SIRfyN: Single Image Relighting from your Neighbors

D.A. Forsyth, Anand Bhattad, Pranav Asthana, Yuanyi Zhong, Yuxiong Wang.
UIUC, Urbana, Illinois

Abstract—We show how to relight a scene, depicted in a single image, such that (a) the overall shading has changed and (b) the resulting image looks like a natural image of that scene. Applications for such a procedure include generating training data and building authoring environments. Naive methods for doing this fail. One reason is that shading and albedo are quite strongly related; for example, sharp boundaries in shading tend to appear at depth discontinuities, which usually apparent in albedo. The same scene can be lit in different ways, and established theory shows the different lightings form a cone (the illumination cone). Novel theory shows that one can use similar scenes to estimate the different lightings that apply to a given scene, with bounded expected error. Our method exploits this theory to estimate a representation of the available lighting fields in the form of imputed generators of the illumination cone. Our procedure does not require expensive “inverse graphics” datasets, and sees no ground truth data of any kind.

Qualitative evaluation suggests the method can erase and restore soft indoor shadows, and can “steer” light around a scene. We offer a summary quantitative evaluation of the method with a novel application of the FID. An extension of the FID allows per-generated-image evaluation. Furthermore, we offer qualitative evaluation with a user study, and show that our method produces images that can successfully be used for data augmentation.

1 INTRODUCTION

When someone switches on a light, the image at the back of your eye changes even when there is no other change. The new shading of the scene in front of you is determined by the shape and material composition of the scene, and by the light source. This paper shows how to relight a scene – we demonstrate a model that can generate images of the scene with random (but plausible) changes in light source. One might expect scene relighting to be easy – decompose the scene into albedo and shading, and replace the shading with another shading field. This approach doesn’t work because shading fields are very strongly related to albedo fields and to shape. For example, sharp boundaries in shading tend to appear at depth discontinuities, which are usually apparent in albedo. This means that it is hard to produce shading fields that apply to a particular scene with elementary methods.

An alternative strategy is to use “inverse graphics” – learn to regress geometry, material, etc. against an image using training set that consists of ground truth geometry, material etc. together with computer generated images (CGI). While this strategy may offer an approach to building some kinds of vision system, it is laborious (one must produce the training data) and is not yet reliable. Furthermore, the inverse graphics strategy is intellectually unsatisfying, because it doesn’t explain how a visual agent learns to interpret the rich visual world they see if the agent doesn’t have a CGI training set conveniently to hand.

Instead, we see scene relighting as an unsupervised problem. We train a neural network to produce a set of per-pixel scalings for an image. Each scaling produces a new image under some, likely extreme, illuminant (and, implicitly, a new shading field). A non-negative linear combination of the scalings produces other relighting. Two novel theorems establish that it is possible to estimate these likely relightings from observations of other, similar, scenes. These theorems motivate using a loss that measures how well the predicted shading fields represent shading fields from other, similar scenes. Extensive evaluation shows that (a) our relighted images are diverse and realistic; (b) rejection sampling methods are available that can produce sharp improvements.
in quality; (c) our predicted shading fields are consistent with observed shading fields; and (d) our strongest method can be used successfully as a data augmentation.

**Contributions:** We offer the first scene relighting method that requires no data marked up in any form. We offer a novel body of theory showing how perturbations of geometry, albedo and luminaire (a luminaire is an object that emits, rather than simply reflects, light) affect scene radiosity, and that lightings of similar scenes can be used to infer lighting of a given scene. We show how to evaluate our models quantitatively as well as qualitatively. We introduce a variant of the FID that applies to image transformations and gives a per-image estimate of quality.

### 2 RELATED WORK

**Intrinsic image decomposition** Decomposition into albedo and shading is well established; evaluation is by comparing predicted lightness differences to human labelled versions, the WHHDR score. We use the method of [1] (see that paper for a detailed review of an area whose history dates to the mid 19th century), which requires no labelled training data or CGI and produces a WHHDR of 16.9%, current SOTA for such a method.

**Inverse rendering** seeks to recover all required data for rendering from an image. Reconstruction methods applied to multiple images are one approach (eg [2]); another is to impose strong parametric models [3]. Alternatively, one could train a regressor using a fixed object class (faces [4]; furniture [5]; birds [6]), CGI of general shapes [7], [8], or multiple images [9]. Very strong reconstructions can be obtained from rendered images [10]; isolated object reconstructions can be relit [11]. Inverse rendering recovers strong scene models for indoor scenes from CGI [12], but we are aware of no successful attempt to change luminaires in images of scenes by relighting an inverse rendering result. In contrast to inverse rendering methods, we use no CGI and no marked up data.

**Precomputed radiance transfer** (PRT) is a method from computer graphics that builds a relightable representation of a scene by rendering versions with multiple distinct illuminants, producing an estimate of the light transport matrix mapping illuminant representation to relit scene (originally [13]; review in [14]; compare a PRT operator with our EGM, below). The operator can be estimated for real scenes with a projector and a large number of images (a generalization of [15]) or from a collection of outdoor images [16]. There are strong regularizers which can massively reduce the amount of scene data needed (eg [17], [18]). Given relatively few lightings of a scene, a neural network can produce a very good light transport matrix estimate [19], [20]; this is consistent with theorems 1 and 2. In contrast to PRT methods, our estimate requires only one image of a given scene, but must be a great deal less accurate.

**Image relighting** is now an established task. For scenes, there are workshop tracks (eg [21], [22]), challenges [23], [24] and datasets [25], [26]. Existing work learns image mappings (pure image mappings as in [27]; depth guided, as in [28]; using wavelets, as in [29]; shadow priors, as in [30]). In all cases, methods are learned with paired data (ie images of the same scene under different illuminations), available in the VIDIT dataset [25] and the MIE dataset [26]. VIDIT data is CGI, and emphasizes point light sources with strong shadows, which are uncommon in indoor scenes. Pairing is necessary to ensure that their method preserves scene characteristics [31], [32]. Chogovadze et al use a multi image dataset and lightprobes to learn a scene relighter, and show the resulting augmentations improve two patch matching tasks [33]. Philip et al learn geometry aware models to relight outdoor video from renderings under multiple light conditions [34]. Similarly, face image relighting requires multi-illumination data (from real light stages [35]; made using deformable models [36]). In contrast to these methods, we do not use paired data.

**Shadow synthesis** is now an established task. People in video can be used as scene probes to learn to create shadows for objects inserted in outdoor scenes [37]; methods with sufficient training data can learn to create soft attached shadows for objects that have been inserted into indoor scenes by relighting an inverse rendering result.
At run time, we want to be able to produce multiple realistic 
scenes [38], [39]. In contrast to these methods, we use no 
paired or synthesized data, and we create entire lighting 
fields.

Global illumination bounds: Section 4 establishes that 
scenes with similar albedo, luminaire and geometry have 
similar radiosity. Bounds on the effect of luminaire difference 
are well known (eg [40], sec 5.3, [41]). Previous results on the 
effects of albedo changes are rare ([42] sketch the case for an 
inTEGRATING sphere; [43], p175, has a bound). There are few 
results on the effect of perturbing geometry; [44] estimate the 
effect of flat ports on an integrating sphere; [42] argue that 
sufficiently small deformations of an integrating sphere do 
not affect its function; and [43], p176, says such results are 
hard to get.

The illumination cone and its generators (ICGs): The 
family of images created by lighting a single convex diffuse 
object, viewed in a fixed camera, with a point source is 
naturally a convex cone (reflection is linear; irradiance is 
non-negative). As Kriegman and Belhumeur [15] point out, 
this fact extends to any diffuse scene. If the observed scene 
has only a discrete set of normals – for example, a polyhedral 
object – then the cone is polyhedral and has a finite set of 
generators (i.e. any element of the cone is a non-negative sum 
of a finite set of generators). In simple cases, one can recover 
generators by illuminating the object with different sources. 
Interreflection and shadowing may mean that the cone is 
not polyhedral, but experiments suggest that (a) relatively 
few generators provide a good representation of the cone 
anyhow ([15], p8; [45], p644) (b) the cone is relatively “flat” 
([15], conjecture 1; p12; for convex objects [46]) and (c) 
illumination cone methods offer strong models of extreme 
relighting for face images [45].

3 SIRfYn: METHODS AND NETWORK

At run time, we want to be able to produce multiple realistic 
relightings from a single image of a scene, without requiring 
intrinsic image decomposition etc. We will build a network 
that maps a 256x256x3 color image \( I_V \) of a scene \( V \) to 
a set of \( N_g \) multipliers \( M_i(I_V) \) (256x256x1 scalar fields). For 
non-negative \( \theta \) and \( \theta \)-broadcast element-wise multiplication, 
these multipliers will have the property that \( M(\theta) \cdot I_V = 
\left[ \sum_i M_i(I_V) \theta \right] \cdot I_V \) is a relighted version of \( I_V \). One can 
think of these multipliers in terms of shading fields, by 
modelling the image as a product of albedo and shading 
\( I_V = \alpha V \cdot S_V \); then relighting \( I_V \) by \( M(\theta) \) implies a new 
shading field \( S_V \left[ \sum_i M_i(\theta) \right] \). So predicting \( M(\theta) \) from \( I_V \) 
yields an encoding of the shading fields we expect for that 
scene – equivalently, estimates of generators for the scene’s 
illumination cone.

Network: We predict multipliers using a network to obtain 
\( \left[ M_1(I), \ldots, M_{N_g}(I) \right] = N(I; \phi) \). We implement 
\( N(I; \phi) \) using a u-net, with skip connections (details in 
supplementary; \( \phi \) parameters). The output layer forms a 
multiplier directly from activations, with housekeeping 
losses to ensure that multipliers are non-negative.

The key trick is obtaining losses to train such a network. 
Section 4 sketches two novel theorems we rely on (proved 
in detail in supplementary). Theorem 1 states that, for 
two scenes \( V \) and \( V' \) with similar geometry, illumination 
and albedo, the difference between radiosity \( B_V \) and \( B'_{V} \) 
is bounded (despite the effects of interreflections). Now 
radiosity is linear in illumination and images are linear in 
radiosity, so we can model different relighted images of 
the same scene with some linear operator, which is easily 
estimated if we have multiple images of the same scene under 
different lights. We don’t have such images, but Theorem 
2 states that the expected future error of using images of 
similar scenes in place of images of the original scene is 
bounded.

We can now test whether the network is producing 
sensible multipliers with theorem 2. Theorem 2 means 
that, for the shading field \( S_V \) of each image of a similar 
scene \( V' \), there should be some non-negative \( \theta_{V'} \) so that 
\( S_V' \approx S_V \left[ \sum_i M_i(I_V) \theta_{V'j} \right] \).

Finding images of similar scenes: It is not practical to 
enforce the exact constraints of section 4, but the intention 
of these constraints are that a nearby image should depict 
a scene of similar shape to the original scene. GIST features 
were hand-tuned to encode the overall shape of a scene [47], 
and we use the 20 nearest neighbors in GIST features. 
We have not attempted to optimize further, but informal 
experiments with learned embeddings were not successful, 
and we expect that incorporating an albedo matching term 
might help.

Exploiting images of similar scenes: Assume we have 
k images which can be used to estimate generators for 
an image \( I_V \). There should be some non-negative \( w \) such 
that the shading field \( S_i \) of the \( i \)’th such image is encoded by 
the multipliers, so \( S_i \approx S_V \left[ \sum_j M_j w_j \right] \). Geometrically, 
we require that each \( S_i \) lie inside the cone generated by 
\( \left[ S_V M_1, \ldots, S_V M_{N_g} \right] \). There are some subtleties involved in 
deriving a loss from this constraint. We do not want \( 256 \) to be 
“further outside” the cone than \( S_i \) is, so the loss needs to be 
independent of the scale of the shading field. We achieve this 
by rescaling all shading fields to have mean 0.7 (the value 
is arbitrary, and affects only the scale of the loss). The nearby 
loss on multipliers can be written as 
\[ L_N(I_V, \phi) = \sum_{i \in N} C(S_i, M(I_i)) \]

for \( N \) the neighbors of \( I_V \) and 
\[ C(S_i, M(I_V)) = \min_{w|w > 0} \left| S_i - \sum_j S_V M_j w_j \right|_2^2 \]

but this is difficult to evaluate (the non-negativity constraint 
means the value is a solution to a quadratic program). We 
estimate \( C(S_i, M(I_V)) \) by solving for \( w \) using least squares; 
clipping to ensure \( w \geq 0 \); then taking a \( N_g \) steps of projected 
gradient descent in \( \theta \). This yields a usable approximate loss.

Realistic scene preserving shading: It is important that, 
for predicted multiplier \( M(\theta) \), the image \( M(\theta) I_V \) should 
look like a relit version of \( I_V \). We impose this constraint 
using an adversarial loss to adjust the joint distribution of 
(predicted shading, predicted image) to match the best 
estimate of the true distribution. Here \( L_R(I_V; \phi) \) is the 
value of an adversary that compares generated tuples 
\( M(\theta) S_V, M(\theta) I_V \) to real tuples \( (S_i, I) \) and \( S_i \) is the 
shading field of image \( I \), estimated using the method of [1]. 
The discriminator score (we use a hinge loss, as in [48]; details 
in supplementary) depends only on local neighborhoods, as
in PatchGAN [49], and the size of the neighborhood has important effects on results.

**Extrapolation with a barrier:** If nearby images for a given image have relatively similar shading fields, the nearby loss will not encourage diversity because it does not force the $M_{ij}$ to be different. But we want to extrapolate and force the new shadings to be as extreme as possible. Write $\text{int}$ for the operator that computes the intensity of an image, $\mathcal{T}$ for the pixelwise mean of an image, $P_V = \text{int}(M_V I_V)$ and $Q_V = \text{int}(I_V)$. We use

$$L_B(I_V; \phi) = \mathcal{T} \log \left[ \frac{P_V}{Q_V} - \frac{Q_V}{P_V} \right]$$

as a diversity loss; this forces multipliers to be different from 1, but may result in unrealistic relightings if overweighted.

**Realism housekeeping:** We require that any pixel in $R$ should be non-negative ($L_U(R)$), and less than one ($L_O(R)$), and achieve this with simple one-sided quadratic losses. Note that this constraint applies to relighted images, rather than shading fields – applying the constraint to shading fields will lead to regions with dark albedo being consistently dark. We show the network $A$ as well as $I$ so that it can anticipate this loss and produce very bright shading at locations where the albedo is dark.

**Overall loss:** the loss is then:

$$\mathcal{L} = \mathcal{L}_N + \lambda_B \mathcal{L}_B + \lambda_O \mathcal{L}_O + \lambda_R \mathcal{L}_R + \lambda_U \mathcal{L}_U$$

### 4 Theory: ICGs and EGMs

**Light in diffuse scenes:** The brightness of each pixel in the camera is determined by the radiosity at the corresponding scene point. Write $\rho(u)$ for albedo at the point on the geometry parametrized by $u$, $E(u)$ for the diffuse emittance of that point (only non-zero when the point is actively emitting light, so a point on a luminaire with non-zero weight), and $B(u)$ for the radiosity at $u$. Write $\mathbf{D}_\rho$ for the linear operator that maps $f(x)$ to $(\rho f)(x)$. Then standard diffuse interreflection theory (see, eg [50], [51], [52], [53]) gives $B = E + \mathbf{D}_\rho \mathbf{K} B$, equivalently $(I - \mathbf{D}_\rho \mathbf{K}) B = \mathbf{R} B = E$ where $\mathbf{K}$ is a bounded linear integral operator. This yields a Neumann series solution $B = (\mathbf{R}^{-1} E = E + \mathbf{D}_\rho \mathbf{KE} + (\mathbf{D}_\rho \mathbf{K})^2 E + \ldots)$ which should be interpreted as “emittance+direct term+one bounce+...”: for most practical geometries and albedos, relatively few terms at relatively poor spatial resolution yield an acceptable solution (eg [54], reviews in [41], [55]).

**ICGs from diffuse interreflection theory:** Assume that a scene has $N_e$ luminaires, so that $E = \sum_i \theta_i E_i$ with $\theta_i \geq 0$ the intensity of the luminaires. We choose $\theta_i \geq 0$ the intensity of the luminaires. We choose $\|E_i\|_2^2 = 1$. Because the solution is linear in the emittance, we can write $B(u; \theta) = \mathbf{R}^{-1} \sum_i \theta_i E_i = \sum_i \theta_i \mathbf{R}^{-1} E_i = \sum_i \theta_i B_i(u)$ if we know the solution for each luminaire separately, we can construct a solution for any linear combination of luminaries. If we have one image of the scene for each luminaire, where the image is illuminated by that luminaire alone, we can generate an image from this viewpoint for any combination of the luminaires. But we do not have these images.

**Similar scenes with similar luminaires have similar radiosity:** We assume that all relevant reflection is diffuse; it is likely that this can be generalized (discussion). A scene $V$ is a tuple of geometry, albedo and luminaire model; write $V = (\mathcal{G}, \rho, \mathcal{E})$, where $\mathcal{E}$ is the luminaire model. The geometry $\mathcal{G}$ consists of a set of surfaces parametrized in some way and a 2D domain for that parametrization; write $\mathcal{G} = (s(x), D)$. The luminaire model captures the idea that lighting in a particular scene may change. We have $\mathcal{E} = (E_1, \ldots, E_N, P(\theta))$ for a set of potentially many basis luminaires and a probability distribution over the coefficients. This information implies a probability distribution over luminaires for the scene $P(\mathcal{E})$.

**Similarity:** Two scenes $V$ and $V'$ are similar if: there is some affine transformation so that $g'(\mathcal{G}, A, b) = (A s(x) + b, D)$. $E_i'(x) = E_i(x)$; and $P_V(E)$ and $P_V'(E)$ the same. Consider some lighting $E_V$ of $V$ and a different lighting $E'_V$ of $V'$: Theorem 1 (below) establishes that if $V$ and $V'$ are similar, and $E_V$ and $E'_V$ are similar, then $B_V$ will be close to $B'_{V'}$.

**Theorem 1:** For $V = (\mathcal{G}, \rho, \mathcal{E})$ and $V' = (\mathcal{G}', \rho, \mathcal{E}')$, where $E_\rho \mathcal{E} = \sup_k \mathcal{E}_k \mathbb{E}|(\rho - \rho')|$, $p = \sup_{\mathcal{E}'} p_k$, $p' = \sup_{\mathcal{E}'} p_k'$, $c$ is the condition number of $\mathcal{A}$ (ratio of largest to smallest eigenvalues), we have:

$$\|B_V - B'_{V'}\| \leq c_1 (E_\rho \mathcal{E}, p, p', c)$$

**Proof:** Elaborate, relegated to supplementary, which gives the form of $c_1$.

**The effective generator matrix (EGM) of a scene:** Choose some $N_r$ element orthonormal basis to represent the possible radiosity functions for the scene (if $N_r = N_e$, the representation is exact, because there is a finite dimensional space of luminaires). In this basis, any particular radiosity $B(x)$ has coefficient vector $b$. Write $b_i$ for the coefficient vector representing $B_i(x)$; and $B = [b_1, \ldots, b_{N_r}]$. For any $r \leq N_r$, an effective generator matrix for scene $V$ is an $N_r \times r$ orthonormal matrix $M_V$ such that for any illumination condition producing radiosity represented by $b$, there is some $w$ such that the radiosity represented by $M_V w$ is similar to the actual radiosity field. In particular, we want

$$\mathcal{L}(M_V) = \mathbb{E}_{\theta \sim P(\theta)} \left[ \inf_w \| M_V w - B \theta \|_2^2 \right]$$

to be small; $M_V$ is clearly not unique. Now if we have $k$ radiosity fields for a scene under different, unknown, illumination conditions, for $k$ very large, where for each $\theta_i \sim P(\theta)$ (which is not known). We could estimate an EGM by replacing the expectation with an average, and minimizing. But we do not have $k$ images of the same scene.

**Estimating an EGM:** An EGM can be estimated from radiosities obtained from similar scenes. Write $\mathcal{L}_T(M)$ for the true expected error of using an EGM $M$ to represent the radiosity of a scene. Theorem 2 shows that substituting an estimate $\hat{M}$ obtained by using $k$ radiosities in total, taken from distinct similar scenes, for the best (but unknown) effective generator matrix $M$ incurs bounded error. This means we can use estimates from similar scenes with confidence, if the scenes are similar enough.

**Theorem 2:** $E_{\mathcal{L}_T(M)} - E_{\mathcal{L}_T(M)}$ is bounded.

**Proof:** Elaborate; relegated to supplementary, which provides the bound.

**Image loss from theory:** Note first that theorem 2 applies to images as well as to scenes. If one views a scene through a fixed camera, the camera is a linear mapping (taking scene radiances to pixel values) and it is bounded. The result is very powerful, because it says that an estimate of the generator matrix for a given image can be obtained from
enough images of similar scenes — take the shading fields $S_i$ of similar scenes and form a $M$ that minimizes

$$\sum_i \inf_w | M w - S_i |_2.$$ 

Using this approach directly is inconvenient because it means we need to find similar images at inference time. Instead, we derive a loss from Theorem 2 and use a network to predict the generator matrix.

5 Quantitative Evaluation for Relighting

Realism in current image generators is typically measured with the FID [56]. One embeds set of generated images, then computes the Frechet distance between the resulting points and an embedding of a set of real images. We use the standard VGG embedding of [56], with dimension $d_f = 2048$. This is an estimate of the difference between the distributions from which the samples were taken; smaller values are better. As [57] show, the FID estimated from a finite set of points is biased by a $1/N$ term that depends on details of the generator, but straightforward extrapolation methods remove this bias successfully to produce $\text{FID}_\infty$. An important difference between image relighting and image generation is that relighting should map the collection of real images to itself, preserving frequencies etc. (loosely, relighting is an isomorphism). In turn, this means that the FID between a set of images $O$ and a relighted set $R$ of images should be zero, so $\text{FID}_\infty(O, R)$ should be zero up to sampling error. Note strong realism scores can be obtained by not changing the original image much.

Diversity: We must also evaluate whether relighted versions of an image are actually different from the original image. We measure diversity by comparing relighted images to the original. We rescale the relit image to have the same mean as the original image, so that just scaling images will get zero diversity score. Write $N_i$ for the total number of pixels in relit images; we score diversity with $\text{MSD}$, computed as

$$\text{MSD}_\theta = \sqrt{\frac{1}{N_t} \sum_{i \in \text{images}} D_i^2},$$

where $D_i = (I_i - R_i(\bar{I}_i/\bar{R}_i))$. Note quite unreasonable shadings may have high diversity scores.

Local FID: Because relighting is an isomorphism, we can recover a per-image measure of quality. Assume we obtain $R$ from $O$ by relighting a single image $I_k$ to obtain $R_k$ and compute $\text{FID}(O, R)$. This value — which is non-negative, and is ideally zero — measures the extent to which the relighted image is realistic. Furthermore, different relightings of the same image will have different values. This value is the local FID of image $R_k$. This local FID strongly reflects human preferences (Figure 4; people reliably prefer images with smaller local FID). Computing the local FID requires care; the supplementary shows how.

6 Experimental Details

Datasets: No markup is required. Training albedos are obtained with the method of [1]. We train with images from the LSUN Bedroom dataset ([58]), resized as described. We use 20,000 images to train the generator (augmented by crop and resize and horizontal flips) and train the discriminator with 20,000 distinct images (to prevent shading “leaking”). Nearest neighbors are drawn from 100,000 LSUN Bedroom images, and are augmented in the same way (so when a training image is cropped, the same crop is applied to each nearest neighbor). We use 20 nearest neighbors for each image to compute $\mathcal{L}_N$.

Relighting weights in training: During training we must ensure that any non-negative combination of multipliers results in a realistic image; we do so by randomly selecting one multiplier per image in a batch, and passing those to the adversary.

Choice of hyperparameters: For model search purposes, we train our network at 128x128. Simple qualitative experiments suggest the model is not particularly sensitive to the choice of $\lambda_R$, $\lambda_U$ and $\lambda_O$, and we have used 1, 100, 100. Simple qualitative experiments show that training without projected gradient descent ($N_{gd} = 0$) fails reliably; more than one step ($N_{gd} > 1$) slows training without offering any obvious benefit. Other hyperparameters that have significant effects are: $N_g$, $\lambda_P$, $\lambda_B$ and the scale of the discriminator. Figure 4 plots $\text{FID}_\infty$ against $\text{MSD}$ for a variety of methods at three different values of $\alpha$ (FID vs MSD and detailed comparisons in supplementary). In summary, we have: a small scale discriminator is better; $\lambda_B = 5$ is about right; there is a bias-variance tradeoff for $N_g$, and $N_g = 10$ is about right (cf [46]).

Inference and equivariance: Our chosen network is trained on 256x256 images. The model should be sensitive to geometry (and is - supplementary), so resizing an image that isn’t square should be dangerous (and is - supplementary), and so we always work with 256x256 images, and crop (rather than resize) as required. We model the weights as distributed according to a symmetric Dirichlet distribution with concentration parameter $\alpha$. Then we generate relit images by: drawing a sample of weights; computing multipliers; computing a weighted combination of multipliers; and multiplying by the image.

Baselines and Controls: There are no meaningfully comparable methods. However, sampling error, etc., means it is important to have baselines for “low” and “high” values. We use two baselines. Spl: we evaluate sampling variance in $\text{FID}_\infty$ by computing the $\text{FID}_\infty$ between random splits $O_i$ of the image set, and obtain $F_{\text{spl}} = \text{FID}_\infty(O_i, O_j)$. This yields a “low” FID estimate. 1D: to exclude the possibility that our relighting method simply maps pixel values, we train a relighting method using only 1x1 convolutions.

7 Results

Qualitative analysis: Figure 10 shows five random relightings of five scenes, together with the relevant multipliers. These random relights, not curated, show: shadows removed or replaced; significant brightening or darkening of surfaces, registered to geometry; suppression or enhancement of gloss effects; and suppression or enhancement of luminaires. The model is clearly capable of substantial, non-obvious relightings that look realistic, and can clearly cope with a very diverse range of surfaces. Relightings include overall
Fig. 3. Images relighted at \( \alpha = 0.1 \) (more variable; left; cf \( \alpha = 0.5 \) less variable in supplementary) and the first five multiplier fields for each right. Top row: shows original images; the following five rows are random relights, not curated. A number of important effects are notable: green shadow removed or replaced; red significant brightening or darkening of one surface, registered to geometry, yellow suppression or enhancement of a gloss effect; blue suppression or enhancement of a luminaire. All relights are shown at the same mean intensity as the original image, to ensure that effects are not masked by overall shift in brightness. Multiplier fields are scaled so that one is mid-grey, less than one darker, greater than one brighter. Recall relights are combinations of multipliers, with coefficients drawn as in Section 3.
changes of brightness, but are not confined to them. The model appears to have some form of geometric knowledge (it often changes the lighting of horizontal and vertical surfaces differently).

**The uses of LFID:** If relighting is to be used for data augmentation, one seeks relightings that make large changes to a scene. Figure 4 shows the results of a rejection sampling method that seeks the best LFID for MSD over a threshold; and the largest MSD for LFID below a threshold.

**Relighting is a successful data augmentation:** Chogovadze et al show relighting augmentations improve two patch matching tasks [33]. Table 1 shows that our relighting augmentations improve performance on a classification task, with greater impact when the dataset is small (to be expected – relighted images contain information from nearest neighbors, redundant when there are enough images). Improvements for the “*room” categories exceed those for others, suggesting the augmentation – which is learned on bedrooms – generalizes, but not to arbitrary scenes. Table 2 shows that our relighting augmentations improve performance on a self-supervised learning task. Note that, while our relighter sees only bedroom images, table 1 suggests strongly that it produces acceptable relightings for other scenes, likely a consequence of the common geometry of rooms (see also supplementary).
### TABLE 2

| Augs | Top-1 acc | Top-1 acc relight |
|------|-----------|------------------|
| A    | 16.1      | 16.3             |
| B    | 16.1      | 16.5             |
| C    | 14.2      | 15.2             |

Relighting augmentations are helpful in a self-supervised learning task. We train a feature stack using SIMCLR [61] with three different augmentation conditions: A - original from the paper; B - without left/right flip; C crop with scale in [0.36, 1]. For relight, we use pairs of images with distinct relights. We train on 10K LSUN bedroom images, freeze, then train a linear readout on places365 36K; then evaluate. Note that using relighted images reliably offers a small gain.

**Fig. 6.** Rough shape – rather than scene label – seems to be key to this method’s success. Left: images (top) from Places365 categories botanical_garden and catacomb and relightings (bottom); because the rough shape is like that of a bedroom, the relightings are quite good. Right: images from ocean; relightings are inaccurate, likely because the model cannot deal with this class of spatial layout; note how the relightings echo the geometry of a room.

### 8 DISCUSSION

Further work will likely include: using these results to regularize PRT estimation; investigating the effects of architectural choice on the network; and reestimating nearest neighbors with generator estimates. **Limitations:** Some relightings are clearly not present, and generalization to other scenes is variable (Figure 6). We have no procedure for improving the model, save collecting more (unlabelled) data. **Negative social implications:** Convincing fake images can pollute (say) the news media; but our method offers only relighting, rather than the much more dangerous insertion or deletion editing. Experiments required approx 250 hours of time on a single GTX3090 from method inception to submission. The datasets used either do not contain, or are not known to contain, personally identifiable or offensive material. A total of $150 was paid to Amazon mechanical Turk workers for image labelling; no personally identifiable information was collected.

### APPENDIX I: FURTHER VARIANTS AND FULLER COMPARISONS

**Equivariance:** A function \( \phi : x \in X \rightarrow y \in Y \) is equivariant under the action of a group \( G \) if there are actions of \( G \) on \( X \) and \( Y \) such that \( \phi(g \cdot x) = g \cdot \phi(x) \). Information being gained or lost at the boundary is an obstacle to applying the theory of group actions exactly (except for certain finite groups [62]). If one relaxes the definition to require only an approximate match, well-known visual feature representations tend to have strong equivariance properties either by design or in practice [63]. Imposing equivariance properties on intrinsic image decompositions (albedo at a point should not depend on image crop) shows sharp improvements [1]. Relighting has a form of equivariance property: small movements of the camera frame should not affect the overall relighting of a scene (apart from rows and columns of the image moving in and out of view). We use that paper’s crude averaging strategy because we know no better. In this approach we: pad an image; estimate multipliers for a number of closely overlapping crops of the padded image; average overlapping estimates to obtain final multipliers. Doing so reliably loses both MSD and FID\(_\infty\), for reasons we do not know (Figure 8).

**Pixel uniformity loss:** We expect there is some lighting that makes any location look bright (meaning the image value at each location should be bright), and that any location might be in darkness (meaning that the shading value at each location might be dark). Write \( \uparrow M_V I_V \) for the largest value at each location over all generators of \( M_V I_V \) and \( \downarrow M_V S_V \) for the smallest value at each location over all generators of \( M_V S_V \). We use:

\[
\mathcal{L}_P(I_V; \phi) = E
\]

where

\[
E = [-\min(0.95 - \uparrow M_V I_V, 0) + \max(\downarrow M_V S_V - 0.05, 0)]
\]

Pixel uniformity loss reliably produces more speculative models with weaker FID\(_\infty\) (Figure 7).
Fig. 7. $\text{FID}_\infty$ against MSD for various parameter choices for model selection. **Left:** PatchGAN adversary scale affects performance; the small receptive field is best. **Center:** $\lambda_B$ encourages variations without much adverse affect on $\text{FID}_\infty$; we chose $\lambda_B = 5$. **Right:** $\lambda_P > 0$ reliably produces more speculative models with weaker FID; the figure is Figure 4 from paper, for reference.

Fig. 8. Averaging for equivariance during inference hurts metrics. $\text{FID}_\infty$ against MSD for **red:** 256x256 images relighted using averaging at inference time; **green:** relighted 256x256 square crops; **yellow:** 256x256 square crops from averaged images.
Fig. 9. No major effects result from using FID rather than FID\(_\infty\). FID against MSD for various parameter choices for model selection. Left: PatchGAN adversary scale affects performance; the small receptive field is best. Center: \(\lambda_B\) encourages variations without much adverse affect on FID\(_\infty\); we chose \(\lambda_B = 5\). Right: \(\lambda_P > 0\) reliably produces more speculative models with weaker FID. Compare Fig. 7.
Fig. 10. Images relighted at $\alpha = 0.1$ (more variable; left; cf $\alpha = 0.5$, less variable, $\alpha = 0.01$, much more variable, in Fig 11) and the first five multiplier fields for each right. Notice how the multipliers are keyed to structures in the image – the method "knows" some geometry and radiometry. Reproduced from Fig 3 of main paper, without arrows, for reference.
Fig. 11. Images relighted at $\alpha = 0.5$ left, less variable and $\alpha = 0.01$, much more variable right. At $\alpha = 0.5$ images have higher quality, but tend to be less variable. At $\alpha = 0.01$, the difficulties presented by extrapolation become more obvious.

Fig. 12. Our method produces qualitative changes in the distribution of light, and appears to generalize well across datasets. Left: Further qualitative results on LSUN-Bedrooms; Right: Results for the first image each of the validation set for 10 scenes with “room” in their name from the PlacesRooms dataset; relights are not curated. Top: Image; other: relights obtained using our method, $\alpha = 0.5$. All relights are shown at the same mean intensity as the original image, to ensure that effects are not masked by overall shift in brightness.
We show all pairs of source/target/least-squares images used to prepare Figure 5 of the main paper. Recall that shows results of: rendering CGI scenes under two lights; computing the multipliers for the source; constructing weighted sum of multipliers that yields a least squares fit to the target; then comparing target to the least squares. The residual error in intensity is shown on a scale where white is 1; mid grey is 0; and black is -1. Note relatively good fits, suggesting that for the cases we have investigated, multipliers are a reasonable fit. Quantitative: over a total of 11 pairs (22 tests), for a dataset we will provide, we find average metrics between the least squares prediction and the target of RMSD: 0.0503 PSNR: 26.6 LPIPS: 0.0648, suggesting quite successful prediction of the target by source multipliers.
9 Appendix II: Proof of Theorem 1

We wish to compare the radiosity of a scene with that of a similar scene, of somewhat different geometry, albedo and luminaire. The key result is the difference between these two radiosities can be described in a sensible way, and can be bounded. In turn, it follows that sufficiently similar scenes will share generators.

9.1 Notation

It is enough to work with a simplified model of a scene $V$ as a tuple $(\mathcal{G}, \rho, E)$ of geometry, albedo and luminaire (rather than stochastic family of luminaires). $\mathcal{G}$ consists of a set of surfaces parametrized in some way and a 2D domain for that parameterization; write $\mathcal{G} = (s(x), D)$. Significant complications in the analysis can occur associated with concave “creases” or at surface interpenetrations. These complications aren’t physically important – every corner in a room is made smooth by at least a thin layer of paint or dirt – and so we avoid them by assuming:

- $\mathcal{D}$ is compact (but need not be connected) with finite area.
- Each connected component of $D$ represents a distinct surface.
- Surfaces are parametrized by $s(x)$, where $s$ is at least $C^1$ and is an embedding (surfaces do not self-intersect).
- Surfaces are orientable, and are oriented.
- For any pair of points $x, y$, we have $k(x, y) \leq C$ for some finite constant $C$ and $k$ defined below. This rules out “throats” – places where the gap between surfaces narrows and then opens out, Figure 14 – in the geometry, and bounds the largest curvatures. Larger $C$ corresponds to narrower throats and tighter curvatures being allowed. The condition is sufficient to ensure the kernel is compact, below.

Note that parametrizing surfaces by atlases of charts complicates notation without changing the analysis; we will suppress these details. Write

$$
\|E\|_{p} = \left[ \int_{D} E^{p} d\mathbf{x} \right]^{1/p}
$$

(1 \leq p \leq \infty) for the $p$ norm of $E$. Recall that for an operator $K$

$$
\|K\|_{p} \leq \|K\|_{\infty} \left( \frac{p-1}{p} \right) \|K\|_{\frac{1}{p}}
$$

(eg [40], [64] p.144 or [65] p.518). All operators here will have both $\|K\|_{1} < 1$ and $\|K\|_{\infty} < 1$, so that results apply for any $p$; we write $\|K\|$.

We wish to compare the radiosity $B_{V}$ of a scene $V$ with the radiosity $B_{V'}$ of a similar scene $V'$. These scenes have somewhat different geometry, albedo and luminaire. We consider affine transformations of the geometry, so that $\mathcal{G}'(\mathcal{G}, A, b) = (A s(x) + b, D)$. We will prove:

**Theorem 1:** For $V = (\mathcal{G}, \rho, E)$ and $V' = (\mathcal{G}', A, b, \rho', E')$, where $\epsilon_{E} \|E\| = \|E - E'\|$, $\epsilon_{\rho} = \sup_{\rho, \rho'} |\rho - \rho'|$, $p = \sup_{D} \rho$, $p' = \sup_{D} \rho'$, $c$ is the condition number of $A$ (ratio of largest to smallest eigenvalues), we have:

$$
\|B_{V} - B_{V'}\| \leq \left[ \frac{\epsilon_{\rho} + 1}{\rho} \right] + \|E\| + \|E'\|
$$

9.2 Standard results

The geometry consists of a set of parametrized surfaces. Write $\mathcal{D}$ for the domain of the parametrization. For a parameter vector $x \in \mathcal{D}$ write $s(x)$ for the corresponding surface point; $B(x)$ for radiosity at that point; $\rho(x)$ for the albedo at that point; $E(x)$ for the radiance at that point; $V(x, y)$ for the visibility function that is 1 when there is a direct line of sight from $s(x)$ to $s(y)$ and 0 otherwise; $d_{s(x)\rightarrow s(y)} = s(x) - s(y)$; and $N(x)$ for a unit normal at $x$. Figure 14 shows some notation commonly used. Standard diffuse interreflection theory (see, eg [50], [52], [53], [66]) gives

$$
B(x) = E(x) + \rho(x) \int_{D} k(x, y) B(y) dA_{y}
$$

where

$$
k(x, y) = \frac{1}{\pi} \frac{V(x, y) \cos \theta_{x} \cos \theta_{y}}{||d_{x-y}||^2} + \text{ (notion in Figure 14)}
$$

= \frac{V(x, y) [N^{T}(x) d_{x-y}] [N^{T}(y) d_{y-x}]}{[d_{x-y} d_{x-y}]^2}
$$

(recalling $N$ is a unit normal). Write $D_{\rho}$ for the linear operator that maps $f(x)$ to $(\rho f)(x)$. We can now abbreviate into $B = E + D_{\rho} T B = E + K B$, where $D_{\rho}, T$ and $K$ are linear integral operators. It is known that

$$
\|T\| \leq 1
$$

(for example, our $T$ is the $K_{G}$ of [40], sec 5.3, with constant diffuse BRDF of 1). Furthermore, writing

$$
p = \|\rho(x)\|_{1} < 1
$$

we have

$$
\|D_{\rho}\| \leq p
$$

by simple calculation and so

$$
\|K\| \leq p.
$$

For our assumptions,

$$
\int_{D} \int_{D} |k(x, y)|^2 dA_{x} dA_{y} \leq C^2 \text{Area}(D)^2,
$$

so that $k$ is a Hilbert-Schmidt kernel. Because the kernel is Hilbert-Schmidt, $K$ is a compact operator (eg [67]). There is then a Neumann series solution

$$
B = M^{-1} E
$$

= $E + K E + K^2 E + \ldots$

= exitance + one bounce + two bounces + \ldots

which converges for all $p < 1$ (and so all practical scenes). It follows that

$$
\|M^{-1}\| \leq \frac{1}{1 - p}
$$
9.3 Lemma 1: Perturbed Luminaire

Assume that the luminaires for \( V' \) differ from those of \( V \) (but geometry and albedo are the same), so \( V = (\mathcal{G}, \rho, E) \) and \( V' = (\mathcal{G}, \rho', E') \). Standard results yield:

**Lemma 1:** using the notation above

\[
|B_V - B_{V'}| \leq \frac{1}{1 - p} \|E - E'\| = \frac{\epsilon_E}{1 - p} |E|.
\]

**Proof:** \( |M^{-1}| \) is bounded by \( \frac{1}{1 - p} \), above. □

9.4 Lemma 3: Perturbed Albedo

Now assume that \( V = (\mathcal{G}, \rho, E) \) and \( V' = (\mathcal{G}, \rho', E) \) and \( \rho' = \rho + \delta \). Write \( \epsilon_\rho = \|D_\delta\| \). We expect \( B' \) to differ from \( B_V \) because it contains multiple bounce terms (above). For example, the shading in a very dark room is notably different from that in a very light room (eg \([50],[66],[68]\)), because low albedos mean that the multiple bounce terms are very small. Equivalently \( M \) and \( M' \) are different. However, \( G \) is bounded and we have \( |M| = p \leq 1 \) and \( |M'| = p' \leq 1 \). The dynamic range of practical albedos is quite limited (eg \([69]\)), so it is reasonable to bound \( \rho(x) \) away from both 0 and 1.

**Lemma 2:** using the notation above, for all pairs of \( 0 < \epsilon \leq \rho \leq 1 - \epsilon < 1 \) and \( 0 < \epsilon' \leq (1 - \epsilon) < 1 \), \( \|D_{\rho - \rho'}\| = \epsilon_\rho < 1 \)

**Proof:** for some test function \( f \),

\[
D_{\rho - \rho'} f = (\rho(x) - \rho'(x)) f(x).
\]

Then

\[
\|D_{\rho - \rho'} f\|_1 = \int_D |(\rho(x) - \rho'(x)) f(x)| \, dA_x
\]

\[
\leq \int_D |(\rho(x) - \rho'(x))| |f(x)| \, dA_x
\]

\[
\leq \sup_x \| (\rho(x) - \rho'(x)) \| \int |f(x)| \, dA_x.
\]

The constraints on \( \rho, \rho' \) ensure

\[
\sup_x \| (\rho(x) - \rho'(x)) \| < 1
\]

and \( \|D_{\rho - \rho'}\|_\infty < 1 \), so the result follows for all norms. □

This yields the novel (but fairly straightforward)

**Lemma 3:** using the notation above

\[
\|B_{V'} - B_V\| \leq \left( \frac{1}{1 - p} \right) \left( \frac{1}{1 - p'} \right) \epsilon_\rho |E|.
\]

**Proof:** We have \( (I - D_\rho G)B_V = E = (I - D_{\rho + \delta} G)B_{V'} \), so

\[
B_{V'} - B_V = D_{\rho'} GB_{V'} - D_{\rho'} GB_V + D_{\rho'} G(B_{V'} - B_V)
\]

\[
= D_{\rho'} GB_V + D_{\rho'} G(B_{V'} - B_V)
\]

\[
= D_{\rho'} GM^{-1} E + D_{\rho'} G(B_{V'} - B_V)
\]

so

\[
B_{V'} - B_V = M'^{-1} D_{\delta} GM^{-1} E.
\]

Now \( \|D_\delta G\| \leq \epsilon_\rho \), so

\[
\|B_{V'} - B_V\| = \|M'^{-1} D_{\delta} GM^{-1} E\|
\]

\[
\leq \left( \frac{1}{1 - p} \right) \left( \frac{1}{1 - p'} \right) \epsilon_\rho |E|.
\]

□

9.5 Theorem 1: Perturbed Geometry

We wish to study the effect of small changes in geometry on the solution \( B \). We have found no results on the effect of perturbing geometry, which appear to be hard to obtain (cf \([43]\), p176 says they are; \([44]\) estimate the effect of flat ports on an integrating sphere).

Small geometric changes can produce major changes in visibility which are difficult to bound. We consider affine transformations of the surfaces, so that \( s(x) \) is mapped to \( T(s) = As(x) + b \) (det(\( A \)) \( \geq 0 \)). We assume that the smallest singular value of \( A \) is positive. There are two advantages of this model: first, it is a reasonable description of small changes between otherwise quite similar rooms (small anisotropic scalings in the right coordinate system); second, the visibility function for \( T(s) \) is the same as the visibility function for \( s \) (lemma 4, below).
9.5.1 Transformations and Functions

The transformation produces a natural bijection between functions of original and transformed geometry. This is so natural that it can be difficult to see at first glance. Assume we have a function \( f \) defined on the first geometry. We must know the value that \( f \) takes at each point on the geometry, which means we know the value at each point in the parameter space. But this parameter space parametrizes the transformed geometry, too, so we have a corresponding function on the transformed geometry (this argument yields that the mapping between function spaces is a bijection). For \( f \) a function defined on the original parametrization, write \( \mathcal{T}(f) \) for the transformed version.

This natural correspondence may not respect the geometric transformation on the embedded surfaces. This means we need to distinguish between a function mapped through the correspondence and one evaluated on the new geometry. For example, consider the function \( \text{sqdist}(x_0, x_1) = \| s(x_0) - s(x_1) \|^2 \). If \( A = 2 \times \text{Id} \), the distance between points on the transformed geometry has doubled. Write \( \text{sqdist}'(x_0, x_1) = \| As(x_0) - As(x_1) \|^2 \) (the prime indicates the function is evaluated on the transformed geometry). We have \( \mathcal{T}(\text{sqdist}) = (1/4)\| As(x_0) - As(x_1) \|^2 = (1/4)\text{sqdist}'(x_0, x_1) \).

9.5.2 Transformations and the Kernel

A similar natural correspondence applies to operators. We must now consider the relationship between \( \mathcal{T}(K) \) and \( K' \). Each is a bounded linear operator defined on functions of \( G' \), but they are different. \( \mathcal{T}(K) \) is obtained by applying the natural correspondence to \( K \); in contrast \( K' \) is obtained by computing the kernel using the geometry of \( G' \), which may give a different operation.

By tracking the effects of affine transformations on the terms of \( k \), we can constrain the difference. First, we have

\[
V(x, y) = \mathcal{T}(V(x, y)) = V'(x, y)
\]

Write \( \sigma_x \) for the largest singular value of \( A \), \( \sigma_n \) for the smallest (which is positive), and \( c = \frac{\sigma_x}{\sigma_n} \) for the condition number. We have

Lemma 6: \( (\sigma_n)N'^T(\mathbf{u} \rightarrow \mathbf{v}) \leq N'^T(\mathbf{u} \rightarrow \mathbf{v}) \leq (\sigma_x)N'^T(\mathbf{u} \rightarrow \mathbf{v}) \)

Proof: \( N'^T(\mathbf{u} \rightarrow \mathbf{v}) = \frac{N^T A^{-1} A_k - \mathbf{v}}{\sqrt{N^T A^{-1} A_k - N^T A^{-1}N}} \) and the bound follows.

Lemma 7: \( \sigma_n^2 \mathcal{T}(dA_y) \leq dA_y' \leq \sigma_x^2 \mathcal{T}(dA_y) \).

Proof: \( dA_y' \) is the infinitesimal area spanned by two infinitesimal vectors \( dx_1 \) and \( dx_2 \). The largest (resp. smallest) scaling of each length is by the largest (resp. smallest) singular value of \( A \).

We then obtain the novel:

Lemma 8: For affine transformations, \( c^{-1}K(\mathcal{T}(f)) \leq K'(\mathcal{T}(f)) \leq cK(\mathcal{T}(f)) \) for any \( f \)

Proof: For any two vectors \( \mathbf{u}^T \mathbf{v} = \mathbf{u}^T A^T A \mathbf{v} \). Then \( \sigma^2_n \mathbf{u}^T \mathbf{v} \leq \mathbf{u}^T A^T A \mathbf{v} \leq \sigma^2_x \mathbf{u}^T \mathbf{v} \). Lemma 6 deals with the angles. Write

\[
\mathbf{K}'(\mathcal{T}(f)) = \int_D k'(x, y) f(y) dA_y'
\]

where

\[
k'(x, y) = \frac{V'(x, y) [N'^T(\mathbf{x} \rightarrow \mathbf{y}) d_{x \rightarrow y}^x | N'^T(\mathbf{y} \rightarrow \mathbf{x}) d_{y \rightarrow x}^y]}{d_{x \rightarrow y}^x d_{y \rightarrow x}^y}
\]

But

\[
\begin{bmatrix}
([\sigma_n]N'(x) d_{x \rightarrow y}^x) \\
([\sigma_n]N'(y) d_{y \rightarrow x}^y)
\end{bmatrix} \leq \begin{bmatrix}
[N'^T(\mathbf{x} \rightarrow \mathbf{y}) d_{x \rightarrow y}^x) \\
N'^T(\mathbf{y} \rightarrow \mathbf{x}) d_{y \rightarrow x}^y)
\end{bmatrix}
\]

and

\[
\sigma^4_n \left[ d_{x \rightarrow y}^x d_{x \rightarrow y}^y \right]^2 \leq \left[ d_{x \rightarrow y}^x d_{x \rightarrow y}^y \right]^2
\]

so that

\[
c^{-4} \mathcal{T}(K)(\mathcal{T}(f)) = \int_D \frac{\sigma^4_n}{\sigma^2_n} k(x, y) f(y) dA_y
\]

Proof: Write \( \Pi_k \) for the tangent plane at \( x \). For \( \mathbf{u} \in \Pi_k \), we have for non-unit normal \( \mathbf{n}, \mathbf{n}(x)^T(\mathbf{u} - \mathbf{x}) = 0 \). But \( \mathbf{u} \) transforms to \( A\mathbf{u} + \mathbf{b} \), etc. and we must have \( \mathbf{n}(x)^T(A(\mathbf{u} - \mathbf{x})) = 0 \), and the result follows by requiring the normal be a unit normal; equivalently, normals are covariant.

□.
which leads to

**Lemma 9:** With the notation above,

\[ \| B_{V'} - B_{V} \| \leq \frac{(c^4 - 1)}{(1 - p^2)} \| E \| \]

**Proof:**

\[
B_{V'} - B_{V} = (\mathcal{T}(E) - E) + (K' B_{V'} - K B_{V})
\]

(\text{recall } \mathcal{T}(E) = E)

\[
= K' B_{V'} - K' B_{V} + K' B_{V} - K B_{V}
\]

so that

\[
(I - K') (B_{V'} - B_{V}) = (K' - K) B_{V}
\]

and

\[
(B_{V'} - B_{V}) = M^{-1} (K' - K) B_{V}
\]

Notice that the kernel is everywhere non-negative, and for every point \( x \) and for any non-negative function \( f \), we have \( 0 \leq (K f)(x) \). Since \( E \) is non-negative, so is \( B_{V} \). From lemma 6, for non-negative \( f \), we have \( (K' f)(x) \leq c^4 (K f)(x) \) at every point \( x \), so that \( K' f - K f \leq (c^4 - 1) K f \) at every point. This means

\[
\| (B_{V'} - B_{V}) \| = \| M^{-1} (K' - K) B_{V} \|
\]

\[
\leq \frac{1}{(1 - p^2)} \| (K' - K) B_{V} \|
\]

\[
\leq \frac{1}{(1 - p^2)} (c^4 - 1) \| B_{V} \|
\]

\[
\leq \frac{1}{(1 - p^2)} (c^4 - 1) \frac{1}{(1 - p)} \| E \|
\]

\[
= \frac{(c^4 - 1)}{(1 - p^2)} \| E \|
\]

where the last step works because \( p' = p \), because \( p' = T(\rho) = \rho \). □

Finally, we can prove the entirely novel

**Theorem 1:** For \( V = (\mathcal{G}, \rho, E) \) and \( V' = (\mathcal{G}', \rho', E') \), with \( \epsilon_E \| E \| = \| E - E' \|, \epsilon_\rho = \sup \sigma |\rho - \rho'|, p = \sup \sigma p, p' = \sup \sigma p', c \) is the condition number of \( A \) (ratio of largest to smallest eigenvalues), we have:

\[
\| B_{V} - B_{V'} \| \leq \left[ \frac{\epsilon_E}{1 - p^2} + \frac{(c^4 - 1)}{(1 - p^2)} \right] \| E \|
\]

**Proof:** Write \( V_1 = (\mathcal{G}, \rho, E'), V_2 = (\mathcal{G}', \rho, E') \). Then

\[
\| B_{V} - B_{V'} \| \leq \left[ \frac{\| B_{V} - B_{V_1} \| + \| B_{V_1} - B_{V_2} \| + \| B_{V_2} - B_{V'} \|}{3} \right] \text{ (triangle inequality)}
\]

and \( \| E' \| \leq (1 + \epsilon_E) \| E \| \) but

\[
\| B_{V} - B_{V_1} \| \leq \frac{\epsilon_\rho}{1 - p} \| E \| \text{ (Lemma 1)}
\]

\[
\| B_{V_1} - B_{V_2} \| \leq \frac{\epsilon_\rho}{(1 - p)(1 - p')} \| E' \| \text{ (Lemma 3)}
\]

\[
\leq \frac{\epsilon_\rho}{(1 - p)(1 - p')} (1 + \epsilon_E) \| E \| \text{ (definitions)}
\]

\[
\| B_{V_2} - B_{V'} \| \leq \frac{(c^4 - 1)}{(1 - p^2)} (1 + \epsilon_E) \| E \| \text{ (Lemma 9)}
\]

\( \Box \)

**10 Appendix III: Proof of Theorem 2**

We now establish that viewing shading fields for a set of similar scenes \( V_i \) is enough to estimate a representation of the generators for a given scene \( V \). It is important to have a representation of the probability that a scene will be illuminated in some particular way. We use the general definition of a scene, above. A scene \( V \) is a tuple of geometry, albedo and luminaire model; write \( V = (\mathcal{G}, \rho, \mathcal{E}) \), where \( \mathcal{E} \) is the luminaire model. The geometry \( \mathcal{G} \) consists of a set of surfaces parametrized in some way and a 2D domain for that parametrization; write \( \mathcal{G} = (s(x), \mathcal{D}) \). The luminaire model captures the idea that lighting in a particular scene may change. We have \( \mathcal{E} = (E_1, \ldots, E_{N_e}, P(\theta)) \) for a set of potentially many basis luminaires and a probability distribution over the coefficients. This information implies a probability distribution over luminaires for the scene \( P(E) \). Two scenes \( V \) and \( V' \) are similar if there is some affine transformation \( \mathcal{A} \) of \( \mathcal{G} \) (smallest singular value of \( \mathcal{A} \) greater than zero) such that \( \mathcal{G}'(\mathcal{G}, \mathcal{A}, \mathcal{B}) = (\mathcal{A}(s(x)) + \mathcal{B}) \) and \( \mathcal{E} = \mathcal{E}' \). Notice this second constraint does not require that the two scenes are lit in the same way, just that the basis of luminaires is the same and the probability distribution over lighting coefficients is the same.

We pass to a finite dimensional representation without loss of generality (to avoid having to deal with function spaces). Choose some \( N_o \) element orthonormal basis to represent the possible radiosity vectors for the scene (if \( N_o = N_e \), the representation is exact, because there is a finite dimensional space of luminaires). In this basis, any particular radiosity \( B(x) \) has coefficient vector \( b \). Write \( \phi_i(x) \) for the \( i \)th element of the basis; \( \Phi = [\phi_1(x), \ldots, \phi_{N_e}(x)] \), \( b_i \) for the coefficient vector representing \( B_i(x) \) (which is the radiosity resulting from \( E_i(x) \)); and \( \mathcal{B} = [b_1, \ldots, b_{N_e}] \).

For any \( r \leq N_o \), an effective generator matrix for scene \( V \) is the \( N_o \times r \) matrix \( \mathcal{G}_{e,V,r} \) such that for any illumination condition producing radiosity represented by \( b \), there is some \( w \) such that the radiosity represented by \( \mathcal{G}_{e,V,r}w \) is similar to the actual radiosity field. In particular, we want

\[
\mathcal{L}(\mathcal{G}_{e,V,r}) = \mathbb{E}_{\theta \sim P(\theta)} \left[ \inf_{w} \| \mathcal{G}_{e,V,r}w - B\theta \|_2^2 \right]
\]

to be small; \( \mathcal{G}_{e,V,r} \) is clearly not unique. We choose to have \( \mathcal{G}_{e,V,r} \) orthonormal.

Assume we have \( k \) radiosity fields \( B_i \) for a scene under different, unknown, illumination conditions, for \( k \) very large. We assume that for the \( i \)th field, \( \theta_i \sim P(\theta) \) (which is not
known). We can clearly estimate an effective generator matrix by replacing the expectation with an average. But we do not have $k$ images of the same scene.

**Lemma 10:** Using the notation above, write

$$C_b = \mathcal{B} \left( \mathbb{E}_{\theta \sim P(\theta)} \left[ \theta G^T \right] \right) B^T.$$

Then:

$$\mathbb{E}_{\theta \sim P(\theta)} \left[ \left\| G_{e,V,r} w - B \theta \right\|^2 \right]$$

is equal to

$$\text{Tr} \left[ C_b \right] - \text{Tr} \left[ G^T_{e,V,r} C_b G_{e,V,r} \right].$$

**Proof:** Manipulation; note that

$$C_b = \mathbb{E}_{\theta \sim P(\theta)} \left[ B \theta \theta^T B^T \right];$$

that

$$\mathbb{E}_{\theta \sim P(\theta)} \left[ \theta^T B^T B \theta \right] = \text{Tr} \left[ C_b \right];$$

and that

$$\mathbb{E}_{\theta \sim P(\theta)} \left[ \theta^T B^T G_{e,V,r} G_{e,V,r}^T B \theta \right] = \text{Tr} \left[ G^T_{e,V,r} C_b G_{e,V,r} \right].$$

□

We cannot know $C_b$, but we do not need to.

**Lemma 11:** Using the notation above; writing $e_i$ for the eigenvector of $C_b$ corresponding to the $i$'th largest eigenvalue; and assuming that the eigenvalues of $C_b$ are distinct, an optimal value of $L(G_{e,V,r})$ is given by

$$[e_1, \ldots, e_r]^T \mathcal{M}$$

for any $r \times r$ matrix $\mathcal{M}$ of full rank such that $\mathcal{M}^T \mathcal{M} = I$

**Proof:** Follows directly from lemma 10. □

**Lemma 12:** Using the notation above, and writing $\lambda_i$ for the $i$'th largest eigenvalue of $C_b$, the optimal value of $L(G_{e,V,r})$ is given by

$$\sum_{i=r+1}^{N_s} \lambda_i$$

**Proof:** Follows directly from lemma 10 and 11. □

Assume we collect $N_s$ samples of radiosity $\mathbf{b} = B \theta$, where $\mathbf{b}$ is the $i$'th such sample, and $\theta \sim P(\theta)$. These yield an estimate of $C_b$, which can be accurate.

**Lemma 13:** Write $\mathbf{b}_i = (1/N_s) \sum_{i=1}^{N_s} \mathbf{b}$. Then there is some constant $C_e$ such that

$$\mathbb{E} \left[ \mathbf{b}_i \right] = \mathbb{E} \left[ \mathbf{C}_b \right]$$

and

$$\mathbb{E} \left[ \left\| \mathbf{b}_i \right\|^2 \right] = (1/N_s) C_e$$

**Proof:** This follows from the weak law of large numbers.

Now collect $N_s$ samples of radiosity from different scenes $V$, each one similar to the scene $V$, where $A_i$ maps the geometry of $V$ to $V_i$. Write $b_{i,v}$ for the $i$'th such sample. Recall that similar scenes have the same basis of luminaires and the same distribution for $\theta$. These samples yield an estimate of $C_b$, too.

**Lemma 14:** Given $b$, $r$ and $\delta$ such that (a) $\left\| b - c \right\|_2 \leq \delta$ and (b) $\left\| b \right\|_2 \leq r$, we have:

$$\left\| b^T - cc^T \right\|_F \leq 3\delta^2 r^2 + 4\delta^3 r + \delta^4$$

**Proof:** Write $\Delta = c - b$. Then

$$\left\| b^T - cc^T \right\|_F = \text{Tr} \left[ \left( b^T - cc^T \right) \left( b^T - cc^T \right)^T \right]$$

$$\leq \frac{(b^T b)^2}{2} + \frac{(b^T b + \Delta^T b)^2}{2}$$

$$\leq 4\delta r^3 + 3\delta^2 r^2 + 2\delta^3 r + \delta^4$$

□.

**Lemma 15:** Write $\mathbf{b}_i = \mathbf{b}_i \mathbf{v}_i$. Using the notation above, and writing $p = \max_i \left( \text{sup} \left| \rho_v - \rho_{v_i} \right| \right); \epsilon = \max_i \left( \text{sup} \left| \rho_v - \rho_{v_i} \right| \right); c$ for the largest condition number of $A_i$. Write

$$\delta = \frac{\epsilon + (c^3 - 1)}{1 - p}$$

and $r = \frac{1}{1 - p}$

Then

$$\left\| \mathbf{b}_i - \mathbf{b}_i \right\|_F \leq N_s \left( 4\delta^2 r^3 + 3\delta^2 r^2 + 2\delta^3 r + \delta^4 \right)$$

**Proof:** The $i$'th sample $\theta_i$ from $P(\theta)$ yields a luminaire $E_i$; write $B_i,v$ for the resulting radiosity in $V$ and $b_i,v$ for its finite dimensional representation. Theorem 1 yields a bound on $\left\| B_i,v - B_i,v_i \right\|$, and so on $\left| b_i,v - b_i,v_i \right|$. We have $\left| b_i,v - b_i,v_i \right| \leq \frac{\epsilon + (c^3 - 1)}{1 - p}$. Furthermore, $\left| b_i,v \right| \leq 1/(1 - p)$. The result then follows from lemma 14. □.

We can now compute an effective generator matrix $\hat{G}$ from an estimate $\hat{C}$ of $C_b$. This estimate uses observations of multiple similar scenes. The $\hat{G}$ we obtain from data like this will not be the true $G$, and we must determine what problems will result from using $\hat{G}$ (which we can get) in place of $G$ (which is the true best EGM).

Theorem 2 establishes that using $\hat{G}$ in place of $G$ incurs bounded error. Recall the expected error of using $\hat{G}$ as the effective generator matrix for estimate $\hat{C}$ is

$$L(G;C) = \left( \text{Tr} \left[ \mathcal{C} \right] - \text{Tr} \left[ \mathcal{G}^T \mathcal{G} \right] \right)$$
Now estimate the effective generator matrix using \( C_{b,m,s} \) to obtain
\[
\hat{G} = \arg \max_{G | G^T G = I} \mathcal{L}(G; C_{b,m,s})
\]
Using \( \hat{G} \) to represent the radiosity of the scene will incur expected error
\[
\mathcal{L}(\hat{G}; C_b) = \left( \text{Tr}[C_b] - \text{Tr}[\hat{G}^T C_b \hat{G}] \right).
\]

**Theorem 2:** \( \mathcal{L}(\hat{G}; C_b) - \mathcal{L}(G; C_b) \) is bounded.

**Proof:**
We have
\[
\mathcal{L}(\hat{G}; C_b) - \mathcal{L}(G; C_b) = \text{Tr}[\hat{G}^T C_b \hat{G}] - \text{Tr}[G^T C_b G]
\]
Recall \( G \) is chosen to minimize \( \mathcal{L}(G; C_b) \), and is orthonormal, so will consist of the eigenvectors of \( C_b \) corresponding to the \( r \) largest eigenvalues. In turn, \( \hat{G} \) is chosen to minimize \( \mathcal{L}(\hat{G}; C_{b,m,s}) \). We must now find the \( \hat{G} \) so that \( \text{Tr}[\hat{G}^T C_b \hat{G}] \) is minimized within the bound on \( C_b - C_{b,m,s} \). Any extremal \( \hat{G} \) must consist of eigenvectors of \( C_b \) (recall \( \hat{G} \) is orthonormal). This yields an immediate but rather loose bound. Label the eigenvalues \( \lambda_i \) of \( C_b \) so that \( \lambda_1 \geq \lambda_2 \geq \ldots \lambda_{N_o} \), then
\[
\inf_{\hat{G} | \hat{G}^T \hat{G} = I} \text{Tr}[\hat{G}^T C_b \hat{G}] = \sum_{i=N_o-r+1}^{N_o} \lambda_i
\]
so
\[
\mathcal{L}(\hat{G}; C_b) - \mathcal{L}(G; C_b) \leq \sum_{i=1}^{N_o} \lambda_i - \sum_{i=N_o-r+1}^{N_o} \lambda_i.
\]

\( \square \)

This bound is quite loose, but for the purposes of this paper, the existence of a bound is enough. Some quite interesting mathematics is possible in tightening the bound, and there may be practical consequences.

Notice that, writing \( \Phi(G; C) = \text{Tr}[C] - \text{Tr}[G^T CG] \), with \( C \) symmetric, diagonalized as \( U^T \Lambda U \); then \( \Phi(UG; \Lambda) = \Phi(G; C) \). This means we can work with diagonal \( C \) WLOG. In turn, determining the value of \( \mathcal{L}(\hat{G}; C_b) - \mathcal{L}(G; C_b) \) requires thinking about shuffling eigenvalues. Any bound will follow from reasoning about permutations in the order of eigenvectors achieved by the error, and the estimation error may not be large enough to permute the eigenvalues so aggressively. A more interesting and tighter bound can be obtained as
\[
\sum_{i=1}^{N_o} \lambda_i - v
\]
where \( v \) is the value of a linear program representing the permutation of the eigenvalues. Write \( 1_r \) for the \( N_o \) dimensional vector consisting of \( r \) ones and \( N_o - r \) zeros in order, \( \lambda = [\lambda_1, \ldots, \lambda_{N_o}] \) (eigenvalues of \( C_b \)), \( \mathcal{P} \) for a permutation matrix, \( \delta \) for a vector of offsets, \( D \) for \( \| C_b - C_{b,m,s} \|_F^2 \), then \( v \) is the value of:
\[
\min_{\mathcal{P}, \delta, m} \frac{1}{r} \sum_{i} p_{ij} \quad \text{subject to} \quad \begin{array}{c}
\sum_{i} p_{ij} = 1 \\
\sum_{i} p_{ij} = 1 \\
p_{ij} \geq 0 \\
u = \mathcal{P} \lambda \\
m = u + \delta \\
m_1 \geq m_2 \geq \ldots m_{N_o} \\\n\sum_{i} \delta_i^2 \leq D.
\end{array}
\]

This problem finds the permutation of eigenvalues that produces the smallest value within the budget \( D \). For \( D \geq 0 \), the program is feasible (\( \mathcal{P} = I \) is always a feasible point), and Birkhoff’s theorem guarantees that a there is a solution with integer \( \mathcal{P} \). Interestingly, for sufficiently small \( D \), \( v = m_i \lambda_i \), meaning the error bound is zero! This occurs when \( D \) is smaller than \( \lambda_{N_o} - \lambda_{N_o+1} \). The quantized character of this bound means that quite small improvements in estimation of \( C_{b,m,s} \) could result in large improvements in the practical behavior of the method.

### 11 APPENDIX IV: THE LOCAL FID

**Definition:** Assume we obtain a collection of images \( \mathcal{R} \) from a ground truth collection \( \mathcal{O} \) by relighting a single image \( I_k \) and compute FID(\( \mathcal{O}, \mathcal{R} \)). This value – which is non-negative, and is ideally zero – measures the extent to which the relighted image is realistic. Furthermore, different relightings of the same image will have different values. Write \( \mu_0 \) for the mean and \( C_O \) for the covariance of \( \mathcal{O} \) in the chosen embedding, etc. We have
\[
\text{FID}(\mathcal{O}, \mathcal{R}) = ||\mu_R - \mu_0||_2^2 + \text{Trace}(C_O + C_R - 2(C_O C_R)^{1/2}).
\]

**Computing difficulties:** Numerical problems occur because \( C_R \) is very similar to \( C_O \), and \( C_O \) has many small eigenvalues; this means there are many eigenvalues of \( C_O C_R \) that are very close to zero, which is a reliable source of trouble for matrix square root methods. One typically gets complex estimates.

**Series estimate:** Write \( x(I_k) \) for the embedding of \( I_k, R_k \) for the result of relighting \( I_k, R_k \), and \( d_k = x(R_k) - x(I_k) \). Then
\[
\mu_R - \mu_O = \frac{d_k}{N}, \quad \text{and}
\]
\[
C_R = C_O + \frac{1}{N} \left( \frac{d_k (x(I_k) - \mu_0)^T +}{(x(I_k) - \mu_0) d_k^T +} \right) \left( \frac{d_k^T d_k}{N^2} \right)
\]
\[
= C_O + \frac{1}{N} U + \frac{1}{N^2} V
\]
We seek an approximation to \((C_O C_R)^{(1/2)}\) of the form \(C_O + \frac{1}{N} M_1 + \frac{1}{N^2} M_2 + \ldots\). By squaring and matching terms, we have
\[
C_O M_1 + M_1 C_O = C_O U
\]
and
\[
C_O M_2 + M_2 C_O + M_1 M_1 = C_O V
\]
so \(2 \text{Trace}(M_1) = \text{Trace}(U)\) and first order terms cancel. Similarly, \(\text{Trace}(V) = 2 \text{Trace}(M_2) + \text{Trace}(C_O^{-1} M_1^2)\), and we have
\[
\text{FID}(O, R) = \frac{1}{N^2}(d_k^2 \cdot d_k + \text{Trace}(C_O^{-1} M_1^2)) + O(\frac{1}{N^3})
\]
and we define the local FID at the \(k\)th image as
\[
\text{FID}_k(O) = (d_k^2 \cdot d_k + \text{Trace}(C_O^{-1} M_1^2))
\]
But we must evaluate \(M_1\). We have \(C_O M_1 + M_1 C_O = C_O U\), which is a symmetric Sylvester equation. While this is a linear system in \(M_1\), it is inconveniently large. But, for \(R\) a rotation and \(G\) general, \((RGR^T)^{(1/2)} = R(G)^{(1/2)} R^T\), so
\[
\text{Trace}(R C_O R^T + R C_R R^T - 2(R C_O R^T R C_R R^T)^{(1/2)})
\]
is equal to
\[
\text{Trace}(C_O + C_R - 2(C_O C_R)^{(1/2)})
\]
and the FID is not affected by a rotation in the embedding space, so we can assume without loss of generality that \(C_O\) is diagonal. In this case, we can compute \(M_1\); write \(a_{ij}\) for the \(i, j\)th component of \(A\) etc., and we have \(m_{ij} = c_{ii} u_{ij} / (c_{ii} + c_{jj})\), and computing the local FID at a particular image is relatively straightforward.

\textbf{LFID is a good predictor of human preference:} We show human subjects a pair of images \((A, B)\) of the same scene, and record which is regarded as more realistic. Images have 5000 runs of 10 choices, with anonymous subjects. The scatter plot shows frequency that anonymous subjects. The scatter plot shows frequency that human subjects a pair of images \((A, B)\), and computing the local FID at a particular image is relatively straightforward.

\section*{12 Appendix V: The Gloss Term}

Experimental results suggest strongly our method can remove or enhance glossy effects. Theorem 1 covers global illumination in a diffuse world. The methods used to prove this theorem extend to cover single-bounce radiance from a luminaire for some BRDF models, including natural gloss models, if all BRDF's have the property
\[
[(\omega' \cdot \omega)(\omega' \cdot \omega^\prime)]^\alpha \leq \frac{b(x, \omega', \omega^\prime)}{b(x, \omega, \omega^\prime)} \leq \left[\frac{1}{(\omega' \cdot \omega)(\omega' \cdot \omega^\prime)}\right]^\alpha.
\]
Notice all diffuse surfaces have this property for any finite \(\alpha \geq 0\).

\section*{References}

[1] D. Forsyth and J. Rock, “Intrinsic image decomposition using paragons,” TPAMI, 2022, in press. 3, 4, 5, 9
[2] K. Kim, A. Torii, and M. Okutomi, “Multi-view inverse rendering under arbitrary illumination and albedo,” in European conference on computer vision, 2016, 2
[3] J. T. Barron and J. Malik, “Shape, illumination, and reflectance from shading,” IEEE transactions on pattern analysis and machine intelligence, 2014, 2
[4] A. Tewari, M. Zollohofer, H. Kim, P. Garrido, F. Bernard, P. Perez, and C. Theobalt, “Mofa: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction,” in Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017, pp. 1274–1283. 2
[5] S. Tulsiani, T. Zhou, A. A. Efros, and J. Malik, “Multi-view supervision for single-view reconstruction via differentiable ray consistency,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, 2
[6] A. Kanazawa, S. Tulsiani, A. A. Efros, and J. Malik, “Learning category-specific mesh reconstruction from image collections,” in Proceedings of the European Conference on Computer Vision (ECCV), 2018, 2
[7] Z. Li and N. Snavely, “Learning intrinsic image decomposition from watching the world,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 9039–9048. 2
[8] Z. Li, M. Shafiei, R. Ramamoorthi, K. Sunkavalli, and M. Chandraker, “Inverse rendering for complex indoor scenes: Shape, spatially-varying lighting and SVBRDF from a single image,” in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020. Computer Vision Foundation / IEEE, 2020, pp. 2472–2481. [Online]. Available: https://openaccess.thecvf.com/content_CVPR_2020/html/Li_Inverse_Rendering_for_Complex_Indoor_Scenes_Shape_Spatially-Varying_Lighting_and_CVPR_2020_paper.html 2
[9] Y. Yu and W. A. Smith, “Inverserendering: Learning single image inverse rendering,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, 2
[10] Z. Li, Z. Xu, R. Ramamoorthi, K. Sunkavalli, and M. Chandraker, “Learning to reconstruct shape and spatially-varying reflectance from a single image,” ACM Transactions on Graphics (TOG), 2018, 2
[11] K. Zhang, F. Luan, Q. Wang, K. Bala, and N. Snavely, “Physf: Inverse rendering with spherical gaussians for physics-based material editing and relighting,” in IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021. Computer Vision Foundation / IEEE, 2021, pp. 5453–5462. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2021/html/Zhang_Physf_Inverse_Rendering_With_Spherical_Gaussians_for_Phyhysics-Based_Material_Editing_CVPR_2021_paper.html 2
[12] Z. Wang, J. Philion, S. Fidler, and J. Kautz, “Learning indoor inverse rendering with 3d spatially-varying lighting,” CoRR, vol. abs/2109.06061, 2021. [Online]. Available: https://arxiv.org/abs/2109.06061 2
[13] P.-P. Sloan, J. Kautz, and J. Snyder, “Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments,” ACM TOG - SIGGRAPH, 2002, 2
[14] R. Ramamoorthi, Precomputation-Based Rendering. NOW Publishers Inc, 2009. [Online]. Available: http://graphics.cs.berkeley.edu/papers/Ramamoorthi-PBR-2009-04/ 2
[15] P. Belhumour and D. Kriegman, “What is the set of images of an object under all possible illumination conditions?” Int. Journal Comp. Vision, vol. 28, no. 3, pp. 1–16, 1998, 2, 3
[16] T. Haber, C. Fuchs, P. Bekaer, H.-P. Seidel, M. Goesele, and H. P. A. Lensch, “Relighting objects from image collections,” in 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 627–634. 2
[17] D. Reddy, R. Ramamoorthi, and B. Curless, “Frequency-space decomposition and acquisition of light transport under spatially varying illumination,” in ECCV, 2012, 2
[18] P.-P. Sloan, J. Hall, J. Hart, and J. Snyder, “Clustered principal components for precomputed radiance transfer,” ACM TOG - SIGGRAPH, 2003, 2
[19] R. Ren, Y. Dong, S. Lin, X. Tong, and B. Guo, “Image based relighting using neural networks,” ACM Transactions on Graphics (TOG), vol. 34, no. 4, August 2015. [Online]. Available: https://www.microsoft.com/en-us/research/publication/image-based-relighting-using-neural-networks-3-2

...
