EUROGRAPHICS 2012 / P. Cignoni, T. Ertl
(Guest Editors)

Volume 31 (2012), Number 2

3D Material Style Transfer

Chuong H. Nguyen[1] Tobias Ritschel[2] Karol Myszkowski[1] Elmar Eisemann[2] Hans-Peter Seidel[1]

[1]MPI Informatik [2]Télécom ParisTech

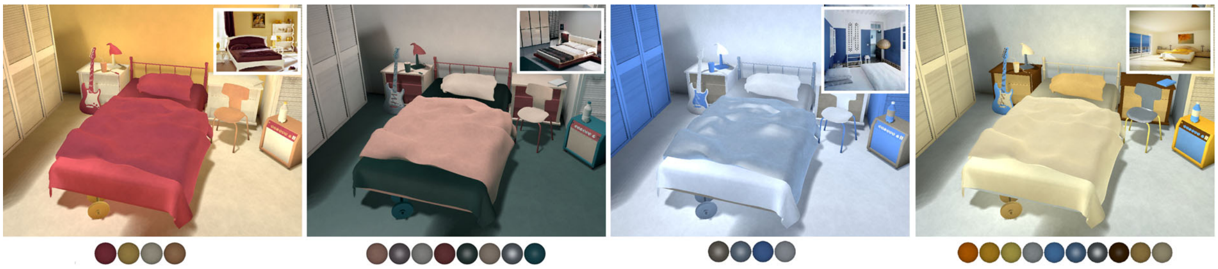

Figure 1: *Automatic 3D material style transfer from different source images (insets) to a target 3D scene using our approach.*

Abstract

This work proposes a technique to transfer the material style or mood from a guide source such as an image or video onto a target 3D scene. It formulates the problem as a combinatorial optimization of assigning discrete materials extracted from the guide source to discrete objects in the target 3D scene. The assignment is optimized to fulfill multiple goals: overall image mood based on several image statistics; spatial material organization and grouping as well as geometric similarity between objects that were assigned to similar materials. To be able to use common uncalibrated images and videos with unknown geometry and lighting as guides, a material estimation derives perceptually plausible reflectance, specularity, glossiness, and texture. Finally, results produced by our method are compared to manual material assignments in a perceptual study.

1. Introduction

Not all 3D scenes come with assigned materials (i. e., reflectance properties); a 3D scanner might not deliver colors or a model from the internet was simply crafted without. Images rendered with such scenes do not appear realistic, as users expect certain materials for certain objects.

When creating thumbnails to browse large databases materials are often a must. When materials are not available, a manual assignment is tedious. Especially, when a high number of objects and materials are involved, selecting, grouping, and navigating is challenging. Further, deciding on an object's material is not obvious, for example when colored light, e. g., due to indirect illumination is involved. The resulting appearance depends on the spatial context; a white object near a colored wall will exhibit color bleeding and not appear white anymore. In other situations, non-obvious rules, that might even be unknown to the user, should be applied, e.g., for architectural models, a common principle is to assign darker materials to the floor than the walls to convey a feeling of higher ceilings. In particular, when a number of design variants should be examined (Fig. 1) from different viewpoints, transferring mood from examples can serve as an inspiration. Similarly, organizing material assignments according to a style, might be beyond the user's capabilities, but can be important, as humans make certain color assumptions depending on the context (i. e., the Stroop [Str35] effect for chroma). Automatically accounting for such expectations makes scenes easier to understand and more pleasant.

Addressing these issues, our work proposes an automated system that extracts materials from a guide source, such as an image or video and assigns them to a target 3D scene. The material extraction approximately captures appearance from uncalibrated images of unknown geometry and lighting using image-filtering heuristics. The material assignment is formalized as an optimization problem that assigns discrete materials to discrete objects in order to minimize a cost function that evaluates the perceptual difference to a guide source. The cost function accounts for image differences as well as geometric matching between objects in the target 3D scene and the guide source. Furthermore, to a certain extent, the cost function grasps inter object relations including global illumination aspects. The user can therefore, by providing a simple guide source, attribute plausible materials to an entire 3D scene fully automatically.

2. Related Work

High-level style transfer within a single or amongst different media has been a longstanding challenge in computer graphics.

Materials Materials are reflectance properties shared by several surfaces that can be acquired from images [TT00] and clustered automatically [LKG*03], or via user interaction [PL07]. Nonetheless, dealing with appearance in a perceptually meaningful way stays challenging [PFG00, MPBM03].

One can produce plausible results for material modifications [KRFB06] in a single image, but, in contrast, our approach needs to map an image to materials of a 3D scene that can be explored afterwards (including light and view changes or animation). To this extent, we extract a Phong model [Pho75]. It cannot reproduce all forms of appearance, but many important effects.

Material and mood perception We are concerned with material style perception (i. e., its mood), and how style can be extracted and transferred. Humans perform remarkably when quickly categorizing natural images into a classes (its *gist* [OT01]). Possible cues include spatial organization of textures [OT01, WM02], or color [CH10], but also the organization of shapes, colors, spatial proximity [Ber48], congruence [Str35] and grouping (e. g., the scene's Gestalt) can influence the scene's mood. Human observers are extremely efficient in material categorization, which is a rapid effortless process; robust performance was reported even for a 160 ms image display [Lav08], and only slightly more for grey-leveled, blurred, or inverted-contrast stimuli.

Colorization Transferring material style includes a two-fold abstraction of the classic color transfer problem [RAGS01, ICOL05, LWCO*07, LWQ*08, WYW*10]. First, instead of transferring colors, we transfer materials, that change their impact on the final result depending on their spatial context, and include a much higher number of parameters (such as glossiness and textures). Second, a guide source image does only provide approximate material information, but the target 3D scene has to provide a similar gist even when it is seen from several different views. We take advantage of 3D scene geometry to make our assignment more reliable than it could be with images only.

Detail synthesis Another important part of style are surface details often in form of texture. Texture synthesis [HB95] takes a statistical approach and produces new instances from a training example or manual statistical settings. The Image-Analogies framework of Hertzmann et al. [HJO*01] can produce new exemplars with details that fit specified content. In the context of 3D models Mertens et al. [MKCD07] generate reflectance details by learning the geometric correlation from another 3D model with reflectance. In a similar fashion, Chajdas et al. [CLS10] consider local geometric structure to assist the user in assigning textures. In contrast to our work, their transfer is from 3D to 3D and does not consider the resulting perceived appearance, but solely statistical physical qualities. Extraction of a single texture to be then assigned to a target 3D object is considered by Lagae et al. [LVLD10], and we will extend their approach when extracting multiple textures for multiple materials from an image. The CG2Real system [JDA*10] uses large image collections to decorate a synthetic image with details. In contrast to our work, their results are image compositions and cannot be used easily used for 3D scenes (the output of our system).

3. Our Approach

Our system consists of two key components: *material extraction* (Sec. 3.2) and *material assignment* (Sec. 3.3). From a guide source, which can be an image or video, a set of materials is extracted. Then the system finds the best assignment of these materials to a target 3D scene (Fig. 2).

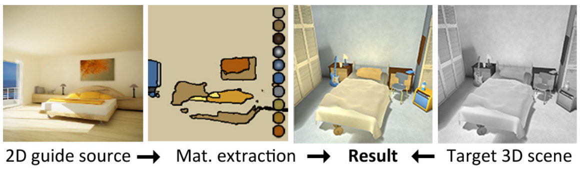

Figure 2: *Our approach extracts materials from a 2D guide source and assigns it to a resulting target 3D scene.*

3.1. Definitions

Here, we introduce the basic definitions that our system builds upon: materials, objects, material assignments and rendering.

We represent materials by (slightly modified) Phong coefficients [Pho75]: diffuse color, monochromatic specular intensity and glossiness. Textures are represented using anisotropic noise with a particular frequency spectrum [Per85] modulating the diffuse color. We decided to not optimize for the number of materials, but assume it to be n_m, a user-controlled parameter.

A scene is hand-crafted and therefore usually segmented into meaningful entities, or it needs to be segmented into objects that will receive the extracted materials.We assume a segmentation into n_o objects a given.

A material assignment $A := \{a_i \in \mathbb{N}^+ | a_i < n_\mathrm{m}, i < n_\mathrm{o}\} \in \mathcal{A}$ is a mapping from every object i to a material a_i. It is neither assumed that every material is used, nor do we enforce that it has to be used more than a certain number of times (or any other similar restriction).

We assume a set $\mathcal{V}$ of views on the scene which we will simply call "the views" for which the assigned materials should be optimal. It can either be a set with only one element defining a single camera position and orientation, a camera path, or a volume in space describing the potential viewpoints. When the views fill the complete space, no view-specific optimizations will be made.

Rendering in our context is an operator $r(A,V) := \mathcal{A} \times \mathcal{V} \to (\mathbb{R}^2 \to \mathbb{R}^3)$ that converts our fixed target 3D scene under some material assignment A and some views V into a two-dimensional RGB image.

3.2. Material Extraction

Extracting physical materials from images is a very ill-posed problem that we try to avoid. Instead, we extract perceptually plausible materials by splitting the image into several components that map to different shading parameters: Diffuse color, specularity and texture (Fig. 3).

First, the input image $\hat{L}$ is (inverse-gamma) converted into linear units and white-balanced to L. Then, L is split into a diffuse L_d and a specular scalar radiance L_s. For fast performance and simplicity, the method of Yang [YWA10] is used, but others are possible. Next, the diffuse radiance L_d is split into a base L_b and a texture part L_t. Bilateral filtering [DD02] is used to decompose L_b in L_d and $L_\mathrm{t} = L_\mathrm{d} - L_\mathrm{b}$.

Accounting for a material's specularity is challenging: highlights are only visible in some parts (white dots in Fig. 4) of an object although the specularity is the same everywhere on the object. We make the assumption that a continuous diffuse base color L_b implies the continuity of material (material segmentation). Conceptually, we can hereby associate a highlight to all places where it could have appeared for a different combination of light and viewpoint. The material segmentation is computed as follows: First, the CIE-LAB values of all base diffuse pixels L_b are clustered using k-means [Mac67] applying the CIEDE2000 [SWD05] color difference metric. Next, disconnected k-means clusters are split into individual segments. Finally, to compensate for over-segmentation (e. g., large gradients in the background), neighboring segments are merged if the average color difference on their boundary pixels is less than a certain threshold.

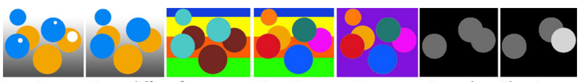

Figure 4: *From left to right: Starting from an input image L, with diffuse colors (yellow, blue) and different highlights (white spots), the diffuse base color L_b is used for the material segmentation to propagate specularity and glossiness.*

To extract texture information, we use a bilateral Laplacian pyramid [FAR07]. The pyramid is computed by convolving L_t with a bank of n_b bilateral filters and subtracting subsequent levels. Doing so, strong edges in L_t, are excluded from the frequency response. In practice, we use $n_\mathrm{b} = 4$ levels. An additional filtering using a 5×5 Gaussian filtering accounts for the local phase insensitivity of the HVS. This results in a spatially varying map $S : \mathbb{R}^2 \to \mathbb{R}^{n_\mathrm{b}}$.

Using the material segmentation, the diffuse color, specularity and texture coefficients are calculated within each segment using robust statistics [OAH11] resulting in k' materials. Depending on the scene structure, k' might be very different from the desired number of materials n_c. Therefore, we interpret the the k' materials once again as a high-dimensional (8 D) point cloud and cluster it once more, resulting in the desired n_c final materials. In practice, segmenting into a handful of discrete materials is sufficient to well-match the guide images.

We also support other guide inputs, such as image sequences that are processed on a per-frame basis and allow for direct sketching of a guide sources that we designed a simple user interface for. The system then updates the scene according to sketched materials (e. g., a red circle with a highlight and a blue square). To ease explanations, we will in the following consider a single guide image as input.

3.3. Material Assignment

An appropriate assignment A is found by minimizing a cost function $d(A,V)$. d is high, if the guide source is perceptually different for the material assignment A under view V. The cost is the sum of two components

$$d(A,V) = w_\mathrm{i} d_\mathrm{i}(r(A,V)) + w_\mathrm{g} d_\mathrm{g}(A). \tag{1}$$

A view-dependent image cost d_i compares the guide source to a rendered image using the current view and assignment. A view-independent geometric cost d_g verifies that the assignment is consistent with the geometry. $w_\mathrm{i}, w_\mathrm{g}$ determine the relative weighting between d_i and d_g respectively. We detail both components in the following, before showing how to efficiently optimize for this cost function.

Image cost d_i penalizes perceived difference of material appearance for all rendered views with respect to the guide source. Histogram calculations are used to compare the material mood of two images. To account for spatial variation, different *local* histograms are used in different parts of the image. We use a grid of 3×3. The CIE LAB color space is used, defining a 3D histogram with 10 bins in each dimension. Continuous bins are computed using Parzen [Par62] windows. Finally, the distance between two histograms is computed using their intersection [SB91].

The final cost is then defined as the integral of the image cost over all views. We could compute this integral using

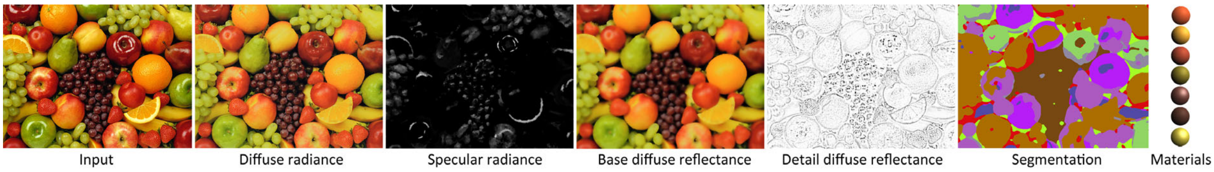

Figure 3: *Our material extraction pipeline is shown from left to right. The input, a single image, is decomposed into diffuse and specular radiance. Next, the diffuse radiance is split into base- and detail diffuse reflectance. The latter will be represented as local statistics. Finally, the full high-dimensional material info (diffuse reflectance, specular intensity, and the detail statistics) is segmented into discrete clusters which are assigned to the target 3d scene.*

quadrature, i. e., computing it for a high number of views, but, as shown later, picking just random views is a sufficient Monte Carlo approximation.

Geometry cost d_g penalizes geometrical differences between objects in the scene and the guide source that share the same material. A simple example illustrates the idea: A still-life-like guide source with a green apple and a yellow banana being is assigned to a 3D scene with an apple and a banana object. In terms of image costs, a green banana and a yellow apple are as plausible as a green apple and a yellow banana. Nonetheless, it is more intuitive to assign materials to similar shapes in the target and guide source, i. e., a yellow material to banana-like and a green material to apple-like shapes. Formally, we define:

$$d_{\mathrm{g}}(A) = \frac{1}{n_{\mathrm{o}}} \sum_{j=0}^{n_{\mathrm{o}}} \min_{0 \leq i < n'_{\mathrm{o}}} (d_{\mathrm{a}}(i,j,A_i,A'_j))$$

with

$$d_{\mathrm{a}}(i,j,A_i,A'_j) := \begin{cases} d_{\mathrm{s}}(i,j) & \text{if } A_i = A'_j \text{ and} \\ \mathrm{float}_{\mathrm{max}} & \text{else,} \end{cases}$$

where n_{o} and n'_{o} are the numbers of objects in the target and source, i and j are shapes of target and source, A and A' are the corresponding material assignments and d_{s} is a shape difference metric. To compute the difference between two shapes we employ a 2D Angular Radial Transform (ART) [Bob01], and use the corresponding shape descriptor vector composed of 35 elements. We use ART because of its simplicity, quick computation, and robustness for shape-based geometry matching [CTSO03]. We use the minimal Euclidean distance between the descriptor vectors over a range of orthographic views [CTSO03]. To make sure only meaningful matches contribute to the geometry cost, we redefine $d_g(A)$ as

$$d_g(A) = \frac{1}{n_{\mathrm{o}}} \sum_{j=0}^{n_{\mathrm{o}}} q_\rho(\min_{0 \leq i < n'_o}(d_a(i,j,A_i,A'_j))) \tag{2}$$

where ρ is a constant to define the minimum requirement for meaningful shape matching and

$$q_\rho(x) := \begin{cases} x & \text{if } x \leq \rho \text{ and} \\ 1 & \text{else,} \end{cases}$$

Note, that the geometric cost is view-independent.

If the guide source does not provide 3D shapes, i. e., for images and videos, the 2D shape is directly used in the shape metric. The shape extraction makes use of a different segmentation than can be deduced from the derived materials because, to use an example, two bananas in a guide source might form one material cluster, yet their shape is still the one of two single bananas, so $n_{\mathrm{m}} = 1$ while $n'_{\mathrm{o}} = 2$.

We use a multi-segmentation to decompose the guide source [RFE*06] because no single parameter setting for a segmentation approach will be sufficient to produce all meaningful segments. Mean shift [CM02] is used and the bandwidth parameter is varied to three different values to produce a fine, a medium, and a coarse segmentation. The resulting segments are different from the material segmentation, but only used to calculcate the geometry cost (Eq. 2).

3.4. Optimization

Given a cost function, we find the best assignment via simulated annealing (SA). SA is an iterative improvement algorithm, in which permuted "neighboring" solutions are used to evaluate the cost function and submitted to an acceptance test. Permutations leading to a cost decrement are always accepted while permutations resulting in an increase are accepted according to a probability based on the Boltzmann factor [KJV83].

In general, a cost function is *sampled* for the current solution $c^i = d(A^i, V)$ as defined in Eq. 1. Next, the current solution A^i is *permuted* to a neighboring solution B in a way that depends on the cost. If the cost c^i is high, a remote neighbor is chosen, if it is low, a close neighbor is selected. Once, a neighboring B is chosen, the next solution A^{i+1}.

To apply this principle to materials, we have to define a neighborhood on our solution domain, i. e., the space of material assignments $\mathcal{A}$. Intuitively, we want remote neighbors to change the material of many objects, and close neighbors to change the material of fewer objects, formally

$$B_j := \begin{cases} p(A^i_j) & \text{if } c^i > \xi_j \\ A^i_j, & \text{else,} \end{cases}$$

where $0 < \xi_j < c_{\mathrm{max}}$ (a user parameter) is a random number and $p(k) \in \{1, \ldots, n_{\mathrm{m}}\}^2$ a material index permutation function to be defined next. The permutation $1 < p(k) < n_{\mathrm{m}}$

maps a current material index k onto a new material index. In a simple implementation, $p(k)$ can be chosen to produce a random number between 1 and n_m. Our more advanced implementation permutes material assignments while maintaining a similar cost (Fig. 5). To find similar, i. e., nearby,

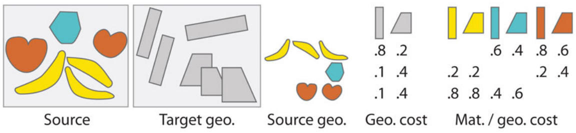

Figure 5: *Left to right: Source image and target 3D scene; the cost matrix of assigning materials to geometry; the cost $p(k)$ of permuting a material k*

materials, the metric of Pellacini and Lawrence [PL07] is used; it computes the average intensity difference over a high number of random incoming and outgoing sampling directions. To accelerate the AS optimization, we can write the distance between all materials as a $n_\mathrm{m} \times n_\mathrm{m}$ matrix $\mathtt{M}$. For each material, a row $\mathtt{M}_k$ in this matrix represents the similarity to all other materials. In the matrix $\mathtt{F}$ a row

$$\mathtt{F}_k := (\sum_{l=1}^{l=n_\mathrm{m}} \mathtt{M}_{k,l})^{-1}\mathtt{M}_k$$

is a probability density function. Finally,

$$f_k(x) := \mathrm{argmin}_l \sum_{m=1}^{m=l} F_{k,l} > x$$

is a cumulative density function. Using f, another random number $0 < \xi' < 1$ allows us to choose a similar material $p(k) := f_k(\xi')$ with a probability proportional to its similarity.

3.5. Implementation Details

Evaluating the cost function – which involves rendering the scene and computing global illumination for every sample – naïvely, is far too costly. A GPU in combination with precomputation is used to accelerate this computation. First, a deferred buffer [ST90] storing position, normal and material ID for all views is pre-computed. As described above, we pick a random view for every sampling. We use 256 views, each in a resolution of 256×256 pixels. At runtime, only the assignment of materials to material IDs in the buffer is changed. It is done in parallel over all views and all pixels using a simple shader program. To simulate light transport, Instant Radiosity [Kel97] is used. 256 virtual point lights with low-resolution shadow maps have shown to provide sufficient accuracy here. The shadow maps for all VPLs are independent of the samples and can be pre-computed as well.

4. Results

Results produced by our system are shown in Fig. 6. In the first and second rows, kitchen scenes are stylized, notice how different guide sources result in different moods (column). In the third row, the different combinations of specularity and texture frequency from the stones in the guide source are captured and transferred to the target. The scene shows a strong resemblance to the input photo despite the process being fully automatic. In the fourth row, even though the scene is mostly lit by indirect lighting, our system successfully captures the guide images' mood. While in Fig. 10 a guide painting is used, the construction of our Phong BRDF ensures that materials stay realistic. Fig. 7 uses a guide video. Our algorithm managed to detect the round shape of the juggling balls and associated them adequately to the scene. The results presented are produced without human intervention and computed in under 2 minutes.

Figure 7: *Transfer from a video to a target 3D scene. Note how the juggling balls are mapped to the small spheres.*

Our approach can also serve as the basis of a rough user sketch-based system (Fig. 8). By analyzing the drawn shapes, similarly shaped 3D objects are attributed the material that is indicated by the user. By drawing specularities, the user can further modify the specular coefficient. The process is interactive and the transfer on target scenes can facilitate material attribution in 3D scenes. One can handle all similar objects can at once, or restrict modifications to a given target view. When our system is presented a guide source where ground truth material information is available, it successfully recovers those materials to a large extent (Fig. 9). The two additional materials in our reconstruction, are due to shadows that were segmented into additional individual materials.

5. Discussion

Perceptual study To validate the performance of our approach, a perceptual study was performed in two steps. In the first step, human performance in terms of quality and speed of assigning materials to a 3D scene according to a guide source was analyzed. Using the free software "Blender", 14 subjects were asked to assign materials to a 3D scene according to a guide image. Both expert (computer graphics graduate students, faster half) and non-expert (university graduate students, slower half) participated. The subjects received instructions and training on how to perform assignments and their results were captured after finishing (usually 20 minutes). There was a total of 13 assignment tasks and on average

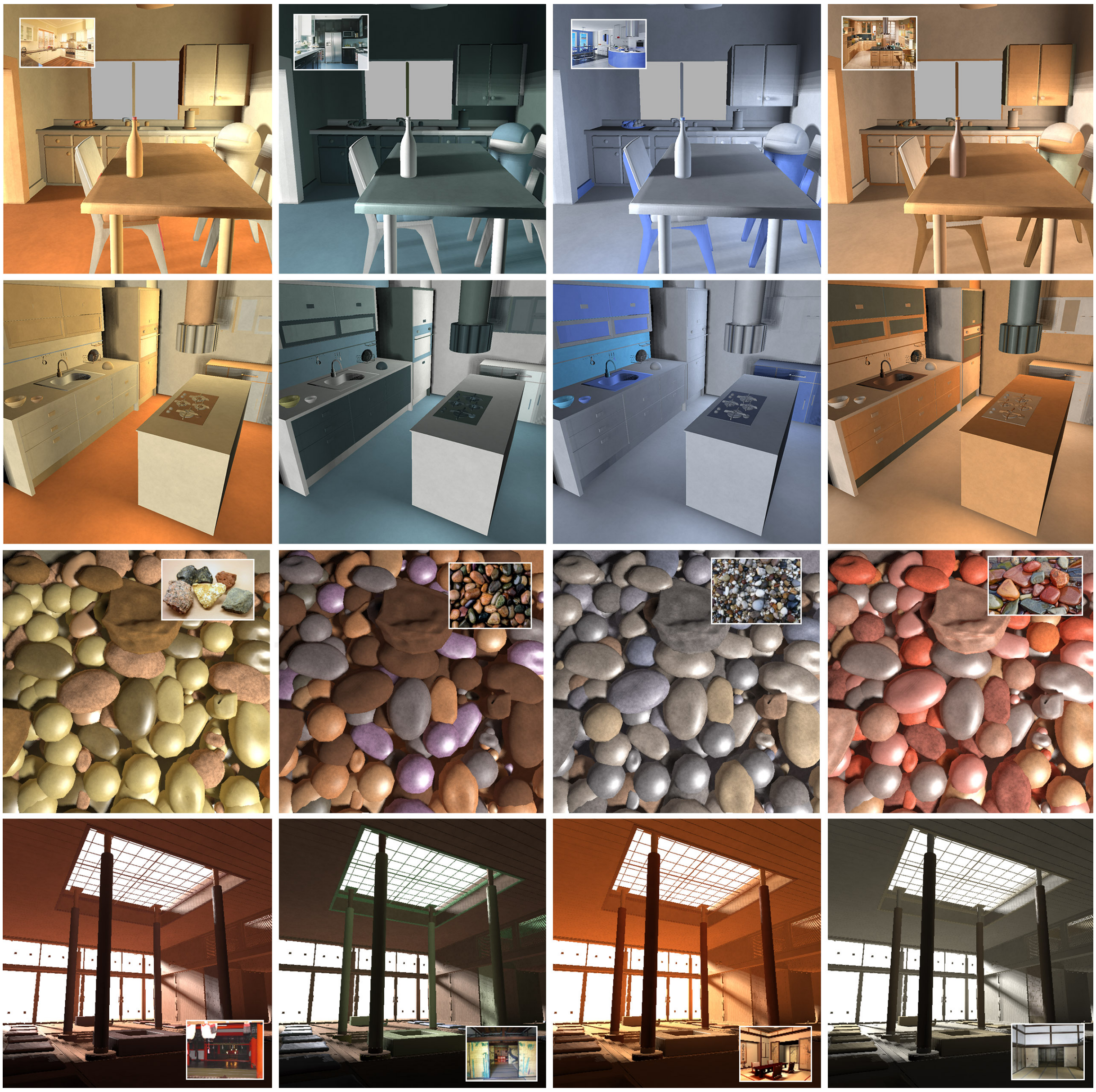

Figure 6: *3D Material style transfer on different target scenes (rows) from different sources (insets).*

every subjects performed 4 assignments. Please see the supplemental material for details. At the end of the study, when asked "How pleasant was the material assignment tasks?" the common answer was "Conveying mood using material assignment is not intuitive, especially in case of dominant indirect lighting". There were a couple of comments indicating that material assignment task was really difficult, such as "I think, my results were not good enough but had no idea how to improve them anymore, searching for the right colors was so difficult and tedious".

In a second step, 20 other subjects were asked to rank the result images obtained in the first step. For each of the 13 assignment tasks we grouped our result and all results of the manual assignment after 5 mintues (avg. group size: 5 images). Subjects were then asked to "sort the images from *best* to *worst* in order of mood similarity to the guide image". Our algorithm can produce an assignment in less than 2 minutes, however, 64 % rank our automatic assignment best (84 % best or second-best). Asking the same question for a different group of images that contained our result and all

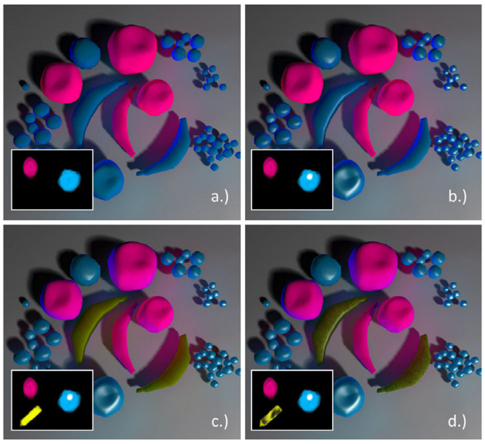

Figure 8: *Four steps of user interaction.* a.) *A user draw two spheres in two colors, being mapped to distinct shapes.* b.) *A highlight is added to the blue patch, resulting in highlights to appear on blue shapes in the target.* c.) *An elongated yellow shape is added that is mapped to the banana.* d.) *Noise on the yellow shape makes the banana appear textured.*

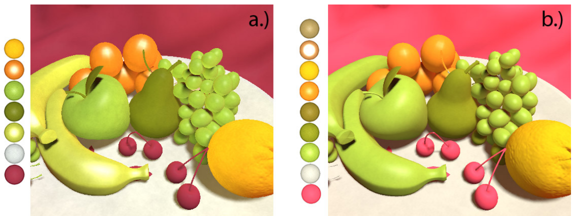

Figure 9: a.): *Our material extraction applied to* b.): *a rendering of a 3D scene where ground truth materials are known.*

results of the final assignment, 60 % rank our automatic result better then all other assignments (80 % best or second-best). One conclusion from the relatively low improvement between 5 and over 20 minutes achieved by manual assignment is that, it is very hard to converge from an acceptable initial result to a final, global illumination-compatible assigment that captures the scene's mood such as produced by our system.

Figure 10: *Comparison between using our combination of image and geometry cost* (Left), *using only the image cost* (Middle) *and using only the geometric cost* (Right).

Image and geometry cost Our cost function consist of an image and a geometry term. Both play an important role and only the combination of the both aspects will lead to a successful assignment results (Fig. 10). Using only the image term will give a similar image appearance, but wrong individual object materials. Only using the geometric term will assign the right materials to objects, but with no consistent global organization.

Limitations and Assumptions While the system often succeeds in transferring the mood from a source image onto a 3D scene, it is subject to several limitations. In summary, our system performs best for input images and target 3D scenes, which can be well-segmented with neutral directional lighting and similarities between source and target.

Certain assumptions about the image and the 3D scene have to be made. The target 3D scene has to be segmented into meaningful objects, i. e., objects that can be assigned a discrete material label. In future work, spatial variation should be addressed by generalizing the discrete assignment into a continuous mapping. User-annotation (such as labels) and structure information (such as symmetry) could be included in future work. The source image has to be automatically segmented, which is a hard problem and we can not expect it to always work. While material extraction works well with imperfect segmentations, shape similarity is more sensitive. To circumvent this problem, we use multiple segmentations. If no matches are found, the geometry cost (Eq. 2) will approach a constant for all assignments and optimizing Eq. 1 will revert to optimizing for the image cost alone.

Many limitations stem from the difficulty to robustly extract materials from images. We focus on visually plausible materials in combination with an optimization based on similarity to a guide image. Perfectly reconstructing physically-correct reflectance from a single image is very difficult without extra assumptions on the object geometry and scene lighting (Sec. 2) that we avoid. Nonetheless, our assumption of material appearance constancy might not always hold; hard shadows or unusual highlights happen to be erroneously interpreted as individual objects. Also, textures under strong perspective transformations are sometimes grouped into different materials. Special materials like glass and other transparent materials are not yet supported.

6. Conclusion

We proposed a technique to transfer material style from a guide source to a target 3D scene. The problem was stated as a combinatorial optimization of assigning discrete materials (approximately extracted from the guide source) to discrete objects in the target according to an image and geometry cost function. Our system can effectively capture materials in terms of highlights and diffuse detail textures and efficiently apply them to a target scene or even whole databases in a meaningful and automatic way.

References

[Ber48] BERG E. A.: A simple objective technique for measuring flexibility in thinking. *J General Psychology 39* (1948), 15–22. 2

[Bob01] BOBER M.: MPEG-7 visual shape descriptors. *IEEE Circuits and Sys. Vid. Tech. 11*, 6 (2001), 716–719. 4

[CH10] CASTELHANO M. S., HENDERSON J. M.: The influence of color on perception of scene gist. *J Vision 5*, 8 (2010), 68–8. 2

[CLS10] CHAJDAS M. G., LEFEBVRE S., STAMMINGER M.: Assisted texture assignment. In *Proc. I3D* (2010), p. 173. 2

[CM02] COMANICIU D., MEER P.: Mean shift: a robust approach toward feature space analysis. *IEEE PAMI 24*, 5 (2002), 603–19. 4

[CTSO03] CHEN D.-Y., TIAN X.-P., SHEN Y.-T., OUHYOUNG M.: On Visual Similarity Based 3D Model Retrieval. *Computer Graphics Forum 22*, 3 (2003), 223–232. 4

[DD02] DURAND F., DORSEY J.: Fast bilateral filtering for the display of high-dynamic-range images. *ACM Trans. Graph. (Proc. SIGGRAPH) 21*, 3 (2002), 257–266. 3

[FAR07] FATTAL R., AGRAWALA M., RUSINKIEWICZ S.: Multiscale shape and detail enhancement from multi-light image collections. *ACM Trans. Graph (Proc. SIGRAPH) 26*, 3 (2007), 51. 3

[HB95] HEEGER D. J., BERGEN J. R.: Pyramid-based texture analysis/synthesis. In *Proc. SIGGRAPH* (1995), pp. 229–238. 2

[HJO*01] HERTZMANN A., JACOBS C. E., OLIVER N., CURLESS B., SALESIN D. H.: Image analogies. In *Proc. SIGGRAPH* (2001), pp. 327–340. 2

[ICOL05] IRONY R., COHEN-OR D., LISCHINSKI D.: Colorization by Example. *Proc. EGSR* (2005), 201–210. 2

[JDA*10] JOHNSON M. K., DALE K., AVIDAN S., PFISTER H., FREEMAN W. T., MATUSIK W.: CG2Real: Improving the Realism of Computer Generated Images using a Large Collection of Photographs. *IEEE TVCG* (2010). 2

[Kel97] KELLER A.: Instant radiosity. *Proc. SIGGRAPH 31*, 3 (1997), 49–56. 5

[KJV83] KIRKPATRICK S., JR C. D. G., VECCHI M. P.: Optimization by simmulated annealing. *Science 220*, 4598 (1983), 671–680. 4

[KRFB06] KHAN E. A., REINHARD E., FLEMING R. W., BÜLTHOFF H. H.: Image-based material editing. *ACM Trans. Graph. (Proc. SIGGRAPH) 25*, 3 (2006), 654. 2

[Lav08] LAVANYA SHARAN R. R. . E. H. A.: Rapid visual perception of material properties. In *Proc. Workshop on Object Perception, Attention & Memory* (2008). 2

[LKG*03] LENSCH H. P. A., KAUTZ J., GOESELE M., HEIDRICH W., SEIDEL H.-P.: Image-based reconstruction of spatial appearance and geometric detail. *ACM Trans. Graph. 22*, 2 (2003), 234–257. 2

[LVLD10] LAGAE A., VANGORP P., LENAERTS T., DUTRÉ P.: Procedural isotropic stochastic textures by example. *Computers & Graphics 34*, 4 (2010), 312–321. 2

[LWCO*07] LUAN Q., WEN F., COHEN-OR D., LIANG L., XU Y.-Q., SHUM H.-Y.: Natural Image Colorization. In *Proc. EGSR* (2007). 2

[LWQ*08] LIU X., WAN L., QU Y., WONG T.-T., LIN S., LEUNG C.-S., HENG P.-A.: Intrinsic colorization. *ACM Trans. Graph. (Proc. SIGGRAPH Asia) 27*, 5 (2008), 1. 2

[Mac67] MACQUEEN J.: Some methods for classification and analysis of multivariate observations. In *Proc. Berkeley Symp. Math. Stat. Prob.* (1967), vol. 1, p. 14. 3

[MKCD07] MERTENS T., KAUTZ J., CHEN J., DURAND F.: Texture Transfer Using Geometry Correlation. In *Proc. EGSR* (2007). 2

[MPBM03] MATUSIK W., PFISTER H., BRAND M., MCMILLAN L.: A data-driven reflectance model. *ACM Trans. Graph. 22*, 3 (2003), 759. 2

[OAH11] O'DONOVAN P., AGARWALA A., HERTZMANN A.: Color compatibility from large datasets. *ACM Trans. Graph. (Proc. SIGGRAPH) 30*, 4 (2011), 1. 3

[OT01] OLIVA A., TORRALBA A.: Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. *Int. J. Computer Vision 42*, 3 (2001), 145–175. 2

[Par62] PARZEN E.: On estimation of a probability density function and mode. *Ann. Math. Statist 33* (1962), 1065–1076. 3

[Per85] PERLIN K.: An image synthesizer. *ACM SIGGRAPH Computer Graphics 19*, 3 (1985), 287–296. 2

[PFG00] PELLACINI F., FERWERDA J. A., GREENBERG D. P.: Toward a psychophysically-based light reflection model for image synthesis. In *Proc. SIGGRAPH* (2000), pp. 55–64. 2

[Pho75] PHONG B. T.: Illumination for computer generated pictures. *Communications of the ACM 18*, 6 (1975), 311–317. 2

[PL07] PELLACINI F., LAWRENCE J.: AppWand: editing measured materials using appearance-driven optimization. *ACM Trans. Graph. (Proc. SIGGRAPH) 26*, 3 (2007), 54. 2, 5

[RAGS01] REINHARD E., ASHIKHMIN M., GOOCH B., SHIRLEY P.: Color Transfer between Images. *IEEE Computer Graphics and Applications 21*, 5 (2001), 34–41. 2

[RFE*06] RUSSELL B., FREEMAN W., EFROS A., SIVIC J., ZISSERMAN A.: Using Multiple Segmentations to Discover Objects and their Extent in Image Collections. In *Proc. CVPR* (2006), pp. 1605–1614. 4

[SB91] SWAIN M. J., BALLARD D. H.: Color indexing. *Int. J Computer Vision 7*, 1 (Nov. 1991), 11–32. 3

[ST90] SAITO T., TAKAHASHI T.: Comprehensible rendering of 3-D shapes. *ACM SIGGRAPH Computer Graphics 24*, 4 (1990), 197–206. 5

[Str35] STROOP J. R.: Studies of interference in serial verbal reactions. *J Exp. Psy.: General 18*, 6 (1935), 643–662. 1, 2

[SWD05] SHARMA G., WU W., DALAL E. N.: The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application 30*, 1 (2005), 21–30. 3

[TT00] TOMINAGA S., TANAKA N.: Estimating reflection parameters from a single color image. *IEEE Computer Graphics and Applications 20*, 5 (2000), 58–66. 2

[WM02] WALKER L. L., MALIK J.: When is scene recognition just texture recognition? *J Vision 2*, 7 (2002), 255–255. 2

[WYW*10] WANG B., YU Y., WONG T.-T., CHEN C., XU Y.-Q.: Data-driven image color theme enhancement. *ACM Trans. Graph. 29*, 6 (2010), 1. 2

[YWA10] YANG Q., WANG S., AHUJA N.: Real-time Specular Highlight Removal Using Bilateral Filtering. In *Proc. ECCV* (2010), pp. 87–100. 3