

Features	Precision
HSV	43.6%
BGLAMs	14.4%
HSV and BGLAMs	47.2%
HSV and object tags	85.4%
BGLAMs and object tags	26.8%
HSV, BGLAMs and object tags	93.2%

Table 1: *Material annotation precision for different features.*

better results, we further downloaded around 600 texture images using keywords such as ‘wood texture’, ‘marble texture’, and ‘fabric texture’. Each texture image is assigned a labeled region in *OpenSurfaces* if their appearance distance (Equation 1) is smaller than 0.2. In Equation 1, $s(r_i, r_j)$ is set to one if the texture and its assigned region have the same material type; and it is set to infinity otherwise. These texture images are used as texture maps for their corresponding materials in *OpenSurfaces* during scene rendering. Figure 6 shows a subset of these textures.

Database Alternatives There are various existing image databases for indoor scenes, such as *LabelMe* [Russell et al. 2008] and *OpenSurfaces* [Bell et al. 2013]. However, none of them meets the requirements of our application. The most widely used image database is *LabelMe* [Russell et al. 2008], but this database focuses on labeling object categories, rather than materials. Although there are works that extend the annotations in *LabelMe* to material names([Endres et al. 2010]), such annotations are not sufficient for our work. *OpenSurfaces* is a very good material database with material types, reflectance parameters, object labels, and surface exemplars. However, it provides little pairwise compatibility information for materials, i.e. scenes that contain multiple objects with different materials. Recently, Bell et al. [2015] augmented *OpenSurfaces* with newly gathered Internet images and built a larger and more diverse database. However, their database primarily aims at facilitating the classification of materials in real-world images, thereby, its annotation only includes material types. As object category and reflection parameters are also essential in 3D scene decoration, their database, again, fails to suit our needs. Although *OpenSurfaces* cannot be directly used as our database, it is useful for inferring material parameters from images. As noted, we relied on *OpenSurfaces* when building our own database.

4.2 Local Material Rules

We represent a material as follows, $m = (t, \mathbf{c}, \mathbf{s}, r)$, where t denotes the material type, $\mathbf{c}$, $\mathbf{s}$ and r denotes the diffuse, specular colors and roughness coefficient, respectively. This notation is the same as that in *OpenSurfaces* [Bell et al. 2013].

Unary Relationship We estimate the *unary likelihood*, f_1, of a specific object part category p having a specific material m as follows,

$$f_1(p, m) = c(p, m)/c(p), \tag{2}$$

where $c(p, m)$ denotes the number of object parts of category p with material m in the database, and $c(p)$ denotes the total number of object parts in category p with *any* material in the database.

Pairwise Relationship Despite the fact that there exist a variety of material choices for each individual indoor object, the combination of materials of multiple objects in a real-world indoor scene still follows certain compatibility rules. For instance, cabinets in a kitchen may have different sizes but they most likely share the same material; a night-table and a dresser may have very different geometric shapes but their materials are typically similar if they appear in the same bedroom. To account for such context dependent material properties, we define a *binary likelihood*, f_2, for a pairwise combination as follows,

$$f_2(p_1, m_1, p_2, m_2) = \sum_{m_1', m_2'} \frac{c(p_1, p_2, m_1', m_2')}{c(p_1, p_2)} \prod_{i=1,2} s(m_i, m_i'), \tag{3}$$

where p_1 and p_2 denote two object part categories, and m_1 and m_2 denote their corresponding materials. $c(\cdot)$ denotes the total number of times that the two object categories co-occur with specific/any materials in the database. The summation operator iterates over all pairs of possible materials. $s(m, m')$ denotes a proximity function, valued 1 if the two materials m and m' are similar, and valued 0 otherwise. We discretize the hue and saturation of a diffuse color into 12 bins and 5 bins respectively. m and m' are considered as similar if and only if they share the same material type and their corresponding hue and saturation values each fall into the same bins. We ignore the value channel of the diffuse color here because this channel is not defined on an absolute scale and is not directly comparable in *OpenSurfaces*. The roughness coefficient in *OpenSurfaces* is taken from 16 glossiness levels. In our experiments, it was found to be not very accurate when applied to 3D models. As the roughness is highly correlated with material types, we ignore this parameter when computing material similarities.

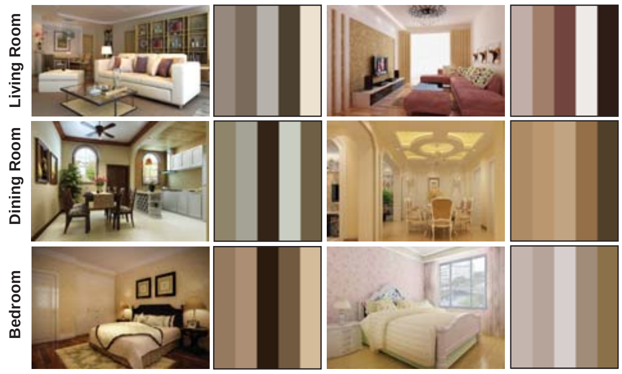

Figure 7: *Color themes extracted from images of living rooms, dining rooms and bedrooms.*

4.3 Global Aesthetic Rules

Local material compatibility rules make sure that the assigned materials are reasonable and consistent with the functionality of the objects. Nonetheless, the color combination of all objects in a scene as a whole is also important. The choice of color combination contributes significantly to the aesthetic, emotional, and stylistic impressions of what is being visualized [Csurka et al. 2011]. Different color distributions also embody different design styles. However, the impression given by a color combination is highly dependent on the perception of different subjects. Different people may have completely opposite preferences on the color combination of a room. Even two professional interior designers may not easily reach an agreement on the colors chosen for the furniture in a room. As we cannot make any assumptions about user preferences, our system must be capable of producing results in a variety of color themes. This means that until the user provides guidance, our global aesthetic rules should not be narrowed to any specific styles. To meet this requirement, we built our color compatibility model from thousands of images of interior designs.

Color Compatibility Model We adopt the color compatibility model proposed in [O’Donovan et al. 2011]. It represents a color theme with a five-color palette, and rates a color theme using a data-driven approach. Specifically, it is trained over a labeled dataset of five-color palettes, where each palette is already given a rating typically between 0 and 1. After being trained over the dataset, the model is able to predict a rating of new color themes through the learned LASSO [Tibshirani 1996] regressor.

(a) initial, $F(M) = 0.664$ (b) 1k iterations, $F(M) = 0.402$ (c) 10k iterations, $F(M) = 0.165$ (d) 100k iterations, $F(M) = 0.058$

(e) initial, $F(M) = 0.431$ (f) 1k iterations, $F(M) = 0.135$ (g) 10k iterations, $F(M) = 0.064$ (h) 100k iterations, $F(M) = 0.039$

Figure 8: *Material suggestion results with different numbers of iterations on two examples ((a)-(d) and (e)-(h))*

.

Training Dataset In the original work [O'Donovan et al. 2011], they primarily use a public color theme dataset, named COLOURLovers, which has 383,938 five-color palettes. This dataset is not tied to any specific context, hence, the model trained using this database predicts generic color compatibility. To customize it for our specific scenario of interior decoration, we train a new model using a modified dataset. Specifically, we collected around 2,500 images of interior designs (see section 4.1), and extracted a five-color palette from each of them using the method proposed in [O'Donovan et al. 2011]. Then, we arranged the five colors in each palette in a decreasing order according to their frequencies in the image. The complete set of extracted five-color palettes is denoted as the *reference dataset* T. Figure 7 illustrates some of the extracted five-color palettes. After that, we re-rate all color themes in the COLOURLovers dataset, and the new rating of a color theme in this database is given by

$$S'(t) = \delta S(t) + (1-\delta) e^{-\min\limits_{t' \in T} d(t,t')}, \tag{4}$$

where t denotes a five-color palette in COLOURLovers; t' denotes a five-color palette in the reference dataset T; $S(t)$ and $S'(t)$ represent the normalized original and new ratings of t, respectively; δ is a balancing weight, and we set $\delta = 0.6$ in all our experiments; $d(t,t')$ represents the squared distance between two palettes, and it is defined as

$$d(t,t') = \sum_{1 \le i \le 5, j \in \{h,s,v\}} w_j (t_{i,j} - t'_{i,j})^2, \tag{5}$$

where $t_{i,j}$ denotes the j-th color channel (i.e., hue. saturation and value) of the i-th color in the palette; w_j ($j \in \{h,s,v\}$) is a weight for each color channel, and we set $w_h = w_s = 1.0, w_v = 0.3$ to reduce the influence of illumination.

Our training dataset includes all color themes in both the COLOURLovers dataset and the reference dataset. All color themes in the COLOURLovers dataset have been re-rated using Equation 4, and all color themes in the reference dataset are assigned a rating of 1.

Scene Aesthetic Score Given a scene with material assignment, we extract its five-color palette as follows. We randomly sample a number of points (i.e. typically one million) on the surfaces of the scene, and the color of each point is set to the texture value (i.e. if having textures) or diffuse color (i.e. if not having textures) at that point. As we need to update the color theme of the scene during every iteration of our material labeling process, and the color theme extraction algorithm in [O'Donovan et al. 2011] is quite slow, we simply cluster all point colors using the method in [Lin et al. 2013] as our run-time solution for extracting the color theme. Specifically, representative colors from the five largest color groups are taken as the five-color palette. The aesthetic score of the scene is set to the predicted rating of its five-color palette using the color compatibility model described above.

5 Material Labeling

Given a semantically segmented 3D scene with object tags for each part, our material suggestion problem is formulated as a combinatorial optimization problem: find a combination of materials for all parts from a list of candidate materials, which best satisfies both the local material rules (Sec. 4.2) and the global aesthetic rules (Sec. 4.3). In the following, we first define the cost function (Sec. 5.1). Since the cost function is highly nonlinear and the number of unknowns is also large (a typical 3D indoor scene may contain hundreds of parts), the entire search space of our problem is too large to be enumerated. In order to quickly obtain a suitable solution, we use a classic heuristic search algorithm, simulated annealing [Kirkpatrick et al. 1983], as does in furniture arrangement [Yu et al. 2011] and building layout planning [Merrell et al. 2010](Sec. 5.2).

5.1 Cost Function

Given an input scene $P = \{p_i, 1 \le i \le n\}$, where p_i denotes the i-th object part in the scene and n denotes the total number of object parts, we denote a material configuration of P as $M = \{m_i, 1 \le i \le n\}$, where m_i denotes the material of the i-th object part. As shown below, our cost function for a material configuration consists of three terms: a unary term F_1, a binary term F_2, a global aesthetic term F_3:

$$F(M) = \lambda_1 F_1(P,M) + \lambda_2 F_2(P,M) + \lambda_3 F_3(P,M), \tag{6}$$

where λ_1, λ_2 and λ_3 are balancing weights for the three terms. In our experiments, we assign $\lambda_1 = \lambda_2 = \lambda_3 = 1/3$. We will evaluate the importance of each term in Section 6.

Unary Term The unary term considers how suitable the material of each part is, i.e. how well the unary relationship in Sec. 4.2 is sat-

Unary Rules

(a) $\lambda_1 = 0$ versus $\lambda_1 \neq 0$

Pairwise Rules

(b) $\lambda_2 = 0$ versus $\lambda_2 \neq 0$

Global Rules

(c) $\lambda_3 = 0$ versus $\lambda_3 \neq 0$

Figure 9: *Effects of the unary, binary and global aesthetic terms.*

isfied. It is defined according to the unary likelihoods (Equation 2) of all parts:

$$F_1(P, M) = -\frac{1}{n} \sum_{1 \leq i \leq n} \log(f_1(p_i, m_i) + \epsilon), \quad (7)$$

where $\epsilon = 1e - 5$ is included to avoid a zero value.

Binary Term Similarly, the binary term considers how compatible the material of one part is with other nearby parts. It is defined according to the binary likelihood (Equation 3) of all pairs:

$$F_2(P, M) = -\frac{2}{n \cdot (n-1)} \sum_{1 \leq i < j \leq n} \log(f_2(p_i, m_i, p_j, m_j) + \epsilon), \quad (8)$$

Global Aesthetic Term The global aesthetic term considers global color compatibility of a material configuration.

$$F_3(P, M) = D(P, M). \quad (9)$$

where $D(P, M)$ denotes the aesthetic score of the scene P decorated with material configuration M described in Section 4.3.

5.2 Simulated Annealing

The goal of simulated annealing is to find the best material configuration M that minimizes the cost function in Equation 6. In a typical simulated annealing process, the initial configuration M_0 is generated by randomly assigning a material to each object or part in the scene. Then in the k-th iteration, we find a new configuration M'_k in the neighbourhood of M_k. M'_k is accepted as M_{k+1} with probability

$$P_{M'_k \to M_{k+1}} = \min[1, \exp(-\frac{F(M'_k) - F(M_k)}{T_0 - \delta_T * k})], \quad (10)$$

where T_0 is the initial temperature, δ_T is the temperature drop at each iteration and k is the number of the current iteration. Note that Φ'_k can be accepted even if $F(M'_k) > F(M_k)$. As the temperature drops, the probability of accepting a worse solution becomes lower. In our system, we set $T_0 = -\log \epsilon$ which is an upper bound of $F(M)$ and $\delta_T = T_0/20n^2$.

Neighborhood The neighborhood of a configuration indicates the potential moves at each iteration. In each iteration, we randomly choose one part, and change its currently assigned material to another with a similar diffuse color. Specifically, the new material is randomly chosen from a set of N materials closest to the current material in terms of the hue value of the diffuse color. We set the neighborhood size $N = 20$ in our experiments. To avoid meaningless moves, we filter out materials whose unary score $f_1(p_i, m_i)$ is zero. Figure 8 gives two examples of the simulated annealing process. As the number of iterations increases, the cost function progressively decreases while the overall material suggestion becomes better.

6 Evaluation and Results

Effects of energy terms Figure 9 evaluates the effects of the unary, binary and global aesthetic terms in our cost function (Equation 6). As shown in Figure 9(a) left, without the unary term, the resulting material configuration of the living room includes metal cabinets, a leather rug and a marble chair, and the material configuration of the bedroom includes marble walls, a wooden quilt, metal night tables, dresser and bed frame, all of which are unlikely to appear in the real world. As shown in Figure 9(b) left, without the binary term, there exists obvious color inconsistency among the refrigerator, oven and microwave oven, and also among the bed frame, night table and dresser. Figure 9(b) right looks better because the binary term is included. As shown in Figure 9(c) left, without the global aesthetic term, the color combination of the entire scene no longer looks harmonious. Thus, the three terms are all important in obtaining a reasonable material configuration.

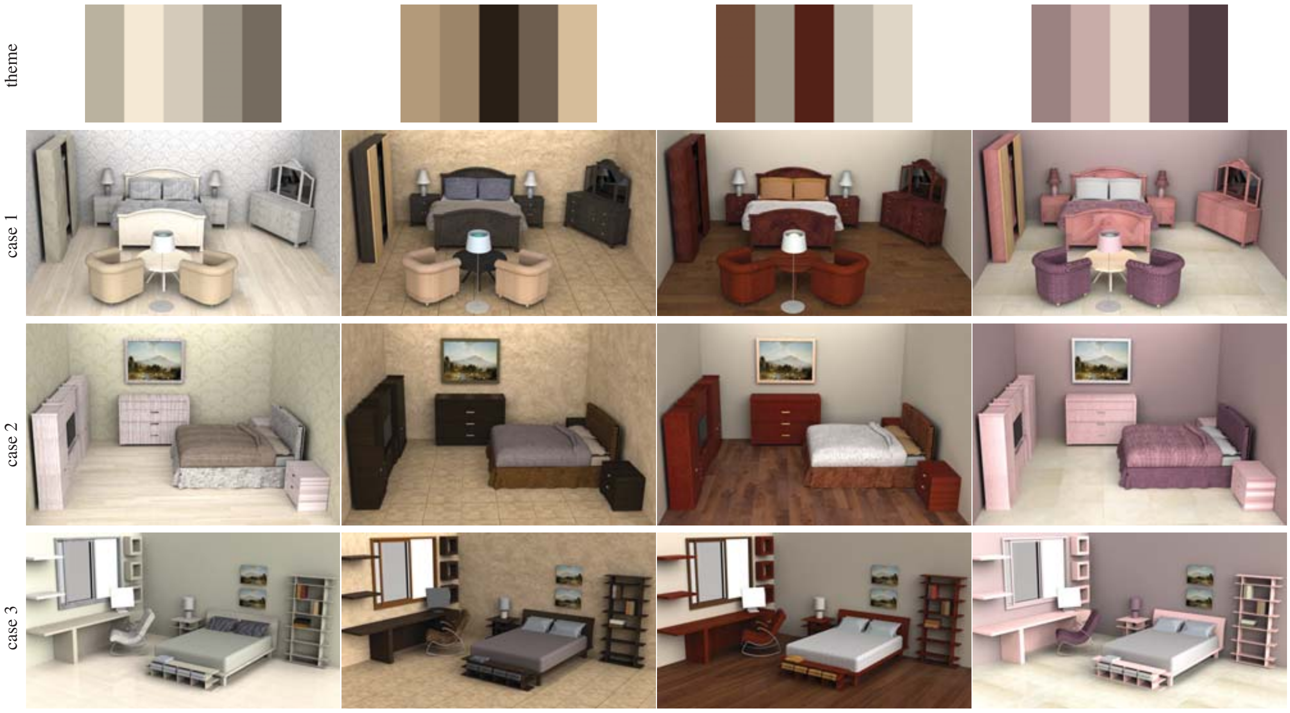

Figure 10: *Stylized decoration results. Color themes of user-provided styles are shown in the top row.*

Figure 11: *Constrained decoration results.*

Constrained Decoration Our system also allows users to specify constraints. A material constraint for one part can be specified as a value constraint by directly supplying desired material type, parameters, and diffuse color, or an image constraint by providing a reference region from an image. In the latter case, we obtain the material parameters of the reference region by applying the material annotation process described in Sec. 4.1. We enforce such constraints by modifying the unary score (Equation 2) in the cost function as follows,

$$f_1(p, m) = (t \equiv t_d) \cdot \exp(-\lambda_d(\mathbf{c} - \mathbf{c}_d)^2), \qquad (11)$$

where t_d and $\mathbf{c}_d$ are the type and diffuse color of the desired material specified by the user. λ_d is set to 1 in our experiments. The first term $t \equiv t_d$ ensures that the chosen material has the same tag as the desired material tag. Figure 11 shows two constrained decoration results, one with image constraints (left) and one with value (color) constraints (right). We can see that our system is capable of producing reasonable decoration results satisfying user-provided constraints.

Stylized Decoration To better reflect user preferences and produce customized results, our system can also generate decorations guided by a user-specified style. To be consistent with the color compatibility model we use, we encode a decoration style into a 5-color palette. Specifically, given a target 5-color palette t_g, we modify the global aesthetic score in Equation 9 as follows,

$$D(P, M) = e^{-d(t_g, t(P,M))}, \qquad (12)$$

where $t(P, M)$ denotes the extracted 5-color palette of the scene, $d(\cdot)$ is the squared distance between two palettes defined in Equation 5. In practice, the user-specified style can be provided in the form of a reference image or a previous decoration result, and our system can extract its corresponding color theme in both cases. Optionally, users can also directly provide five colors as a style, which is quite useful for experienced designers. Figure 10 shows sample results of stylized decoration.

Performance We implemented our system on a desktop computer with a Core i3 3.30GHz CPU and 8GB RAM. It takes about 8 hours for precomputation, including material annotation of representative regions in the database (Sec. 4.1), extracting unary and pairwise rules (Sec. 4.2), extracting color themes from images and training the final color compatibility model (Sec. 4.3). Note that the precomputation process is only needed once when our database is built. At run-time, for a typical scene containing about 150 parts, our system takes about 1-2 seconds to obtain a good decoration after the default number of $100,000$ iterations.

Experiments on Material Memex Our method is quite different and sometimes advantageous in comparison to *Material Memex* [Jain et al. 2012] in the context of material assignment for complete scenes. First, our method considers semantic information (i.e. object part categories) when defining contexts, while Material Memex primarily considers low-level geometric features; second, we include an extra global aesthetic term to ensure color compatibility within an entire scene; third, since we adopt *OpenSurfaces* as our material database, our assigned materials are more realistic, and have textures. Most importantly, our method only uses images as training data, which are much easier to obtain than a database of 3D scenes with high-quality materials, required by Material Memex. We have tested our method on a kitchen scene, which is also used by Material Memex. As shown in Figure 14, decoration results generated by our system are comparable to that produced by Material Memex. This test demonstrates the feasibility of our pure image-based approach.

Scene Decoration Results We have tested our system on a variety of indoor scenes, including bedrooms, living rooms, kitchens and a banquet hall. These scenes were assembled from a collection of 3D models, which were downloaded from Trimble 3D Warehouse. Most of these 3D models already included texture coordinates and had associated materials and textures at the time of downloading. Figure 15 shows a collection of results. The first column shows results manually chosen by artists, and the other columns are results automatically generated by our system using 100,000 iterations. The original materials and textures of all the scenes can be found in the supplemental document.

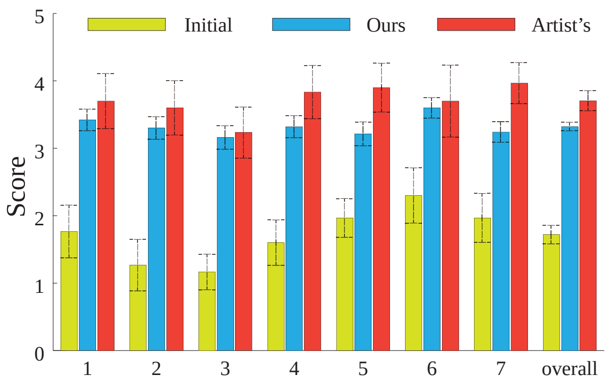

Figure 12: *Average user ratings of the initial decorations, decorations designed by an artist, and decorations automatically generated by our system. Confidence intervals are shown on top of average ratings.*

User Study We invited 30 users to participate in the following user study. For each scene, each participant was given 7 rendered images of the scene: one with the initial decoration of the scene, one with a decoration designed by an artist, and five different material assignment results generated by our system. The *initial decoration* of a 3D scene refers to the original material properties possessed by the 3D models in the scene when these models were downloaded. These 7 images were rendered under the same viewing and lighting conditions, and presented to the participants in a random order. The participants were asked to rate each rendered image between 0 (worst) and 5 (best), indicating the degree of visual quality. Figure 12 plots the average score along with the associated 95% confidence interval for each case. A t-test shows that our results outperform the initial scene decorations by more than 78% (one-tail p = 0.039) while results designed by the artist outperform ours by no more than 17% (one-tail p = 0.029), which demonstrates the effectiveness of our system. Furthermore, our system can produce results in a variety of styles, and it only takes 1-2 seconds per scene, which is much faster than a manual approach. It typically takes an artist 20-30 minutes to decorate a scene manually.

Figure 13: *At present, our system does not take spatial information into consideration, and thus is unable to make material suggestions for spatially related material patterns, e.g. the interleaved colors of the cabinet doors (left) and the black and white chairs (right).*

Limitations Since we use images as training data and robustly recovering 3D spatial relations from 2D images is still an unsolved problem, our cost function does not take spatial information into consideration. Figure 13 shows two examples of spatially related material patterns, the interleaving colors of the cabinet doors in the left image and the white and black chairs in the right. Currently, our system is unable to make material suggestions for such cases.

7 Conclusions

This paper has presented a novel system that automatically generates material suggestions for 3D scenes. We cast this problem as a combinatorial optimization that finds suitable material configurations under the guidance of both local material rules and global aesthetic rules. Both types of rules are learned from a database of annotated images of objects and complex indoor scenes for scene-level material suggestion. Our system is able to generate material suggestions efficiently with a quality most often comparable to those manually produced by artists, as demonstrated by a user study.

Our work still has much room for improvement. One promising direction is to incorporate geometric features into our material rules. Integrating our work with [Liu et al. 2015] to obtain product images with better material decoration would be another interesting possibility.

8 Acknowledgments

We thank the reviewers for their constructive comments. This work was supported by the National Basic Research Project of China (Project Number 2011CB302205), the Natural Science Foundation of China (Project Number 61120106007), Hong Kong Research Grants Council under General Research Funds (HKU17209714), Research Grant of Beijing Higher Institution Engineering Research Center, and Tsinghua University Initiative Scientific Research Program.

References

AN, X., AND PELLACINI, F. 2008. Appprop: all-pairs appearance-space edit propagation. *ACM Trans. Graph. 27*, 3, 40:1–40:9.

ASHA, V., BHAJANTRI, N. U., AND NAGABHUSHAN, P. 2011. Glcm-based chi-square histogram distance for automatic detection of defects on patterned textures. *Int. J. Comput. Vision Robot. 2*, 4 (Feb.), 302–313.

BELL, S., UPCHURCH, P., SNAVELY, N., AND BALA, K. 2013. Opensurfaces: A richly annotated catalog of surface appearance. *ACM Trans. Graph. 32*, 4 (July), 111:1–111:17.

BELL, S., UPCHURCH, P., SNAVELY, N., AND BALA, K. 2015. Material recognition in the wild with the materials in context database. *Computer Vision and Pattern Recognition (CVPR)*.

CHAJDAS, M. G., LEFEBVRE, S., AND STAMMINGER, M. 2010. Assisted texture assignment. In *Proceedings of the 2010 ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, ACM, New York, NY, USA, I3D '10, 173–179.

CHEN, T., ZHU, Z., SHAMIR, A., HU, S.-M., AND COHEN-OR, D. 2013. 3-sweep: Extracting editable objects from a single photo. *ACM Trans. Graph. 32*, 6 (Nov.), 195:1–195:10.

CHEN, K., LAI, Y.-K., WU, Y.-X., MARTIN, R., AND HU, S.-M. 2014. Automatic semantic modeling of indoor scenes from low-quality rgb-d data using contextual information. *ACM Trans. Graph. 33*, 6 (Nov.), 208:1–208:12.

CHIA, A. Y.-S., ZHUO, S., GUPTA, R. K., TAI, Y.-W., CHO, S.-Y., TAN, P., AND LIN, S. 2011. Semantic colorization with internet images. *ACM Trans. Graph. 30*, 6 (Dec.), 156:1–156:8.

COHEN-OR, D., SORKINE, O., GAL, R., LEYVAND, T., AND XU, Y.-Q. 2006. Color harmonization. *ACM Trans. Graph. 25*, 3 (July), 624–630.

CSURKA, G., SKAFF, S., MARCHESOTTI, L., AND SAUNDERS, C. 2011. Building look & feel concept models from color combinations. *The Visual Computer 27*, 12, 1039–1053.

ENDRES, I., FARHADI, A., HOIEM, D., AND FORSYTH, D. 2010. The benefits and challenges of collecting richer object annotations. In *CVPR. Workshops, 2010 IEEE Computer Society Conference on*, 1–8.

FARIDUL, H. S., POULI, T., CHAMARET, C., STAUDER, J., TRÉMEAU, A., REINHARD, E., ET AL. 2014. A survey of color mapping and its applications. In *Eurographics 2014-State of the Art Reports*, The Eurographics Association, 43–67.

FISHER, M., AND HANRAHAN, P. 2010. Context-based search for 3d models. *ACM Trans. Graph. 29*, 6 (Dec.), 182:1–182:10.

FISHER, M., RITCHIE, D., SAVVA, M., FUNKHOUSER, T., AND HANRAHAN, P. 2012. Example-based synthesis of 3d object arrangements. *ACM Trans. Graph. 31*, 6 (Nov.), 135:1–135:11.

HUANG, H.-Z., ZHANG, S.-H., MARTIN, R. R., AND HU, S.-M. 2014. Learning natural colors for image recoloring. *Computer Graphics Forum 33*, 7, 299–308.

HUANG, Q., WANG, H., AND KOLTUN, V. 2015. Single-view reconstruction via joint analysis of image and shape collections. *ACM Trans. Graph. 34*, 4 (July), 87:1–87:10.

JAIN, A., THORMÄHLEN, T., RITSCHEL, T., AND SEIDEL, H.-P. 2012. Material memex: Automatic material suggestions for 3d objects. *ACM Trans. Graph. 31*, 6 (Nov.), 143:1–143:8.

KIRKPATRICK, S., GELATT, C. D., AND VECCHI, M. P. 1983. Optimization by simulated annealing. *SCIENCE 220*, 4598, 671–680.

LEIFMAN, G., AND TAL, A. 2012. Mesh colorization. *Computer Graphics Forum 31*, 2pt2, 421–430.

LEVIN, A., LISCHINSKI, D., AND WEISS, Y. 2004. Colorization using optimization. *ACM Trans. Graph. 23*, 3, 689–694.

LIN, S., RITCHIE, D., FISHER, M., AND HANRAHAN, P. 2013. Probabilistic color-by-numbers: Suggesting pattern colorizations using factor graphs. *ACM Trans. Graph. 32*, 4 (July), 37:1–37:12.

LIU, T., MCCANN, J., LI, W., AND FUNKHOUSER, T. 2015. Composition-aware scene optimization for product images. *Computer Graphics Forum 34*, 2, 13–24.

MERRELL, P., SCHKUFZA, E., AND KOLTUN, V. 2010. Computer-generated residential building layouts. *ACM Trans. Graph. 29*, 6 (Dec.), 181:1–181:12.

MIAO, Y., HU, F., ZHANG, X., CHEN, J., AND PAJAROLA, R. 2015. Symmsketch: Creating symmetric 3D free-form shapes from 2D sketches. *Computational Visual Media 1*, 1, 3–16.

NGUYEN, C. H., RITSCHEL, T., MYSZKOWSKI, K., EISEMANN, E., AND SEIDEL, H.-P. 2012. 3D Material Style Transfer. *Computer Graphics Forum (Proc. EUROGRAPHICS 2012) 2*, 31.

O'DONOVAN, P., AGARWALA, A., AND HERTZMANN, A. 2011. Color compatibility from large datasets. *ACM Trans. Graph. 30*, 4 (July), 63:1–63:12.

OLIVA, A., AND TORRALBA, A. 2001. Modeling the shape of the scene: A holistic representation of the spatial envelope. *IJCV 42*, 145–175.

QIN, X., AND YANG, Y.-H. 2005. Basic gray level aura matrices: theory and its application to texture synthesis. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, vol. 1, 128–135 Vol. 1.

RUSSELL, B., TORRALBA, A., MURPHY, K., AND FREEMAN, W. 2008. Labelme: A database and web-based tool for image annotation. *IJCV. 77*, 1-3, 157–173.

TIBSHIRANI, R. 1996. Regression Shrinkage and Selection Via the Lasso. *J. Royal. Statist. Soc B. 58*, 1, 267–288.

WANG, B., YU, Y., WONG, T.-T., CHEN, C., AND XU, Y.-Q. 2010. Data-driven image color theme enhancement. *ACM Trans. Graph. 29*, 6 (Dec.), 146:1–146:10.

WELSH, T., ASHIKHMIN, M., AND MUELLER, K. 2002. Transferring color to greyscale images. *ACM Trans. Graph. 21*, 3, 277–280.

XU, K., LI, Y., JU, T., HU, S.-M., AND LIU, T.-Q. 2009. Efficient affinity-based edit propagation using k-d tree. *ACM Trans. Graph. 28*, 5, 118:1–118:6.

XU, K., ZHENG, H., ZHANG, H., COHEN-OR, D., LIU, L., AND XIONG, Y. 2011. Photo-inspired model-driven 3d object modeling. *ACM Trans. Graph. 30*, 4, 80:1–10.

XU, K., CHEN, K., FU, H., SUN, W.-L., AND HU, S.-M. 2013. Sketch2Scene: Sketch-based co-retrieval and co-placement of 3D models. *ACM Trans. Graph. 32*, 4, 123:1–123:15.

YU, L.-F., YEUNG, S.-K., TANG, C.-K., TERZOPOULOS, D., CHAN, T. F., AND OSHER, S. J. 2011. Make it home: automatic optimization of furniture arrangement. *ACM Trans. Graph. 30*, 4, 86:1–86:12.

YU, L.-F., YEUNG, S. K., TERZOPOULOS, D., AND CHAN, T. F. 2012. Dressup!: Outfit synthesis through automatic optimization. *ACM Trans. Graph. 31*, 6, 134:1–134:14.

(a) Material Memex (b) Our result 1 (c) Our result 2

Figure 14: *Experiments on a kitchen scene from Material Memex [Jain et al. 2012].*

Figure 15: *Gallery of material suggestion results. Scene 7 used in the user study is shown in figure 1.*