Composition-Aware Scene Optimization for Product Images

Tianqiang Liu[1] Jim McCann[†2,3] Wilmot Li[2] Thomas Funkhouser[1]

[1]Princeton University [2]Adobe Research [3]Disney Research

Abstract

Increasingly, companies are creating product advertisements and catalog images using computer renderings of 3D scenes. A common goal for these companies is to create aesthetically appealing compositions that highlight objects of interest within the context of a scene. Unfortunately, this goal is challenging, not only due to the need to balance the trade-off among aesthetic principles and design constraints, but also because of the huge search space induced by possible camera parameters, object placement, material choices, etc. Previous methods have investigated only optimization of camera parameters. In this paper, we develop a tool that starts from an initial scene description and a set of high-level constraints provided by a stylist and then automatically generates an optimized scene whose 2D composition is improved. It does so by locally adjusting the 3D object transformations, surface materials, and camera parameters. The value of this tool is demonstrated in a variety of applications motivated by product catalogs, including rough layout refinement, detail image creation, home planning, cultural customization, and text inlay placement. Results of a perceptual study indicate that our system produces images preferable for product advertisement compared to a more traditional camera-only optimization.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Animation I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—Physically based modeling

1. Introduction

A growth application for computer graphics is creation of images for product advertisements and catalogs [Ent12, Sou12, Stu14]. As photorealistic rendering algorithms have improved, it has become practical to synthesize images that are indistinguishable from photographs for many types of scenes commonly found in product advertisements (e.g., kitchens, bathrooms, living rooms, etc.). As a result, several furniture and home goods companies are beginning to create product images for their catalogs by rendering 3D models rather than photographing physical objects [Sou12]. For example, IKEA has reported that 75% of scenes shown in its most recent catalog were rendered from 3D models [Stu14].

There are many advantages to creating catalog images from 3D models [Ent12]. Rendering virtual scenes is much less expensive than photographing real scenes because it does not require building physical sets in large photo studios, storing physical objects in large warehouses, and scheduling actors, stylists, and photographers to meet for photo shoots. Moreover, digital assets make it easier to customize images in a variety of ways, such as producing multiple images of the same scene with different objects of interest, adapting scene composition to the resolution and aspect ratio of the display device, adapting the size and placement of text labels for different languages.

Despite these advantages, producing good product images from 3D models is still difficult. Based on interviews with professionals who work on product catalogs, we learned that the typical workflow for a *stylist* (person who designs scenes for product images) is to first create an approximate scene layout with the appropriate objects (furniture, accessories, etc.) in roughly the desired configuration (e.g., a sofa with an end table on the left, a lounge chair on the right, and a coffee table with some plants and coasters sitting on top). Based on this rough layout, the stylist will then carefully refine the positions, orientations and materials of objects as well as the camera viewpoint to compose several different images of the scene that highlight different objects of inter-

---

† work performed while at Adobe

est, fit different display devices (e.g., iPhone, desktop, print, etc.), and in some cases, target different cultures (e.g., IKEA makes 62 variants of its catalog for 43 countries). This refinement step is challenging because it requires taking into account the image-space position, size, visibility, and color contrast of objects of interest, as well as the overall composition of each image. As a result, stylists often spend multiple days refining the initial rough scene layouts to produce all the necessary product images.

In this work, we present a tool that facilitates the creation of product images by automatically refining rough scene layouts to produce good compositions. Our tool starts with an approximate scene description provided by a stylist that includes which objects should appear in the scene, which objects rest upon which other objects, and which materials can be used for which objects, plus an initial configuration for object positions/orientations, surface materials, lighting parameters, and (optionally) camera views. Our tool then optimizes the camera view, object transformations and surface materials to meet user-specified design goals (e.g., highlight these objects of interest, fit the image within a specific form factor, leave space for a text box) while maintaining the compositional quality of the image.

The main contribution of our work is an optimization approach that takes into account a large variety of design constraints and image composition rules that are important for producing effective product images. We define a (highly non-convex) energy function that models these criteria and then introduce an iterative optimization procedure that minimizes the energy. One key feature of our method is that we simultaneously manipulate object transformations, surface materials, and the camera view, all of which have a significant impact on the quality of the resulting image. We show that our approach produces better results than a more traditional camera-only optimization, and we also demonstrate how our tool can help create product images for a variety of practical applications.

2. Related Work

Our work draws upon previous work in image composition and aesthetics, image analysis and optimization, virtual camera control, and automatic scene synthesis.

Image composition and aesthetics: Our work is inspired by composition "rules" that have been established to guide photographers and graphics designers towards better scene compositions and aesthetics [Arn88] [Bet56] [Cli73] [GT90] [Kra05] [MB88] [Tay38]. Well-known examples include the "rule of thirds," visual balance, diagonal dominance, and color contrast. Although previous work has considered subsets of these rules for automatic image composition, one of our key contributions is in determining a set of rules suitable for product images and applying these rules to a challenging 3D scene optimization problem. Compared to existing

methods, our approach considers a larger set of 3D scene parameters — object transformations, surface materials, and the camera view — that are important for generating effective product images.

Image analysis and optimization: Several papers have used these rules to quantify [DJLW06, DW10] and enhance the compositional and aesthetic quality of images. However, all previous methods operate only on edits to 2D images: for example, rotating and cropping images [KMM08, LCWCO10, JWL12], or adjusting locations of foreground regions [BSS10, LFN04] – they do not optimize 3D scene parameters, such as cameras, materials, and/or object transformations, as our system does. Our work is also related to the technique that allows the user to alter the composition after 3D rendering [Gri01], but our system is able to automatically optimize the composition by refining 3D scene layouts and materials.

Camera optimization: Several methods have incorporated principles of image aesthetics and composition in optimization algorithms for camera control in 3D rendering systems [OHPL99] [GRMS01] [CON08] [BMBT00] [Bar06]. While these papers provide motivation for our work, they consider only camera control – we additionally optimize object transformations and surface materials, which can significantly improve image compositions, but require solving a more difficult optimization problem.

Scene optimization: Several recent papers have proposed methods for automatically placing objects in scenes to produce plausible furniture layouts based on examples and design guidelines [YYT*11, FRS*12, MSL*11]. These methods focus on scene plausibility without concern for any particular camera viewpoint and/or image composition principles. As a result, they produce scenes that may not support generation of aesthetic images from any camera viewpoint. Our work is also related to arrangement synthesis based on aesthetic relations [MSSH13]. Instead of optimizing relations among objects, our goal is to optimize the 2D composition produced by a 3D scene layout.

We believe that ours is the first system to optimize aesthetics and composition of rendered images with simultaneous control over camera parameters, object transformations, and surface materials. We investigate this optimization problem for the novel application of image synthesis for product catalogs.

3. Overview

The core of our work is a method for optimizing 2D compositions of rendered 3D scenes by adjusting camera parameters, object transformations, and surface materials.

Our system works with a database of 3D object models, where each model is annotated with an object class

(e.g. chair, table) and a list of possible materials. Each object class is associated with a set of 1-4 canonical views, which can be manually specified or predicted automatically [BTBV99, GRMS01, SLF*11]. Some object classes are also associated with a set of semantic spatial constraints. For examples, picture frames on walls should not be rotated, and the distance between a dining chair and a dining table should be maintained.

Given the model database, a stylist begins the process of creating a set of product images by specifying a *scene configuration*, which consists of a set of object models, light sources, and a rough location/orientation for every object. With the rough scene in place, the stylist can instruct our system to create a specific image by specifying an *image configuration*, which consists of desired *aspect ratio* and *focus objects* $\mathbf{O_F}$ to highlight. The stylist may optionlly specify initial camera parameters and *context objects* $\mathbf{O_C}$ that should remain visible for context.

From this input, our system optimizes the scene description to generate a set of rendered images. In particular, our method optimizes the following scene parameters (plus other application-specific variables described in Section 6), with the degrees of freedom listed in parentheses:

- *Camera (6):* position (3), direction (2), and field of view (1) (camera roll is constrained to be zero).
- *Object transformations (3 per object):* position of the object centroid on its support surface (2) and rotation of the object around the normal of its support surface (1).
- *Materials (1 per object):* choice of a material/texture definition amongst a list of possible candidates.

Our system optimizes these parameters according to an energy function that accounts for image composition, aesthetic principles, and object focus, while maintaining 3D spatial constraints and 2D image space layout constraints.

The following sections describe our energy function and optimization procedure in detail.

4. Energy Function

Our energy function estimates how effectively an image advertises the product(s) it depicts.

To design the energy function, we interviewed professionals responsible for creating scenes for popular product catalogs, worked with them to identify a set of principles important for product image composition, and encoded how well those principles are satisfied into a mathematical error function. We interviewed two professionals: 1) a professional stylist who lays out scenes for Pottery Barn catalogues, and 2) the lead of image synthesis for IKEA catalogues. They identified twelve different factors as critical for composition of product images. The following lists these factors, grouped roughly into six categories, along with an explanation of why each is

important (Figure 1). Mathematical details of how each one is encoded as a term in our energy function are deferred to Appendix A for brevity.

- **Object placement within the 2D frame.** For product advertisements, some of the most significant factors affecting image quality are the positions of focus objects, which we encode with the following terms.

  - **Rule of thirds.** In general, focus objects should align with vertical or horizontal lines that divide the viewport into thirds and/or be centered at the intersections formed by them [BE04] [Bar06] [BDSG04] [DJLW06] [GRMS01] [LCWCO10] [War03].
  - **Centeredness.** For some product images (e.g., zoomed views of single objects), focus objects should appear in the center of the image [Arn88].
  - **Clearance.** Since other objects should not compete with the focus objects in the composition, we introduce a term to penalize objects that are close to focus objects.

- **Object saliency within the 2D frame.** The visibility and image-space size of objects contribute to their perceived importance [BMBT00] [Bar06] [BFH01] [OHPL99].

  - **Visibility.** An object is perceived as less important if it is partially occluded by another object or partially clipped by the frame. Thus, focus objects should be more visible than context objects.
  - **Object size.** Similarly, focus objects should take up more image space than context objects, since larger objects are perceived as more important.

- **Object constraints within the 3D scene.** Scene plausibility and physical laws both impose constraints on object positions. We therefore introduce several terms to enforce these constraints.

  - **Semantic constraints.** Semantic spatial constraints that are specified in the model database should be satisfied during the optimization.
  - **Collision relationships.** Object inter-penetrations should be avoided to improve the physical plausibility of the scene.
  - **Support relationships.** The support relationships that exist in the input scene layout must be maintained during the optimization.

- **Camera placement.** Product images generally depict scenes from viewpoints that are "natural" for people. We introduce two terms that capture the notion of natural viewpoints.

  - **Canonical views.** In most product images with one focus object, stylists favor canonical views of the object class [BTBV99, GRMS01, SLF*11].
  - **Typical views.** Product images that depict large scenes with multiple focus objects often use camera view-
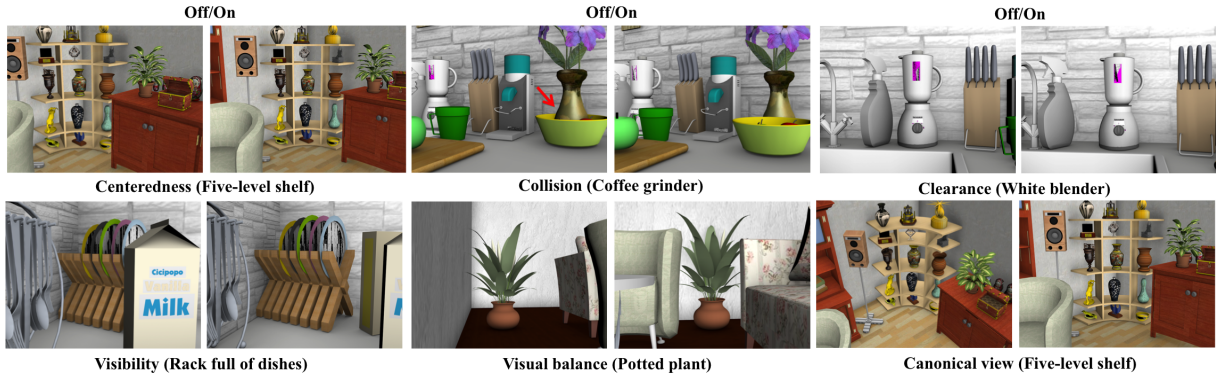
Figure 1: *Effects of disabling energy function terms. For each energy term, we compare the result with the term disabled (left) to our result (right). The focus object(s) is specified in the parentheses. Examples showing the effect of all terms can be found in the supplemental material.*

points that match how a human would typically see the scene.

- **Image composition.** Several well-established composition guidelines are used by stylists to create aesthetically pleasing images. We have included several in our system.

  - **Visual balance.** Images whose "center of mass" is close to the center of the image frame generally have better aesthetics [Arn88] [LCWCO10] [LFN04].
  - **Color contrast.** Greater color contrast at object contours can help a viewer understand boundaries between shapes in a scene [KHRO01] [WL11].

- **Regularization.** Finally, we add a regularization term that encourages small changes to the scene with respect to the initial configuration provided by the stylist.

The large number of terms in this energy function reflects the inherent complexity of composing a good product image. The necessity of each and every term has been confirmed by our experts and validated experimentally. As an example, Figure 1 shows the effects of removing some of the energy terms: each omission yields a product image of inferior quality (results of experiments omitting every individual term can be found in the supplemental material). Overall, our conclusion is that composition of product images is very difficult as many competing considerations must be balanced.

5. Optimization

Our optimization procedure searches for camera parameters, object placements, and surface materials that minimize the energy function.

This is a difficult optimization problem for several reasons: 1) there are many free variables (six for the camera, three for each object transformation, one for each surface with multiple candidate materials); 2) some of the variables are continuous (camera and object transformations) while others are discrete (surface materials); and 3) the energy function is

highly non-convex, with strong dependencies between multiple variables (e.g., camera and object movements). As a result, we can only hope to find a good local minimum.

Our approach is to decompose the problem into two simpler, more tractable optimizations that we interleave in an iterative algorithm. Within each iteration, the first step is to optimize the discrete choices of materials with camera parameters and object transformations fixed. Then, we optimize the continuous camera parameters and object transformations with the materials fixed. The iterations terminate when neither step changes the scene significantly in the same iteration.

5.1. Discrete optimization

We use a discrete steepest-descent algorithm to optimize materials. The input to the algorithm is a scene and a list of candidate definitions for each surface material, and the output is a selection of one candidate definition for each material that minimizes the energy function. Note that the color contrast term (E_{cc}, defined in Appendix A) is the only one affected by material switches. Higher color contrast for focus objects corresponds to a lower overall energy.

The algorithm first builds a list of visible objects with multiple candidate materials. Then, it iteratively optimizes the materials for each such object in order from the one with the lowest to the highest color contrast. For each object, the algorithm selects the material switch that produces the highest color constrast. The algorithm stops iterating over the objects when no material switches are possible to improve the color contrast further, which usually occurs within 2-3 iterations through all objects.

5.2. Continuous optimization

We use a continuous steepest-descent algorithm to optimize camera parameters and object transformations. The following paragraph describes how the direction and magnitude of each step is computed.
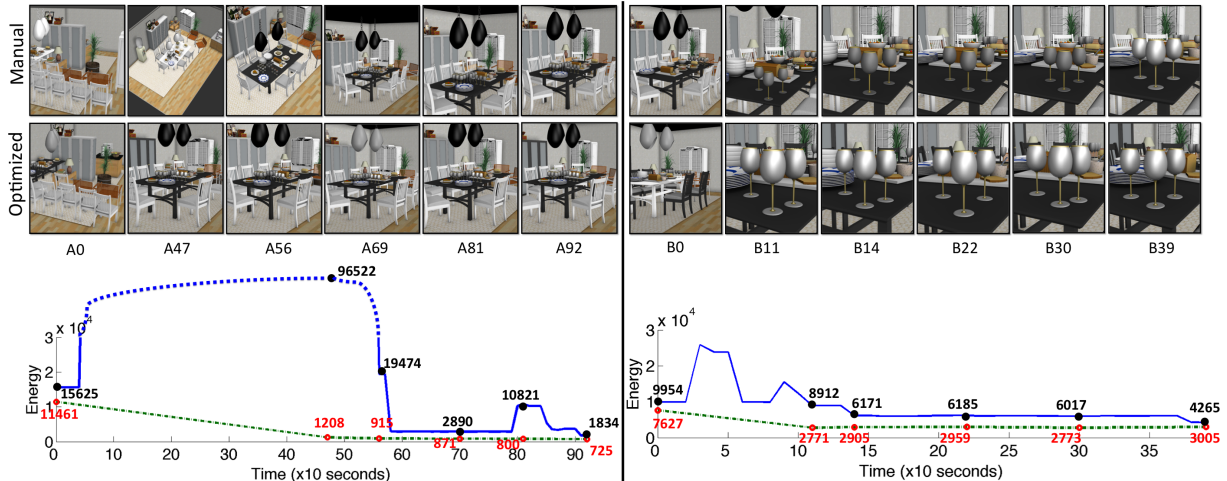
Figure 2: *Snapshots of an interactive session (top row) and the results of refining them by our optimization tool (second row). In the first session (left), the user's goal is to achieve the composition in Figure 3 left, while in the second session (right), her goal is to achieve Figure 3 right. The plots on the bottom show the evaluation of these scenes using our energy function, with the blue representing the energy of interactive snapshots and the red points representing our optimized results. Note that the dotted section of the lefthand blue curve has been compressed to save space.*

Since the energy function contains terms whose partial derivatives are difficult to compute analytically (e.g., visibility), we compute the derivative of the energy with respect to each free variable via finite differences. Of course, a brute force implementation of centered differences for each variable would be extremely slow: a typical scene has approximately 150 free variables (3 for each of ~50 object transformations plus 6 camera parameters), and thus the energy would have to be computed 300 times for each steepest descent step. Instead, we use a technique similar to stochastic gradient descent. Specifically, we keep estimates for all partial derivatives and re-estimate only a subset after most steps. Specifically, every k steps, we estimate partial derivatives for all variables, except ones for transformations of objects outside the view frustum, and make a move along the direction of steepest descent determined by all partial derivatives. We also build a list of objects T that have non-zero partial derivatives. Then, during the intervening steps, we re-estimate partial derivatives only for the camera parameters and k randomly selected objects from T and make a steepest descent move based on these derivatives. We choose $k = \sqrt{|T|}$, which provides a nice trade-off between efficiency and accuracy, leveraging the fact that fewer objects have significant effect on the energy as the optimization converges. To compute the magnitude of each steepest descent step, we conduct a line search along the direction of the estimated derivative.

5.3. Timing

The full optimization procedure takes approximately 20 minutes for the most complex examples in this paper. The discrete optimization step is usually very fast (<10 sec-

onds), since there are relatively few (~10) candidate materials in most scenes. The continuous optimizations are slower, since there are many possible object transformations in most scenes (~60 objects per scene in our examples) and computing partial derivatives for each transformation variable requires rendering the scene multiple times. In our experience, computing partial derivatives takes ~90 seconds for all variables (every k steps), but only ~10 seconds for our randomly chosen subsets (intervening steps), at no observed accuracy difference. All times are reported for a 2660 MHz Intel Core i7 processor with 8 GB of memory.

6. Applications

In this section, we describe several applications of our scene optimization framework. These applications were chosen based on the suggestions of the same two professionals we interviewed to determine the relevant rules for creating effective product images.

6.1. Refining rough compositions

The primary application of our system is to facilitate the refinement stage of digital catalog image creation. Given a set of focus objects and a rough scene configuration as initialization, we can apply the optimization procedure described in the previous section to automatically adjust the camera, object positions, and materials.

To evaluate whether our system can assist this application, we ran an informal experiment in which we asked a user with formal training in image composition to go through the full process of creating and refining a scene for a product image using an interactive modeling tool, and then investigated

how our tool could have helped during the modeling process. We instructed the user to create a 3D scene motivated by an image highlighting a dining room table and chair in the IKEA catalog (Figure 3 left), which she could refer to as she modeled. During the session, she started with a set of objects, candidate materials, and a random camera viewpoint (Figure 2, A0), and then edited the scene interactively to achieve the final result shown in Figure 3 left. We then asked her to further refine the composition to recreate the image in Figure 3 right, which highlights the three goblets on the dining table. The experiment was performed exactly once with no feedback from the system regarding composition quality.

During these interactive sessions, we logged a "snapshot" scene file every 10 seconds representing the user's progress (several examples are shown in the top row of Figure 2). After the session was finished, we used the snapshot scenes to: 1) analyze whether our energy function explains changes made interactively by the user, and 2) to study at what point in the modeling process our optimization procedure could have been used to assist the user by refining the scene automatically.

The blue curve in the plot at the bottom of Figure 2 shows the value of our energy function for each snapshot of the user's interactive session. Note that for each session, the curve reveals two phases: a period of "large-scale layout" when the scene energy goes up and down ($A0 \rightarrow A47$ and $B0 \rightarrow B11$), followed by a period of "fine-scale refinements" where the energy decreases almost steadily ($A47 \rightarrow A92$ and $B11 \rightarrow B39$). This behavior suggests that the energy function correctly captures image quality differences of improvements made by the user.

The second row of images in Figure 2 shows the results of running our optimization procedure on each of the snapshot scenes shown in the top row, and the red dots in the plot below show the energy function of the optimized results (connected by a green curve). Note that the optimized results of the snapshots (bottom row) captured in the latter half of each user session (A47-A92 and B11-B39) are qualitatively similar to the final scene created by the user (top-right image), and their corresponding energy function values are comparable, or even less. These results suggest that more than half of the time the user spent on scene refinement could have been off-loaded to the computer.

6.2. Generating detail images from an overview

In many cases, catalogs provide an overview image that shows how various objects can fit together in a room, and then one or more detail images that focus on individual products of interest. Detail images are almost never simply cropped and zoomed-in versions of the overview image. Stylists typically choose different viewpoints and move ob-

Figure 3: *Overview and detail images in IKEA catalog. In addition to the overview image on the left, IKEA provides a detail image that advertises the glasses on the table. Note how the viewpoint and object positions are adjusted from the overview image (reprinted with permission from the 2013 IKEA catalog).*


(a) *Overview* **(b)** *Speaker* **(c)** *Shelf*

Figure 4: *Detail images generated from overview. From an overview image of a living room (a), we automatically generate detail images that highlight the speaker (b) and shelf (c). Notice how the chair moves to the right in (b) and to the left in (c) to provide an unobstructed view of the focus object (results without moving objects can be found in Figure 9).*

jects slightly in order to highlight the shape and relevant features of the focus object (e.g. Figure 3).

To reduce this effort, stylists can use our system to automatically create detail images. For each detail object, our optimization framework initializes all object positions to the arrangement in the overview image and generates a set of candidate detail images using each canonical view of the detail object as a different starting point for the camera. By default, we choose the candidate image with the lowest energy as the result.

As a test of our method, we generated detail images for three scenes: a kitchen, study, and living room. For each scene, we generated detail images for 20 random objects. In many cases, our optimization was able to put the camera close to canonical views only by moving objects that would otherwise occlude the detail object. For example, the gray chair is moved to different positions in Figure 4b and Figure 4c in order to reduce occlusions for different focus objects. Overall, most of our detail images produce reasonable compositions, and our perceptual study (Section 7) indicates that our full optimization produces better results than camera-only optimizations in a large majority of cases.
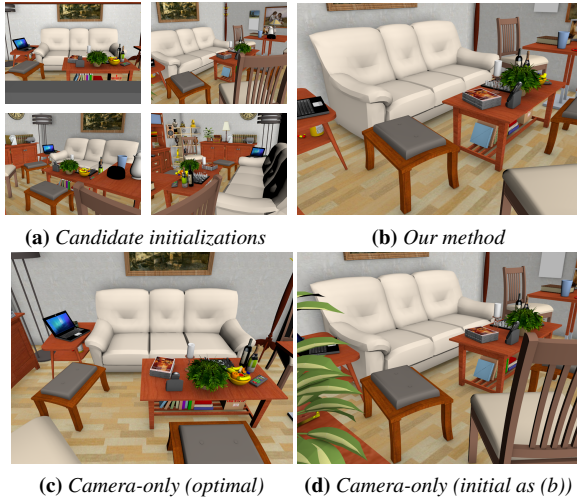
(a) *Candidate initializations*   **(b)** *Our method*

(c) *Camera-only (optimal)*   **(d)** *Camera-only (initial as (b))*

Figure 5: *Optimizing without a starting camera for a room planner application. Our system starts by sampling plausible cameras to generate a set of initial views (a), which are then optimized (b). Running camera-only optimization over all initial views yields (c), while starting from the same initialization as (b) yields (d).*

6.3. 3D views for room planner

Home furnishing companies have recently started to provide online tools that let users create arrangements of furniture customized for their own rooms, e.g. IKEA's Home Planner. After designing a room in this manner, users often want images of the room to share with others and to help them evaluate the design. Thus, another application of our system is to provide an automated solution for generating well composed images of user-designed rooms.

After generating the 3D arrangement of objects, the user selects a set of focus objects (likely the objects he is considering for purchase) and then asks our system to generate a composition. Unlike the previous two applications, we do not expect the user to provide an initial viewpoint for the scene. As a result, we modify our optimization to first search globally for the best camera parameters, which we then use as an initialization to our full optimization.

For our global camera search, we first generate a set of "plausible" initial camera parameters. We sample camera positions within the walls of the room at roughly 2 ft intervals and restrict the height to be at human eye level; we take 100 samples for the camera direction with a polar angle between $\pi/2$ (looking horizontally) and $\pi/2 + 0.28$ (looking slightly down); and we consider 4 uniformly spaced field-of-view values from 0.3 to 0.6. We then prune very poor samples based on the fractional image-space area of every focus object bounding box within the viewport (threshold = 50%). Next, we do k-means clustering (with $k = 4$) of the camera parameters. Within each cluster, we pick the camera with the lowest energy as the initial viewpoint and run our opti-

mization. Finally, we pick the optimized composition with the lowest energy as our result. Without parallelizing the optimization for different initial viewpoints, the entire process took 40 to 60 minutes in our experiments.

We used this optimization procedure to generate the results in Figure 5. Here, we chose the couch, coffee table and ottoman as the focus objects. In the final composition (Figure 5b), all of the focus objects are visible and the image provides a good overview of the scene from a plausible camera angle. For comparison, we show the result of a camera-only optimization in Figure 5c, which uses a different initial viewpoint than Figure 5b. In Figure 5d, we show the camera-only result generated from the same initial viewpoint as image (b). With the capability of moving objects, our full optimization is able to achieve a better balance between multiple factors, and achieve a better overall composition.

6.4. Object replacement

Multinational furniture companies like IKEA usually customize their catalog images for different countries to match cultural preferences. This customization often involves choosing different materials or replacing objects within a scene. In many cases, the size and shape of new objects can be significantly different from the original, which means that a stylist will have to spend a significant amount of additional effort adjusting the camera parameters and object positions to achieve a good composition for each customization. As with the previous applications, our optimization framework can automatically make these adjustments to reduce the amount of human effort required to perform these cultural customizations.

Figure 6 shows an example where we replace the grey seat, side table and coffee table. When we swap in the new objects, there is a collision between the chair and plant, and in general, the composition feels cramped. When we optimize the composition, the collision is resolved and the camera pulls back to keep all the relevant objects in the frame.
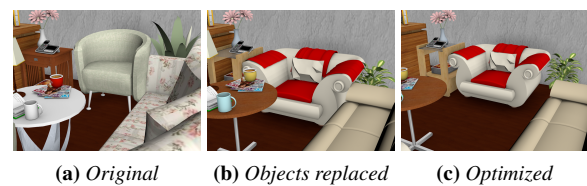
(a) *Original*   **(b)** *Objects replaced*   **(c)** *Optimized*

Figure 6: *Object replacement. From the original composition (a) the chair, side table and coffee table are replaced (b). Our optimization eliminates collisions and produces a better composition for these objects (c).*

6.5. Text-incorporated composition

Most catalog images have text overlays that describe the depicted scene. Such text is typically positioned over regions with nearly constant color so that it is easy to read and often appears in roughly the same location on every page (e.g.,

(a) *Input*   **(b)** *Ours*   **(c)** *Camera-only*

Figure 7: *Retargeting for different text layouts. The artist provides a rough position for the text box and specifies the champaign bottle and the goblet as focus objects. Our optimization adjusts object positions, viewpoint and text positions to increase contrast, reduce clutter and remove occlusion of the focus objects. By contrast, only optimizing the camera produces an inferior result. Note how the readability of the text is reduced due to the fruit bowl.*

corners) so that the viewer knows where to look to find textual information. Our optimization framework can automatically position text based on all of these criteria.

In addition to a set of focus objects and an initial composition, the stylist also specifies a set of rectangles **R** where she would like overlay text to appear in the frame. Our system treats each rectangle as just another object in the scene, but one that only has a 2D position and can only move within the viewport.

We apply the visibility and inertia terms to text rectangles as well. Specifically, the overlapping region of a focus object or a context object with a text rectangle is treated as occlusion.

We introduce two extra terms to make the text with constant color stand out. First, we use a contrast term E_{tc} to keep the background light or dark, so the text can be made a contrasting color. Second, we use a variance term E_{tv} to keep a low variance in luminance within each rectangle to reduce clutter behind overlaid text. The mathematical definitions of E_{tc} and E_{tv} are given in Appendix B.

Figure 7 presents a composition optimized with two different initial positions for the text. Notice how the objects in the scene are moved to create low contrast, low variance regions of the image where the text can be overlaid.

6.6. Retargeting for different aspect ratios

Our system can also automatically retarget catalog images to aspect ratios that are appropriate for different display formats. Simple cropping is usually not sufficient to create a good retargeted composition because the relative arrangement of objects in image space remains fixed. In contrast, our optimization framework has the ability to adjust camera parameters and object positions to produce good compositions for different aspect ratios. For this application, we use the viewpoint and object positions from the input composition as initialization and solve for a new image with the specified dimensions.

In Figure 8, we start with the optimal composition in one aspect ratio and then retarget it to another. For comparison, we generate images using our optimization method but without adjusting object positions. Notably, the greater flexibility creates better retargeted images.

(a) *Input*   **(b)** *Ours*   **(c)** *Camera-only*

Figure 8: *Retargeting for different aspect ratios (focus objects: champagne bottle and goblet). We start with the initial aspect ratio 1:2 (left), and retarget it to a different aspect ratio 4:3 (middle). We compare our result to the one where only the camera is optimized (right).*

7. Perceptual Study

A natural question to ask when considering our system is whether the additional freedom afforded by moving objects makes a positive impact on the results, or if – to the contrary – similarly good results could be obtained by performing a camera-only optimization. We investigated this question by asking people to compare 36 pairs of compositions created using our optimization procedure with object movement enabled (our method) and disabled (camera-only).

Our method   Camera Only

Figure 9: *A subset of the image pairs compared in our perceptual study. Our system is able to satisfy multiple composition constraints simultanuously, which cannot be achieved by changing viewpoint only. Note visibility in row 1 and clearance for milk carton in row 2.*

Study design: Our selection of scene compositions to compare includes all the examples shown in Section 6 of the paper, plus thirty detail images generated for different objects in three scenes (living room, study, and kitchen). For each of the three scenes, we selected – from among all the large furniture and a random subset of the smaller objects – the ten
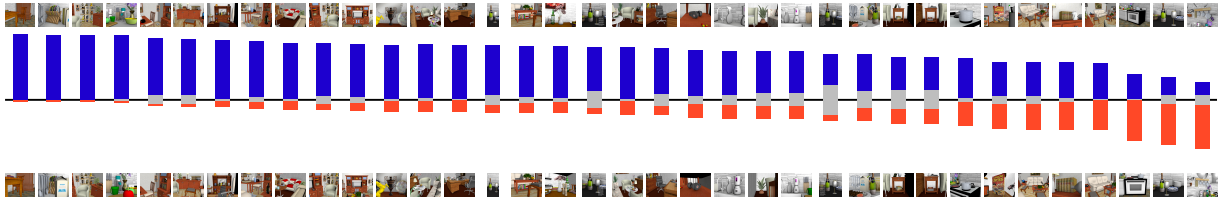
Figure 10: *Amazon Mechanical Turk Study results. We asked participants to compare 36 pair of images generated with full optimization (top) and camera-only optimization (bottom). Bars represent the proportion of participants who favored each image (dark blue: full, red: camera-only, light grey: no preference).*

objects where the detail image generated with camera-only optimization differed most from the full optimization.

We showed these pairs of scene compositions to study participants in randomized order, with images in each pair flipped left-right randomly. For each pair, the user selected a radio button to indicate that one composition was better at showcasing the specified focus objects (listed in the title), or that the two compositions are of the same quality (see a sample perceptual study web page in the supplemental material).

We administered the study to two groups: *experts* who work professionally on scene layout for catalog images, and *non-experts*.

The first group was recruited through personal contacts and completed a single-page web-based survey without compensation. Given the small number of experts in this field, we were only able to administer the survey to four participants.

The second group was recruited through Amazon Mechanical Turk, and each participant was compensated 10 cents. To exclude 'lazy' participants from our results, we tested the consistency of each participant's responses. Specifically, each participant completed a multiple-page (one comparison per page) survey, where each comparsion was asked twice, with compositions swapped left-right. We excluded any input from participants for each question where their two answers for the same pair were inconsistent, and we excluded all input from any participant whose answers were inconsistent for more than 25% of the questions. After running the study for 200 participants, these consistency checks yielded 49 to 75 answers per comparison.

Study results: Results of the expert study and the Amazon Mechanical Turk study are summarized in Table 1 and Figure 10 respectively. Generally, we find that full scene optimization is preferred to camera-only optimization. If the "no preference" answer is treated as half a vote for our optimization result and half a vote for camera-only, our optimization results received $\geq$ 75% of the votes in 20 cases, 50%-75% of the votes in 13 cases, and < 50% of the votes in 3 cases in the Mechanical Turk study. With a significance level of 0.05, our method is significantly preferred in 26 cases, while the camera-only optimization is significantly preferred in 2 cases. In the remaining 8 cases, the null hypothesis (i.e. no preference) cannot be rejected. Note that the

Holm-Bonferroni method was used in the analysis to control the familywise error rate in our experiments. According to comments provided by participants, the main benefits of our full optimization are that it moves objects to avoid unwanted occlusions, generates better contrast, and avoids awkward camera views in the final composition.

ID	Ours	Camera only	No preference
Expert 1	22	12	2
Expert 2	17	14	3
Expert 3	22	11	3
Expert 4	21	12	3

Table 1: *Expert study. Our method is preferred in general.*

8. Conclusion and Future Work

In this work, we have introduced a technique for optimizing 2D compositions of 3D scenes that adjusts camera parameters, object transformations, and surface materials. Our results and user evaluation show the benefits of optimizing over all of these scene parameters simultaneously. In particular, the comparisons between images generated by only adjusting the camera and those generated by our full optimization clearly indicate that moving objects and altering materials significantly improves the quality of compositions in many cases. We have demonstrated how our optimization framework benefits a variety of applications related to the creation of digital catalog images.

Our system has several limitations. First, the current implementation of our optimization procedure does not run at interactive rates. This is largely because speed has been sacrificed for flexibility in our prototype; we believe a production-oriented implementation could likely run orders of magnitude faster. In particular, the computation of partial derivatives, which is the bottleneck of the speed of our optimization, is highly parallelizable, because partial derivatives with respect to different free variables could be computed independently. Second, we use OpenGL rendering during our optimization, which does not account for global illumination effects that can impact the composition of the final image. Third, there may be additional composition rules that could improve the quality of the results. Early on in the project, we implemented energy terms for diagonal dominance, symmetry, and focusing with vanishing points, but found them less

useful in our target applications. Using our system to systematically investigate which energy terms are most effective for which applications would be an interesting topic of further study.

Given that companies are increasingly relying on computer-generated imagery for catalogs and other product advertisements, there are many opportunities for future work related to the automated generation of such images. For example, we imagine new advertising applications that choose furniture arrangements based on how a room will look from key viewpoints (e.g., the front door). Film, game, and real-estate companies could automatically optimize scenes for sequences of camera viewpoints (e.g, for movie shots or virtual tours). On-line advertisers could adapt product images to wide varieties (millions) of user preferences with automatically optimized aesthetics. We believe composition-aware scene modeling is a useful approach for all of these applications and as such represents a promising research direction for the computer graphics community.

References

[Arn88] ARNHEIM R.: *The Power of the Center*. University of California Press, 1988. 2, 3, 4

[Bar06] BARES W.: A photographic composition assistant for intelligent virtual 3d camera systems. In *Smart Graphics* (2006), Springer, pp. 172–183. 2, 3

[BDSG04] BYERS Z., DIXON M., SMART W., GRIMM C.: Say cheese! experiences with a robot photographer. *AI magazine 25*, 3 (2004), 37. 3

[BE04] BANERJEE S., EVANS B.: Unsupervised automation of photographic composition rules in digital still cameras. In *SPIE Conference on Sensors, Color, Cameras, and Systems for Digital Photography* (2004), vol. 5301, pp. 364–373. 3

[Bet56] BETHERS R.: *Composition in pictures*. Pitman Pub. Corp., 1956. 2

[BFH01] BELL B., FEINER S., HÖLLERER T.: View management for virtual and augmented reality. In *Proceedings of the 14th annual ACM symposium on User interface software and technology* (2001), ACM, pp. 101–110. 3

[BMBT00] BARES W., MCDERMOTT S., BOUDREAUX C., THAINIMIT S.: Virtual 3d camera composition from frame constraints. In *Proceedings of the eighth ACM international conference on Multimedia* (2000), ACM, pp. 177–186. 2, 3

[BS95] BUKOWSKI R., SÉQUIN C.: Object associations – a simple and practical approach to virtual 3d manipulation. In *ACM Symposium on Interactive 3D Graphics* (1995). 12

[BSS10] BHATTACHARYA S., SUKTHANKAR R., SHAH M.: A framework for photo-quality assessment and enhancement based on visual aesthetics. In *Proceedings of the international conference on Multimedia* (2010), ACM, pp. 271–280. 2

[BTBV99] BLANZ V., TARR M., BÜLTHOFF H., VETTER T.: What object attributes determine canonical views? *Perception-London 28*, 5 (1999), 575–600. 3, 12

[Cli73] CLIFTON J.: *The Eye of the Artist*. North Light Publishers, 1973. 2

[CON08] CHRISTIE M., OLIVIER P., NORMAND J.: Camera control in computer graphics. In *Computer Graphics Forum* (2008), vol. 27, Wiley Online Library, pp. 2197–2218. 2

[DJLW06] DATTA R., JOSHI D., LI J., WANG J.: Studying aesthetics in photographic images using a computational approach. *Computer Vision–ECCV 2006* (2006), 288–301. 2, 3

[DW10] DATTA R., WANG J.: Acquine: Aesthetic quality inference engine - real-time automatic rating of photo aesthetics. *Proceedings of the ACM International Conference on Multimedia Information Retrieval* (2010), 421–424. 2

[Ent12] ENTHED M.: 3D at IKEA. In *3D Modeling Standards*. SIGGRAPH Birds of a Feather, 2012. 1

[FRS*12] FISHER M., RITCHIE D., SAVVA M., FUNKHOUSER T., HANRAHAN P.: Example-based synthesis of 3d object arrangements. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia) 31*, 6 (2012). 2, 12

[FSH10] FISHER M., SAVVA M., HANRAHAN P.: Characterizing structural relationships in scenes using graph kernels. *ACM Transactions on Graphics (Proc. SIGGRAPH)* (2010). 12

[Gri01] GRIMM C.: Post-rendering composition for 3d scenes. *Eurographics short papers 20*, 3 (2001). 2

[GRLM03] GOVINDARAJU N. K., REDON S., LIN M. C., MANOCHA D.: Cullide: Interactive collision detection between complex models in large environments using graphics hardware. In *Proceedings of the ACM SIGGRAPH/EUROGRAPHICS conference on Graphics hardware* (2003), Eurographics Association, pp. 25–32. 11

[GRMS01] GOOCH B., REINHARD E., MOULDING C., SHIRLEY P.: Artistic composition for image creation. In *Rendering Techniques 2001: Proceedings of the Eurographics Workshop in London, United Kingdom, June 25-27, 2001* (2001), Springer Verlag Wien, p. 83. 2, 3, 12

[GT90] GRILL T. S. M.: *Photographic Composition*. Watson-Guptill, 1990. 2

[JWL12] JIN Y., WU Q., LIU L.: Aesthetic photo composition by optimal crop-and-warp. *Computers & Graphics* (2012). 2

[KHRO01] KOWALSKI M., HUGHES J., RUBIN C., OHYA J.: User-guided composition effects for art-based rendering. In *Proceedings of the 2001 symposium on Interactive 3D graphics* (2001), ACM, pp. 99–102. 4

[KMM08] KAO H., MA W., MING Q.: Esthetics-based quantitative analysis of photo composition. In *Pacific Graphics* (2008). 2

[Kra05] KRAGES B.: *Photography: The Art of Composition*. Allworth Press, 2005. 2

[LCWCO10] LIU L., CHEN R., WOLF L., COHEN-OR D.: Optimizing photo composition. In *Computer Graphics Forum* (2010), vol. 29, Wiley Online Library, pp. 469–478. 2, 3, 4

[LFN04] LOK S., FEINER S., NGAI G.: Evaluation of visual balance for automated layout. In *Proceedings of the 9th international conference on Intelligent user interfaces* (2004), ACM, pp. 101–108. 2, 4

[MB88] MARTINEZ B., BLOCK J.: *Visual forces: an introduction to design*. Prentice Hall, 1988. 2

[MSL*11] MERRELL P., SCHKUFZA E., LI Z., AGRAWALA M., KOLTUN V.: Interactive furniture layout using interior design guidelines. In *ACM Transactions on Graphics (TOG)* (2011), vol. 30, ACM, p. 87. 2, 12

[MSSH13] MAJEROWICZ L., SHAMIR A., SHEFFER A., HOOS H.: Filling your shelves: Synthesizing diverse style-preserving artifact arrangements. 2

[OHPL99] OLIVIER P., HALPER N., PICKERING J., LUNA P.: Visual composition as optimisation. In *AISB Symposium on AI*

and Creativity in Entertainment and Visual Art (1999), pp. 22–30. 2, 3

[SLF*11] SECORD A., LU J., FINKELSTEIN A., SINGH M., NEALEN A.: Perceptual models of viewpoint preference. *ACM Transactions on Graphics (TOG) 30*, 5 (2011), 109. 3

[Sou12] SOUTHERN A.: Real or rendered? how 3d imagery is changing the way you shop. *Technomy* (October 2012). 1

[Stu14] STUMP S.: Say it ain't so: Ikea reveals 75 *Today Home* (2014). http://www.today.com/home/say-it-aint-so-ikea-reveals-75-catalog-images-are-1D80127051. 1

[Tay38] TAYLOR E.: *The How and Why of Photographic Composition*. The Galleon Publishers, 1938. 2

[War03] WARD P.: *Picture Composition for Film and Television*. Focal Press, 2003. 3

[WL11] WONG L., LOW K.: Saliency retargeting: An approach to enhance image aesthetics. In *Applications of Computer Vision (WACV), 2011 IEEE Workshop on* (2011), IEEE, pp. 73–80. 4

[YYT*11] YU L., YEUNG S., TANG C., TERZOPOULOS D., CHAN T., OSHER S.: Make it home: automatic optimization of furniture arrangement. *ACM Trans. Graph 30*, 86 (2011), 1–86. 2

Appendix A: Definition of energy terms

Our energy function is the summation of the value of each energy term as introduced in Section 4.

$$E = E_{rt} + E_{ce} + E_{cl} + E_{sa} + \\ E_{sr} + E_{co} + E_{su} + E_{cv} + \\ E_{tv} + E_{vb} + E_{cc} + E_{ir} \quad (1)$$

In this section, we will provide the definition of each energy term. Symbols used in the definition can be found in Table 2. In the rest of this section, we use w to represent the weight of the term with the same subscripts.

- **Rule of thirds.**

$$E_{rt} = \frac{w_{rt}}{\mathcal{R}(F)^2} \sum_{O_i \in \mathbf{O_F}} \left(\frac{w^2}{h^2} d_h(\mathcal{C}_2(P_i))^2 + \frac{h^2}{w^2} d_v(\mathcal{C}_2(P_i)) \right)^2$$

where w and h are the width and height of the bounding box of P_i, respectively, and d_h and d_v stand for minimum distance to the closest horizontal and vertical third lines respectively.

- **Centeredness.**

$$E_{ce} = \frac{w_{ce}}{\mathcal{R}(F)^2} \sum_{O_i \in \mathbf{O_F}} d_2\left(\mathcal{C}_2(F), \mathcal{C}_2(P_i)\right)^2$$

- **Clearance.** For an object O_i, we consider its signed distance to the bounding circle of a focus object in image space (a negative distance means O_i is inside the bounding circle), and normalize the distance by the radius of the bounding circle in order to avoid favoring small objects. We use $r(O_i)$ to denote the minimum of all the distances,

$$r(O_i) = \min_{O_j \in \mathbf{O_F}} \frac{d_2(\mathcal{C}_2(P_i), \mathcal{C}_2(P_j)) - \mathcal{R}(P_j)}{\mathcal{R}(P_j)}$$

The clearance term is then defined as,

$$E_{cl} = \frac{w_{cl}}{N} \sum_{O_i \in \mathbf{O_F}} \left(e^{-\max\{0, r(O_i)\}^2} + \max\{0, -r(O_i)\}^2 \right)$$

F	The 2D image frame (viewport)
$\mathbf{O}$	The set of all objects
$\mathbf{O_F}/\mathbf{O_C}$	Focus/context objects
O_i	An object in set $\mathbf{O}$
P_i	O_i's projection into image space
V_i	the part of P_i visible to the camera
$\mathcal{V}(\cdot)$	Volume in scene space
$\mathcal{A}(\cdot)$	Area in image space
$\mathcal{B}(\cdot)$	Boundary contour in image space
$\mathcal{F}(\cdot)$	Projection onto XY plane in scene space
$\mathcal{R}(\cdot)$	Diagonal radius
$\mathcal{C}_2(\cdot)/\mathcal{C}_3(\cdot)$	Centroid in screen/scene space
$d_2(\cdot,\cdot)/d_3(\cdot,\cdot)$	Euclidean distance in screen/scene space
$c(\cdot,\cdot)$	Color difference in L*ab space

Table 2: *Symbols used in the energy function definition.*

where $N = |\mathbf{O} \setminus \mathbf{O_F}|$. The $\max\{0, -r(O_i)\}^2$ term handles the case where O_i is inside a focus object's bounding circle.

- **Object size.** To quantify the effect of object size on saliency, we introduce the following term:

$$\mathscr{S}_r(O_i) = \max\left\{0, r - \frac{\mathcal{A}(V_i)}{\mathcal{A}(F)}\right\}^2$$

where $A(V_i)$ is the area of the visible part V_i of O_i, r is the minimum required size of O_i. We describe how we choose r later.

- **Visibility.** To quantify the effect of visibility on saliency, we use the following term:

$$\mathscr{V}_r(O_i) = \max\left\{0, r - \frac{\mathcal{A}(V_i)}{\mathcal{A}(P_i)}\right\}^2 + \mathscr{D}(O_i)$$

where $A(P_i)$ is the total area of the object projection P_i assuming no occlusions by other objects or clipping by the image frame. $A(V_i)$ and $A(P_i)$ can be computed efficiently using hardware occlusion queries [GRLM03]. We introduce the term $\mathscr{D}(O_i)$ to encourage objects outside the frame to move towards the center of the frame:

$$\mathscr{D}(O_i) = \begin{cases} d_2(\mathcal{C}_2(F), \mathcal{C}_2(P_i))^2, & \text{if } \mathcal{A}(V_i) = 0 \\ 0, & \text{else} \end{cases}$$

Visibility and object size are both essential for focus objects. However, for context objects, we observe that there are two scenarios. If the context object is small compared to the focus object, visibility is important while its absolute size in the viewport is not (e.g. items on the dining table in Figure 3 left). On the other hand, if the context object is largely occluded, it must maintain some minimum size in the composition. To handle these cases, we compute both energies $\mathscr{V}_r$ and $\mathscr{S}_r$ for each context object and select the minimum. Thus, the complete form of our object saliency energy term is

$$E_{sa} = w_{sf} \sum_{O_i \in \mathbf{O_F}} \left(\mathscr{V}_{v_f}(O_i) + \mathscr{S}_{s_f}(O_i) \right) + \\ w_{sc} \sum_{O_i \in \mathbf{O_C}} \min\{\mathscr{V}_{v_c}(O_i), \mathscr{S}_{s_c}(O_i)\}$$

There are four parameters that reflect parameter r in the visibility term $\mathcal{V}_r$. In all experiments, we set $v_f = 100\%, s_f = 10\%, v_c = 80\%, s_c = 5\%$.

- **Semantic constraints.** We consider two types of constraints. The first type is defined for a single object class, which constrains the object to only transform in a subspace in 3D, e.g. a picture frame cannot rotate. We implemented it as hard constraints – i.e. the degrees of freedom are reduced in the optimization. The second type is defined for a pair of object classes, which constrains the change of the pairwise distance between two objects, e.g. the distance between a dining table and a dining chair cannot change drastically in the optimization. We implemented it as soft constraints using the method similar to [BS95, MSL*11]:

$$E_{sr} = w_{sr} \sum_{\{O_i, O_j\} \in \mathbf{C}} \sigma_{i,j} d_3(\mathcal{C}_3(O_i), T_i^{-1}(\mathcal{C}_3(O_j)))^2$$

where $\mathbf{C}$ is a set of constrained object pairs, and T_i^{-1} is the initial transformation from the scene space into the local coordinate frame of object O_i, and $\sigma_{i,j}$ controls how much the spatial relationship can change. $\sigma_{i,j}$ can be set by the stylist empirically in practice, and we set $\sigma_{i,j} = 1$ by default in all of the results shown in this paper.

- **Collision relationships.**

$$E_{co} = w_{co} \sum_{\substack{O_i \in \mathbf{O} \\ \mathcal{A}(V_i) > 0}} \sum_{\substack{O_j \in \mathbf{O} \\ \mathcal{A}(V_j) > 0}} \frac{\mathcal{V}(O_i \cap O_j)}{\mathcal{V}(O_i)}$$

- **Support relationships.** We infer the support relationships in the input scene layout by using the method similar to [FSH10]. Then we penalize placement of an object off its support object by measuring the fraction of its projected area outside its support surface [FRS*12]:

$$E_{su} = w_{su} \sum_{O_i \in \mathbf{O}} \left(1 - \frac{\mathcal{A}(\mathcal{F}(O_i) \cap \mathcal{F}(S_i))}{\mathcal{A}(\mathcal{F}(O_i))} \right)^2$$

where S_i is the object supporting object O_i.

- **Canonical views.** We manually defined a set of 1-4 canonical view directions for each object *class* [BTBV99, GRMS01], and then deviations from them are measured as:

$$E_{cv} = w_{cv} \min_i ||(\theta, \phi) - (\hat{\theta}_i, \hat{\phi}_i)||$$

where (θ, ϕ) is the view direction of the camera, and $\{(\hat{\theta}_i, \hat{\phi}_i)\}$ are canonical view directions for the object class.

- **Typical views.**

$$E_{tv} = w_{ch}(h - h_0)^2 + w_{ca}\phi^2$$

where h is the height of camera off the floor, $h_0 = 5ft$ is the typical height of a human eye, and ϕ is the pitch of the camera (where 0 is horizontal). This term penalizes viewpoints that deviate from a typical human eye height and tilt the camera upwards/downwards.

- **Visual balance.**

$$E_{vb} = \frac{w_{vb}}{\mathcal{R}(F)^2} d_2 \left(\mathcal{C}_2(F), \frac{\sum \mathcal{C}_2(P_i)\mathcal{A}(P_i)}{\sum \mathcal{A}(P_i)} \right)^2$$

We penalize by the distance between the frame centroid and the center of mass of all objects in the frame.

- **Color contrast.**

$$E_{cc} = \frac{w_{cc}}{\mathcal{R}(F)^2} \sum_{O_i \in \mathbf{O_F}} \sum_{p \in \mathcal{B}(V_i)} \frac{1}{\left(\text{avg}_{q \in N(p) \setminus V_i} c(p, q) \right)^2 + \varepsilon}$$

where $N(p) \setminus V_i$ denotes the neighborhood of pixel p, excluding the visible pixels in V_i.

- **Regularization.**

$$E_{ir} = w_{ir} \sum_{O_i \in \mathbf{O}} \left(\frac{x_i^2}{\sigma_t^2} + \frac{y_i^2}{\sigma_t^2} + \frac{\theta_i^2}{\sigma_r^2} \right) + \sum_{i=0}^{5} \frac{c_i^2}{\sigma_c[i]^2}$$

where (x_i, y_i, θ_i) describe the translation and rotation of object O_i, respectively, and c_i describe the change to camera parameters, with $\sigma_t = 0.5, \sigma_r = 0.5, \sigma_c = [0.17, 0.17, 20, 20, 20, 0.17]$ controlling the flexibility of object movement and camera manipulation.

These energy terms are weighted by coefficients that adjust for scale differences and control their effects on the final results. By default, the weights are set to $w_{rt} = 10000$, $w_{ce} = 10000$, $w_{cl} = 500$, $w_{sf} = 10000$, $w_{sc} = 500$, $w_{sr} = 100$, $w_{co} = 10000$, $w_{su} = 10000$, $w_{cv} = 10000$, $w_{ch} = 10000$, $w_{ca} = 10000$, $w_{vb} = 20000$, $w_{cc} = 1.0$, and $w_{ir} = 1.0$. These weightings were determined empirically and are kept the same for all examples in this paper, except that $w_{ce} = w_{cv} = 0$ for overview images of scenes (e.g., session A in Figure 2, Figure 5, 7 and 8), and $w_{rt} = w_{ch} = w_{ca} = 0$ for zoomed-in images of specific objects (e.g., session B in Figure 2, Figure 4 and 6). *It is not expected that a user has to tweak these weights to get good results for specific scenes.*

Appendix B: Terms for text-incorporated composition

In this section, we define the two extra terms used in text-incorporated composition.

- **Text contrast.**

$$E_{tc} = w_{tc} \sum_{R_i \in \mathbf{R}} \min_{wb=w,b} \frac{1}{\mathcal{A}(R_i)} \sum_{p \in R} \frac{1}{d_l(p, wb)^2 + \varepsilon}$$

where p is a pixel in the rectangle R_i, $d_l(\cdot, \cdot)$ is the difference between the luminance of two pixels, w is white and b is black, $w_{tc} = 2.0$.

- **Text variance.**

$$\bar{L_{R_i}} = \frac{1}{\mathcal{A}(R_i)} \sum_{p \in R_i} L(p)$$

$$E_{tv} = w_{tv} \sum_{R_i \in \mathbf{R}} \frac{1}{\mathcal{A}(R_i)} \sum_{p \in R_i} \left(L(p) - \bar{L_{R_i}} \right)^2$$

where $w_{tv} = 10000$.