

Supplementary Materials for “Bayesian Time-of-Flight for Realtime Shape, Illumination and Albedo”

Amit Adam, Christoph Dann, Omer Yair, Shai Mazor, Sebastian Nowozin

Abstract—Supplementary materials to the main paper, containing further technical details and additional results.

1 VIDEO DEMONSTRATION

We submit a short video showing the robustness of our approach. In the video we show robust live (real-time) inference of depth, reflectance/albedo and illumination. We demonstrate the effective separation between illumination and reflectance by waving a powerful light projector and noting that the albedo and depth map stay invariant to the changing illumination conditions.

The video was captured using four exposures with the exposure profiles corresponding to the far-range design in Figure 9 in the main paper.

Please note that the RGB stream was shot by the person holding the light projector - and is unrelated to the depth camera stream (we added it for general impression of the scene - it is slightly confusing).

The video is available from the authors’ homepages.

2 INFERENCE DETAILS

For this section we will use the compound parameter vector $\vec{\theta} = [t, \rho, \lambda]^T$, or $\vec{\theta} = [t, \rho, \lambda, t_2, \rho_2]^T$ for the single- and two-paths model. This unifies the notation for all unknown imaging conditions we would like to infer.

The response curve function $\vec{C}(t)$ appearing in the expression for the mean photon response $\vec{\mu}$ (see equation (7) in the main paper), is obtained from calibrated measurements of the actual camera, and then approximated by Chebyshev polynomials of degree sixteen [1]. Because the curves are smooth the Chebyshev approximation is compact yet very accurate and evaluation of $\vec{C}(t)$ also provides the derivatives $\frac{\partial}{\partial t} \vec{C}(t)$ and $\frac{\partial}{\partial t^2} \vec{C}(t)$ for no additional computational cost.

- A. Adam, O. Yair, and S. Mazor were with Microsoft AIT, Haifa, Israel. {email.amitadam, omer.yair, smazor.shai}@gmail.com
- C. Dann is with the School of Computer Science, Carnegie Mellon University, USA. cdann@cmu.edu
- S. Nowozin is with Microsoft Research, Cambridge, UK. Sebastian.Nowozin@microsoft.com

2.1 Maximum Likelihood Estimation (MLE)

The standard maximum likelihood estimate are the imaging conditions t, ρ, λ which maximize the likelihood or equivalently minimize the negative log-likelihood

$$\begin{aligned} & \operatorname{argmin}_{\vec{\theta}} -\log P(\vec{R}|\vec{\theta}) \\ & = \operatorname{argmin}_{\vec{\theta}} \sum_{i=1}^n \left[\frac{(R_i - \mu_i(\vec{\theta}))^2}{2(\alpha\mu_i(\vec{\theta}) + K)} + \frac{1}{2} \log(\alpha\mu_i(\vec{\theta}) + K) \right]. \end{aligned} \quad (1)$$

With this Chebyshev polynomial approximation we can also compute derivatives with respect to $\vec{\theta}$ of the log-likelihood function, and the entire log-likelihood function becomes smooth and twice differentiable.

Solving the three-dimensional minimization problem in Equation (1) with standard Quasi-Newton methods such as L-BFGS [2] is possible but often yields unreasonable result if we do not constrain the parameters. For example, negative values of ρ might have the lowest function value but are physically impossible. Another issue is that the response curves \vec{C} are measured only within a reasonable range. Outside of this range, the Chebyshev approximations have arbitrary behavior which leads to implausible solutions.

We therefore constrain the range of parameters using log-barrier terms

$$\begin{aligned} & \operatorname{argmin}_{\vec{\theta}} \sum_{i=1}^n \left[\frac{(R_i - \mu_i(\vec{\theta}))^2}{2(\alpha\mu_i(\vec{\theta}) + K)} + \frac{1}{2} \log(\alpha\mu_i(\vec{\theta}) + K) \right] \\ & + \sum_j b (\log(\theta_j - \theta_{j,\min}) + \log(\theta_{j,\max} - \theta_j)). \end{aligned} \quad (2)$$

The scalar $b = 10^{-2}$ is a *barrier* coefficient and $\vec{\theta}_{\min}, \vec{\theta}_{\max}$ are the smallest and largest values of each parameter we want to consider. The problem remains twice differentiable and quasi-Newton methods can be applied for finding local minima reliably because any local optima has to occur within the relative interior of the rectangle described by $\theta_{j,\min}$ and $\theta_{j,\max}$.

To find the global optimum, we restart the quasi-Newton method ten times with initialization sampled uniformly in the parameter ranges. For producing labeled training data this is more than sufficient. Even during exposure profile

optimization, experiments on good and mediocre shutter designs have shown that after 10 restarts in 97% of the cases the same global solution was found as with 100 restarts.

2.2 Maximum A-Posteriori Estimation (MAP)

The maximum a-posteriori (MAP) estimate is similar to the maximum likelihood estimate but also considers the prior instead of only the likelihood distribution. We determine the estimate by minimizing the negative log posterior

$$\begin{aligned} & \underset{\vec{\theta}}{\operatorname{argmin}} -\log P(\vec{\theta}|\vec{R}) \\ &= \underset{\vec{\theta}}{\operatorname{argmin}} \sum_{i=1}^n \left[\frac{(R_i - \mu_i(\vec{\theta}))^2}{2(\alpha\mu_i(\vec{\theta}) + K)} + \frac{1}{2} \log(\alpha\mu_i(\vec{\theta}) + K) \right] \\ & \quad - \log p(\vec{\theta}). \end{aligned} \quad (4)$$

Due to the particular choices of twice differentiable prior distributions, we can solve this problem right away with quasi-Newton methods. The log-barrier terms used for the maximum likelihood estimate are now implicitly defined in the prior. In fact, the constrained maximum likelihood estimate in the previous subsection can be understood as MAP estimate with an approximately uniform prior on the ranges $\vec{\theta}_{min}$ to $\vec{\theta}_{max}$.

The advantage of the MAP estimate is that when prior knowledge exists - for example a strong belief on the ambient light intensity - then we may incorporate it. In contrast, the MLE does not encode any preference for certain parameter values.

2.3 Bayesian Posterior Inference

The Bayesian point estimate is motivated by statistical decision theory [3, 4]. The Bayes estimator yields the lowest expected error. Assuming the squared loss function, the estimator is characterized as

$$\hat{\theta}_{\text{Bayes}}(\vec{R}) := \underset{\vec{\theta}}{\operatorname{argmin}} \mathbb{E}_{\vec{\theta} \sim P(\vec{\theta}|\vec{R})} [\|\vec{\theta} - \tilde{\theta}\|_2^2], \quad (5)$$

where $\tilde{\theta}$ are the true but uncertain parameters.

This decision problem has a closed form solution: namely the mean parameters under the marginal posterior distributions. Because the squared loss decomposes over parameters, so does the decision problem.

For example, the Bayes estimator \hat{t}_{Bayes} for depth is given by

$$\hat{t}_{\text{Bayes}}(\vec{R}) = \mathbb{E}[t|\vec{R}] = \int t p(t|\vec{R}) dt. \quad (6)$$

The marginal posterior distribution $p(t|\vec{R})$ can be written in terms of the joint distribution as

$$\begin{aligned} p(t|\vec{R}) &= \int p(\vec{\theta}|\vec{R}) d\vec{\theta} \\ &= \int \frac{p(\vec{R}|\vec{\theta})p(\vec{\theta})}{p(\vec{R})} d\vec{\theta}. \end{aligned}$$

The Bayes estimator \hat{t}_{Bayes} is therefore equal to

$$\mathbb{E}[t|\vec{R}] = \frac{\int t p(\vec{\theta})p(\vec{R}|\vec{\theta}) d\vec{\theta}}{\int p(\vec{\theta})p(\vec{R}|\vec{\theta}) d\vec{\theta}}. \quad (7)$$

One way of computing the Bayes estimator is solving the integrals in the numerator and denominator for all

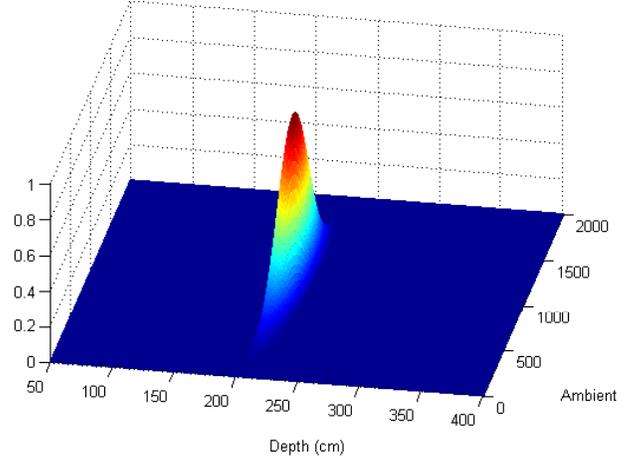


Fig. 1: Slice of the posterior distribution of the single-path model for fixed values of reflectivity ρ . In this case the posterior distribution of the single-path model is dominated by a single Gaussian-like mode.

parameters that we are interested in. We use a state-of-the-art numerical quadrature method [5] for vector-valued integrals over rectangular regions.

However, the numerical quadrature approach is very slow and has numerical issues that yield sub-optimal solutions. We therefore consider an alternative way to compute the Bayes estimators: Monte Carlo using importance sampling [6].

We observed that the posterior distributions of the single-path model are mostly dominated by a few important modes that often have symmetric shape, see Figure 1. The posterior can therefore be approximated well by a mixture of Gaussians. Using importance sampling with a mixture of Gaussians proposal distribution should therefore yield fast convergence to the true Bayes estimator.

The proposal distribution is a mixture of k Gaussians placed at the outputs of k local optima of the MAP problem obtained as described in Section (2.2). The proposal distribution is

$$q(\vec{\theta}) \propto \sum_{i=1}^k p(\vec{\theta}^{(i)}) p(\vec{R}|\vec{\theta}^{(i)}) \mathcal{N}(\vec{\theta}|\vec{\theta}^{(i)}, H^{(i)}), \quad (8)$$

where k is the number of mixture components used and $\vec{\theta}^{(k)}$ are the locations of these mixtures. For the covariance matrices $H^{(k)}$ we use the inverse Hessian of the negative log-posterior (as in a Laplace approximation). Due to the particular choice of twice differentiable priors, the Hessian of the log-posterior are always positive definite in local optima.

We generate samples $\vec{\eta}_1 \dots \vec{\eta}_m$ from q and re-weight each sample by $w_i = \frac{p(\vec{\eta}_i)}{q(\vec{\eta}_i)}$ to account for the errors in the approximation of the posterior by q . These samples are then used to obtain Monte-Carlo estimates of the integrals in equation (7).

We determine the number of samples required to approximate the integrals by the *effective sample size* (ESS) [7, 8],

$$\frac{(\sum_{i=1}^m w_i)^2}{\sum_{i=1}^m w_i^2}. \quad (9)$$

We stop sampling as soon as the ESS exceeds a threshold (usually in the order of 50 – 200). In most cases this threshold is reached with a small number of actual samples. Empirically we observe the importance sampling approach to be much faster and robust in practice than the numerical quadrature approach.

3 TEST-TIME REGRESSION TREE INFERENCE DETAILS

The regression trees approach we described in the main paper has an advantage in terms of flexibility. We used this flexibility to solve several issues that we encountered during development of the prototype camera. While a full description of the issues and their seamless solution within this framework is beyond the scope of this paper, we want to provide one important example.

One notes that all inference is based on the response curve $\vec{C}(\cdot)$ which characterizes the pixel's response to depth. In the physical camera, due to various optical and semiconductor effects, this response curve varies between sensor elements, and this variation is smooth with the position of the pixel on the image sensor. As a result, instead of having a single curve $\vec{C}(\cdot)$ as we described so far, we actually have a set of response curves $\vec{C}_{x,y}(\cdot)$, one for each pixel in the image. Using the regression tree framework, we had a simple seamless solution for this issue as follows:

- During training, instead of sampling responses from a single curve $\vec{C}(\cdot)$, we sample responses from multiple response curves corresponding to different parts of the image. To obtain the label $\hat{t}(\vec{R}_i)$, use slow inference with the actual (position dependent) curve from which \vec{R}_i was sampled.
- We augment the feature vector to include pixel position in addition to the response \vec{R} .
- We extend the leaf model and add linear terms in pixel coordinates x and y .
- Train the regression tree as usual.
- During runtime: just add pixel position to the feature vector used to traverse the tree

This example serves to show the added benefit of a flexible regression mechanism in extending the model to solve new and unexpected problems.

4 MODEL CHECKING AND P-VALUES

This discussion provides additional background for our choice of the posterior predictive p-value used in the main paper. P-values are highly controversial in the field of statistics, in particular for formal hypothesis testing. For example, in [9] Xiaoli Meng writes about the p-value,

“There is perhaps no single notion in statistics, other than the p -value, that has been so widely used and yet so seriously criticized for so long.”

Many works have discussed this controversy and Berger [10] provides a nice formal summary of the issues of disagreement.

For models that are fully observed the choice of a test statistic unambiguously defines the p-value. However, our

model involves unobserved quantities, the imaging conditions t , ρ , and λ , as well as t_2 and ρ_2 for the two-path model. In this case, there is no single p-value to be used and this case is known as “composite null hypothesis” or models with “nuisance parameters”. Also, in this case the test statistic in the classical p-value generally does not depend on the nuisance parameter, which is a drawback as it limits the choice of useful test statistics.

The posterior predictive p-value [9, 11] addresses both problems by integrating the test statistic over the Bayesian posterior of the unknown variables. As test statistic we choose the likelihood of the observation given the unknown parameters, yielding equation (16) in the main paper.

It is easier to understand the usefulness of the likelihood as a test statistic on a model that is fully observed. We visualize such an example with a simple Gaussian mixture model in Figure 2. For the case with unobserved variables the situation is similar except that the distribution changes as a function of the observation \vec{R} .

The posterior predictive p-value has known drawbacks, analyzed in [11, 12]. In particular it is known that it can be too conservative in rejecting the null hypothesis. In our application this implies that we may not detect all detectable deviation from the assumed model. Intuitively the reason for this lack of power is that the p-value is not Bayesian and actually does use the observation twice, once to define the posterior, and once in the computation that defines the p-value, leading to overly optimistic agreement with the model. The so called *partial* posterior predictive p-value [11] does successfully address this issue but it is much more difficult to compute; in fact, although desirable, we have not found a practical method to compute it in our application.

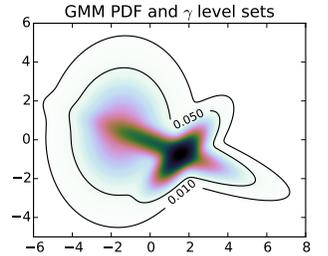


Fig. 2: Invalidation score example. For a Gaussian mixture model level sets at significance levels $\gamma \in \{0.01, 0.05\}$ separate the total probability mass such that outside the level set the integrated probability is γ . Samples outside these sets are rejected.

5 EXPOSURE PROFILES DESIGN DETAILS

To optimize the design objective (19) in the main paper, we use a simulated annealing approach as follows. Let us abbreviate the objective function (19) as

$$f(\mathbf{Z}) = \mathbb{E}_{t,\rho,\lambda} \mathbb{E}_{\vec{R} \sim P(\vec{R}|t,\rho,\lambda,\mathbf{Z})} \left[\ell(\hat{t}(\vec{R}), t) \right]. \quad (10)$$

We introduce an auxiliary Gibbs distribution, parametrized by a temperature $T > 0$,

$$r(\mathbf{Z}, T) \propto \exp\left(-\frac{1}{T}f(\mathbf{Z})\right). \quad (11)$$

We use a sequence of temperature parameters that is slowly decreased for a finite number of steps, that is, $T_0 > T_1 > \dots > T_K$, starting from an initial temperature $T_0 = T_{\text{start}}$ down to a final temperature $T_K = T_{\text{final}}$. The smaller T gets, the more peaked the distribution $r(\cdot, T)$ becomes around

the minimum of f . Given a Markov chain sampler on r , this approach converges to the global minimum of f .

We first discuss the Markov chain that we use, then give details about the temperature schedule.

5.1 Markov Chain

To account for the sparsity constraints on \mathbf{Z} , our Markov chain uses an augmented state space [13] to avoid measure-theoretic difficulties of asserting reversibility in the context of changing dimensionality [14].

We decompose \mathbf{Z} into a binary matrix $\mathbf{B} \in \{0, 1\}^{m \times n}$ and a value matrix $\mathbf{V} \in \mathbb{R}^{m \times n}$ with $Z_{ji} = B_{ji}V_{ji}$. This allows us to easily set weights to zero by setting $B_{ji} = 0$ and have the reversible proposal readily available by setting $B_{ji} = 1$. Our MCMC sampler is a reversible Metropolis-Hastings sampler and consists of the following transition kernels (moves):

- 1) *Move mass*: Choose two matrix entries V_{ji}, V_{kl} randomly (uniform) and move a uniformly sampled value from one entry to another such that their total value stays the same and both are still positive. This kernel is reversible with itself.
- 2) *Swap values*: Choose two matrix entries W_{ji}, W_{ji} randomly (uniform) and swap their values V and binary indicator value B . This kernel is reversible with itself.
- 3) *Set a weight to zero*: Choose a matrix entry with $B_{ji} = 1$ randomly (uniform) and set it to zero. This kernel is reversible with the following kernel.
- 4) *Set a weight to nonzero*: Choose a matrix entry with $B_{ji} = 0$ randomly (uniform) and set its binary indicator value to one. This kernel is reversible with the previous set-to-zero kernel.
- 5) *Perturb weight value*: Choose a matrix entry $V_{ji} = 0$ randomly (uniform) and rescale its value with a log-normal sampled factor. This kernel is reversible with itself.
- 6) *Scale all weight values*: Rescale all values \mathbf{V} with a log-normal sampled scalar. This kernel is reversible with itself.

The above kernels are combined with the following probabilities: 20% for the *move mass* kernel; 20% for the *swap values* kernel; 10% for the *set-to-zero* and *set-to-nonzero* kernels, each; 30% for the *perturb weight* kernel; 10% for the *global scaling* kernel.

5.2 Temperature Schedule

For simulated annealing we use a geometric temperature schedule [15], with the temperature at iteration k being

$$T_k = T_{\text{start}} \beta^k,$$

where we use the initial temperature $T_{\text{start}} = 20$ and a target temperature of $T_{\text{final}} = 0.01$, so that

$$\beta = \exp\left(\frac{1}{K} [\log T_{\text{final}} - \log T_{\text{start}}]\right).$$

This leads to the schedule as shown in Figure 3. We typically use a $K = 20,000$ or $K = 100,000$ iterations.

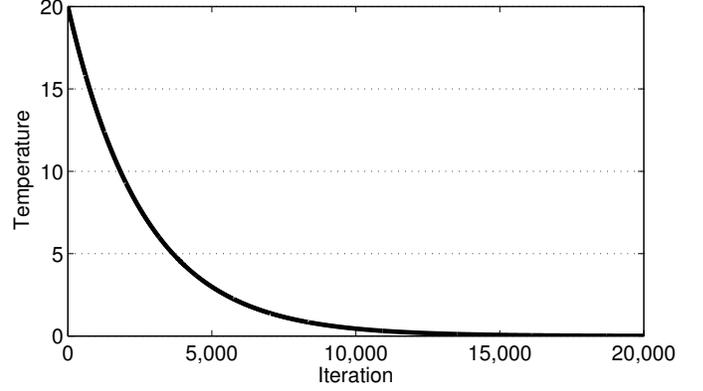


Fig. 3: Simulated annealing schedule used during shutter profile design optimization, here with $K = 20,000$ iterations.

6 TIME OF FLIGHT SIMULATION DETAILS

We now discuss details of the physically accurate light simulation that we use to simulate multipath phenomena. First we recap the basis of both the bidirectional path tracer (BDPT) and the Metropolis light transport (MLT) algorithms [16, 17] and then provide information about the variance reduction techniques we use.

6.1 Light Transport Formulation

Assuming a geometric light model where light travels in straight lines and only interacts with surfaces, the measured light intensity at a pixel in a static scene without active illumination can be formulated as a path integral. This integral accumulates the intensity from light paths $x_0, x_1 \dots x_{k+1}$ that start in a point x_0 on an emitting surface and end in a point x_{k+1} on the pixel's sensor surface. The intermediate nodes of this path $x_1 \dots x_k$ are surfaces in the scene. The integral can be formulated (see [17] for details) as

$$\sum_{k=0}^{\infty} \int_{\mathcal{M}^{k+1}} L_e(x_0 \rightarrow x_1) G(x_0 \leftrightarrow x_1) \prod_{i=1}^k (f(x_{i-1} \rightarrow x_i \rightarrow x_{i+1}) G(x_i \leftrightarrow x_{i+1})) L_s(x_k \rightarrow x_{k+1}) dA(x_0) \dots dA(x_{k+1}). \quad (12)$$

In this equation,

- \mathcal{M} is the set of all surfaces in the scene including emitters and the camera sensor and A is the area measure on \mathcal{M} ;
- $L_e(x_0 \rightarrow x_1)$ is a function representing emitters. It is proportional to the light that is emitted from point x_0 in the direction of x_1 . It takes only non-zero values if x_0 is on emitter surfaces;
- $L_s(x_k \rightarrow x_{k+1})$ is the equivalent of L_e for the sensor. L_s specifies how sensitive the sensor is for photons arriving at x_{k+1} from the direction of x_k .
- $f(x_{i-1} \rightarrow x_i \rightarrow x_{i+1})$ is the bidirectional scattering distribution function (BSDF) describing how much light is scattered at surface point x_i in direction x_{i+1} of an incoming ray from the direction of x_{i-1} ;
- $G(x_i \leftrightarrow x_{i+1}) = V(x_i \leftrightarrow x_{i+1}) \frac{|\cos \phi_i \cos \phi_{i+1}|}{\|x_{i+1} - x_i\|^2}$ is the throughput of a differential beam between $dA(x_i)$

and $dA(x_{i+1})$. $V(x_i \leftrightarrow x_{i+1})$ is an indicator function for mutual visibility of x_i and x_{i+1} , which means V is zero if the direct path between the two inputs is blocked, otherwise 1; The variables ϕ_i, ϕ_{i+1} denote the angle between the beam and the surface normals at x_i and x_{i+1} .

The observed response in a specific pixel of our time-of-flight camera from the emitted light pulse can be modelled by extending the path integral formulation above to

$$R_{\text{active}} = \int \sum_{k=0}^{\infty} \int_{\mathcal{M}^{k+1}} P(u) L_e(x_0 \rightarrow x_1) G(X_0 \leftrightarrow x_1) \prod_{i=1}^k (f(x_{i-1} \rightarrow x_i \rightarrow x_{i+1}) G(x_i \leftrightarrow x_{i+1})) L_s(x_k \rightarrow x_{k+1}) S_j(u + t_l) dA(x_0) \dots dA(x_{k+1}) du. \quad (13)$$

We additionally integrate over time u and include the intensity of the emitted pulse $P(t)$ as well as the shutter function $S_j(t + t_l)$. The time delay $t_l = c l$ of emitted light arriving at the sensor is the total path length,

$$l = \sum_i \|x_{i+1} - x_i\|,$$

times the speed of light c . All terms involving time can be group together into the expression

$$\int P(u) S_j(u + t_l) du = \frac{C_j(t_l)}{d(t_l)}$$

that only depends on the time delay t_l corresponding to total path length. It corresponds to the curve C_j without the decay of light $d(t_l)$ due to distance l (The decay of light is already accounted for in the G terms of the integral). The measured response is then

$$R_{\text{active}} = \sum_{k=0}^{\infty} \int_{\mathcal{M}^{k+1}} \frac{C_j(t_l)}{d(t_l)} L_e(x_0 \rightarrow x_1) G(X_0 \leftrightarrow x_1) \prod_{i=1}^k (f(x_{i-1} \rightarrow x_i \rightarrow x_{i+1}) G(x_i \leftrightarrow x_{i+1})) L_s(x_k \rightarrow x_{k+1}) dA(x_0) \dots dA(x_{k+1}). \quad (14)$$

This formulation is identical to the path integral Equation (12) but with the additional $C_j(t_l)/d(t_l)$ term.

We modified the bidirectional path tracer (BDPT) algorithm [16] and the Metropolis light transport (MLT) algorithm [17] in the Mitsuba renderer [18] to produce a weighted set of samples $\{(w_i, L_i, t_i)\}_{i=1, \dots, N}$ of the path integral in Equation (12). The weight of the path sample is w_i , L_i is the number of edges and t_i is the time corresponding to the total path distance. We can generate samples of R_{active} by

$$\sum_{i=1}^N \frac{w_i}{d(t_i)} C_j(t_i).$$

Considering all shutters C_1, C_2, \dots and adding constant ambient light τ to account for R_{ambient} , we may obtain realistic estimates of the mean response vector

$$\vec{\mu} = \tau \vec{A} + \sum_{i=1}^N \frac{w_i}{d(t_i)} \vec{C}(t_i). \quad (15)$$

6.2 Variance Reduction

The BDPT and MLT rendering techniques are Monte Carlo methods and therefore estimates obtained from them will have a Monte Carlo variance. This variance does not originate with the underlying mechanism that is being simulated, but is due to the finite number of samples that are used for estimation. Whereas normal light transport rendering in computer graphics applications is targeted at estimating mean intensities in three spectral bands (RGB), we are instead interested in the time-of-flight density. Because it is a function instead of a small number of values, it is more difficult to obtain reliable estimates of this function.

To improve the accuracy of our estimate with the given time and memory constraints, we use two variance reduction techniques: *stratification* and *priority sampling*.

The starting point for both methods is a stream of weighted samples (w_i, L_i, t_i) being generated for each pixel.

6.2.1 Stratification

Stratification is a classic variance reduction technique based on prior knowledge of subpopulations which have lower within-population variation. It works by breaking up the estimation problem into one estimation problem per subpopulation and combining the individual estimates into one joint estimate. This reduces the variance of the joint estimate compared to lumping all subpopulations together in only one population and sampling and estimating from only this one population [6, Section 5.5].

We *stratify* the incoming stream of samples into two sets. The first stratum is the set of samples $L_i = 2$ and the second stratum is the set of samples for which $L_i > 2$. For both sets we keep an equal number of samples, typically a few thousand. The following priority sampling is then performed on each of these two sets separately.

6.2.2 Priority Sampling

The output of the simulation is a set of path samples (w_i, L_i, t_i) for each pixel. For large image sizes this can require tens of gigabytes of storage. In particular for the MLT sampler many of these samples contain partially redundant information due to correlated sampling via runnign a Markov chain, and storing all of them is wasteful in terms of storage. Because for MLT the samples are correlated in time, one really does need to generate a large number of samples to get good estimates; it is only the storage that is wasteful. For BDPT the samples are uncorrelated but we can still improve estimates by replacing low-weight samples with more important ones due to the inefficiencies of importance sampling.

Typically, in general Markov chain Monte Carlo simulations the samples are unweighted and one can simply *thin* the samples by taking, for example, every 10th sample only, or by using reservoir sampling to keep a random subset. Here, however, the samples from both BDPT (importance sampling) and MLT (Markov chain simulations) are weighted, and this naive strategy—while still valid in terms of providing an unbiased estimate—yields a high variance estimate because it discards important samples with high weights.

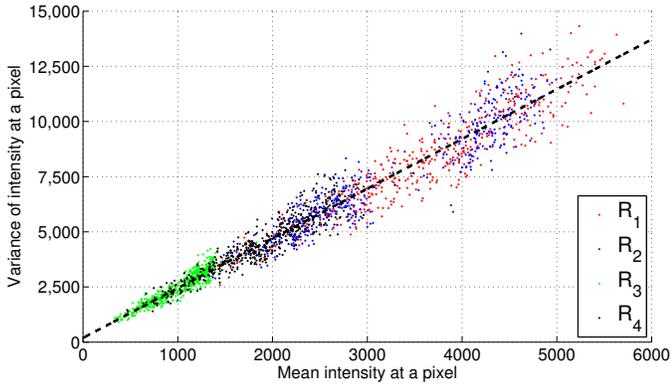


Fig. 4: Validation of the noise model (8). The variance of the response is approximately a linear function of the mean response [20]. We show samples from four responses in different colors and superimpose the least squares fit.

To obtain low-variance estimates from few samples we use *priority sampling* [19], a close to optimal method addressing the above subsampling problem. Intuitively, priority sampling generalizes reservoir sampling to the case of weighted samples. It processes the input sample stream one sample at a time and keeps a fixed number of samples with adjusted weights. The weights are adjusted such that the estimate of any subset sum is unbiased, and the variance of weight subset sums is almost optimal uniformly over all possible subsets.

We use priority sampling to thin the two sample streams for each stratum and after rendering is finished we simply output the kept samples and adjusted weights.

Overall we found that the bidirectional path tracer (BDPT) often produces better results with lower variance and all simulation results in the main paper are obtained by running BDPT with 8192 samples per pixel.

7 NOISE MODEL VALIDATION

To verify the noise model we assumed in Section 3 (equation 8) of the main paper, we used the experimental setup as described in Section 9.2 of the main paper: we sample 500 random pixels and capture 200 frames from a static scene. We then measure the variance of each pixel’s response, as well as estimate the mean response; this provides empirical data about the actual noise present in the input signal.

Figure 4 shows the noise model results in the form of a scatter plot of the variance of responses versus their mean. The data clearly validates the assumed noise model (8) from the main paper, and shows that the signal-dependent Poissonian shot noise component dominates except for very small intensities.

8 ACCURACY OF ALBEDO AND AMBIENT LIGHT LEVEL ESTIMATES

In order to get a sense of the accuracy with which we are able to recover albedo and ambient light level, we present another experiment as follows. We sampled responses obtained under the following imaging conditions: depth is uniform between 70cm and 370cm, ambient light level is uniform between $\lambda = 0$ and $\lambda = 20000$, and reflectivity is

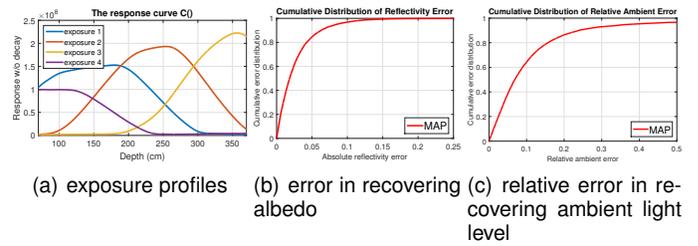


Fig. 5: Accuracy of albedo and ambient light level recovery

uniform between 0% and 100%. The responses are sampled from the response curve shown in part (a) of Figure 5.

We then ran our MAP inference procedure, and compared the point estimates of albedo and ambient light level, with the actual values. Parts (b) and (c) of Figure 5 show the cumulative distribution of errors. For albedo we show the absolute error, and for ambient light level we show the relative error.

We note that the median albedo error is less than 3%, and that the median relative error in recovering ambient light level is around 7%. For the applications we considered this accuracy level is higher than actually required.

9 NO-MULTIPATH SIMULATION EXPERIMENT

In the main paper we discussed inference under two imaging models - namely the single-path and the two-path models. In real world scenes multipath is a common phenomenon. Neither of these two models accurately describes what happens under realistic multipath conditions. Nevertheless, as we saw in Section 9.7 and in Table 1 in the main paper, the (inappropriate) two-path model on the average reduces the median error by 40% with respect to the (even more inappropriate) single-path model.

An interesting experiment¹ would be to evaluate these two models under scene conditions where no multipath exists. We took the five scenes used in Section 9.7 of the main paper, and removed any multipath light paths from the simulator output. In this manner we obtained responses that correspond to direct paths only. We then ran the same SP-Bayes and TP-Bayes inference procedures and obtained the results shown in Table 1.

What do we expect to see in comparison to what we got in Table 1 of the main paper?

- 1) The single-path errors should decrease significantly. A single-path imaging model is completely wrong in the presence of multipath - this is the reason the errors of SP-Bayes were large in the main paper. In contrast, in the current experiment the single-path imaging model is a perfect description of the way responses are generated and therefore the errors are much lower (resulting from shot-noise only). Indeed, the average decrease in median error of the SP-Bayes model is 68% (e.g. 13.46cm going down to 2.49cm for the first Sitting Room scene).
- 2) Likewise, single-path invalidation score should not have any reason to detect model misfit in the new

1. We thank an anonymous reviewer for suggesting this experiment

Scene	Model	Absolute error quantile (cm)		
		25%	50%	75%
Sitting Room	SP-Bayes	1.09	2.49	5.25
	TP-Bayes	1.47	3.22	6.48
Breakfast Room	SP-Bayes	1.15	2.63	5.67
	TP-Bayes	2.13	4.45	8.66
Kitchen Nr 2	SP-Bayes	1.22	2.85	6.56
	TP-Bayes	2.10	4.52	9.09
Country Kitchen	SP-Bayes	1.66	3.97	9.01
	TP-Bayes	2.84	5.89	11.48
Wooden Staircase	SP-Bayes	1.18	2.62	5.07
	TP-Bayes	1.70	3.64	6.79

TABLE 1: Predictive performance of the Bayesian single-path (SP) and two-path (TP) models on rendered simulation data in which only the first-bounce response is used and all multipath effects are removed from the scene.

experiment. When multipath is present we have indication of responses being invalid under the assumed model when we get a low γ invalidation score. When the responses fit the model, the γ score should be uniform.

Figure 6 demonstrates this on the staircase scene shown in Figure 12. In part (a) of Figure 6 we show the invalidation score under the presence of multipath. The structure of the scene is evident with large areas of low γ values - indicating misfit to single-path model. In part (b) we show the invalidation score under the current experiment where no multipath is present: indeed γ is uniform. (The center spot in the scene is due to specular pixel saturation).

- 3) How should the two-path error behave now ? In comparison to the experiment in the main paper, not much can be said: the two-path model is not a correct description of the responses under multipath, but it is also not a correct description of the responses under the current experiment which has zero multipath. In a sense it overfits the imaging model under the current experiment.

Because the two-path model is wrong when no multipath exists, and in contrast the single-path model is perfectly correct, the two-path error in the current experiment should be greater than the single path error. Indeed, looking at Table 1 in the current experiment, the average increase in median error (with respect to running a single-path model) is 49% (e.g 2.49cm going up to 3.22cm in the first Sitting Room scene).

We note that an environment with absolutely no multipath as in this interesting experiment, does not exist. A possible approach to dealing with real environments would be to use a mixture model between the single and two-path imaging models. The prior on which model holds would have to be somehow chosen based on knowledge of the environment.

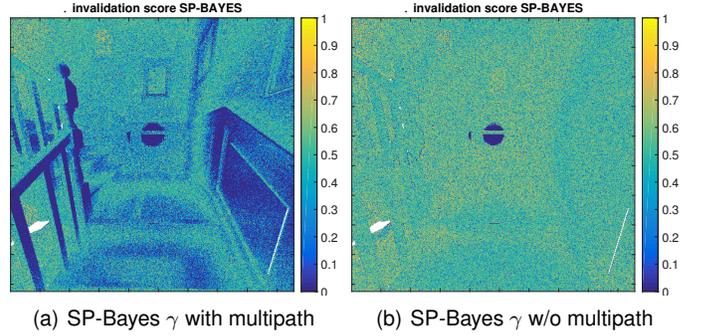


Fig. 6: Effect of removing multipath from simulation on the γ invalidation score. Without multipath the SP-Bayes model is an accurate model of the world and the invalidation score becomes close to uniformly distributed.

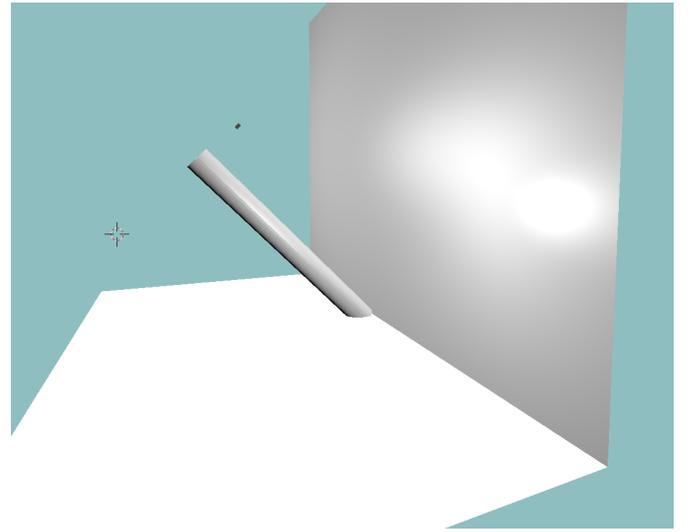


Fig. 7: The scene used for designing a multipath-resistant exposure profile.

10 THE SCENE USED AS AN EXTENDED GENERATIVE MODEL

As described in Section 8.3 in the main paper, we may use a realistic light transport simulation as a more complex generative model Q in order to generate responses containing also multipath components. We used this method to design a multipath-resistant exposure profile. The scene we used is depicted in Figure 7. The camera is pointed towards a reflective wall, and the responses sampled from the cylinder and the floor contain both a direct component and multipath components. We used two copies of this scene at two different scales. The exposure profile designed using this scene was tested on a very different test scene as described in Section 9.5.

11 ROBUST MODEL INVALIDATION

To demonstrate the performance of the proposed invalidation mechanism we use the same real experimental data as in Section 9.6 in the main paper. We compute the γ score for both scenes and for the SP-Bayes and TP-Bayes models.

The results are shown in Figure 8 and show that robust invalidation is possible for this scene. In the next section we also report γ score images for simulation data.

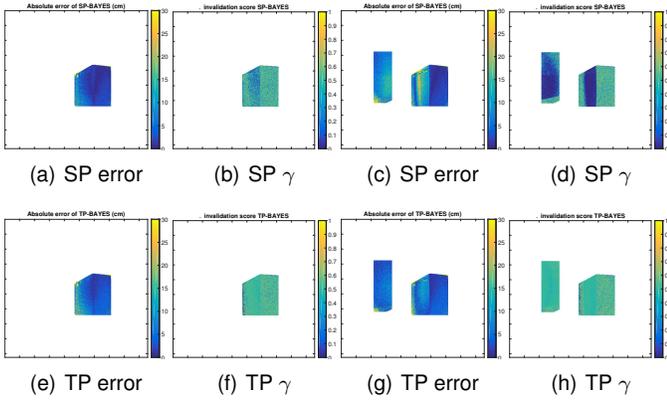


Fig. 8: Robust invalidation of measurements that violate our model assumptions, as explained in Section 6 of the main paper. In the top row we show the single path model SP-Bayes depth inference errors on real data as with the previous experiment in Section 9.6, whereas the bottom row shows the TP-Bayes model. The γ invalidation score robustly highlights pixels affected by multipath in (d) (values close to zero in dark blue) leading to large depth errors in (c). By thresholding the γ score at a suitably chosen threshold these measurements can be excluded from further processing. Note that the two-path (TP) model handles multipath and the γ score in (h) does not invalidate affected pixels.

12 ADDITIONAL RENDERED RESULTS

In Section 9.7 of the main paper we omitted four scenes for space reasons; the results are provided here in Figure 9, Figure 10, Figure 11 and Figure 12 and qualitatively agree with the results shown in the main paper.

12.1 Example Problem with γ

We now give an example where we suffer high depth errors but have no operational method to recognize this: the estimated uncertainty $\hat{\sigma}$ is not overly large, and the invalidation score γ does not indicate deviation from the model.

The situation occurs in the scene shown in Figure 10, and we highlight the area in Figure 13. In essence the problem is due to complex multipath phenomena involving diffuse multipath and multiple bounces as can be seen from Figure 13(b), which leads to an observed response that is well within a high-probability region of the assumed two-path model (no invalidation) and has a strong direct response component (leading to low $\hat{\sigma}$).

In order to improve depth accuracy in regions such as the highlighted one, several approaches are relevant. We are currently considering temporal integration of the observed response and an imaging model that does not assume conditional independence among different sensor elements. For example, many surfaces are planar and recognizing deviations from planarity over multiple pixels could potentially provide a strong cue to recognize and correct for multipath interference.

13 INTRINSIC IMAGING

In this section we provide further examples showing that in addition to depth we also obtain a high-quality reconstruction of effective albedo and ambient light.

In Figure 14 and 15 we show for two scenes our posterior mean estimates of all three unknowns for the single-path (SP-Bayes) and two-path (TP-Bayes) models.

For effective albedo ρ the ground truth is shown in the first row (third column, Figures 14(c) and 15(c)). Our estimates have overall excellent agreement with this ground truth as shown in Figures 14(g) and 15(g) for the single path model and in Figures 14(k) and 15(k) for the two path model. Minor overestimation is visible in regions affected by multipath, for example on the table and in the corner of the room in Figure 14(g). The TP-Bayes model improves this estimate as shown in Figure 14(k), however some multipath artifacts remain.

For ambient λ the ground truth is shown in the first row, fourth column - Figures 14(d) and 15(d). Again our estimates are accurate as shown in Figures 14(h) and 15(h) for the single-path model, and Figures 14(l) and 15(l) for the two-path model. Minor artifacts are in very dark regions where little light is reflected (e.g. within the picture on the wall in Figure 14(a)) and regions affected by multipath (e.g. the nearby floor in Figure 15(l)).

REFERENCES

- [1] L. N. Trefethen, *Approximation Theory and Approximation Practice*. SIAM, 2012.
- [2] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Mathematical programming*, vol. 45, no. 1-3, pp. 503–528, 1989.
- [3] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*. Springer, 1985.
- [4] C. P. Robert, *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer, 2001.
- [5] A. C. Genz and A. Malik, "An adaptive algorithm for numerical integration over an n-dimensional rectangular region," *Journal of Computational and Applied Mathematics*, vol. 6, no. 4, pp. 295–302, 1980.
- [6] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo method*. John Wiley & Sons, 2011.
- [7] J. S. Liu, "Metropolized independent sampling with comparisons to rejection sampling and importance sampling," *Statistics and Computing*, vol. 6, no. 2, pp. 113–119, 1996.
- [8] A. Kong, "A note on importance sampling using renormalized weights," Department of Statistics, University of Chicago, Tech. Rep., 1992.
- [9] X.-L. Meng, "Posterior predictive p-values," *The Annals of Statistics*, pp. 1142–1160, 1994.
- [10] J. O. Berger *et al.*, "Could fisher, jeffreys and neyman have agreed on testing?" *Statistical Science*, vol. 18, no. 1, pp. 1–32, 2003.
- [11] M. J. Bayarri and J. O. Berger, "P values for composite null models," *Journal of the American Statistical Association*, vol. 95, no. 452, pp. 1127–1142, 2000.
- [12] J. M. Robins, A. van der Vaart, and V. Ventura, "Asymptotic distribution of p values in composite null models," *Journal of the American Statistical Association*, vol. 95, no. 452, pp. 1143–1156, 2000.
- [13] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*. Springer, 2001.
- [14] P. J. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, 1995.
- [15] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, 1983.

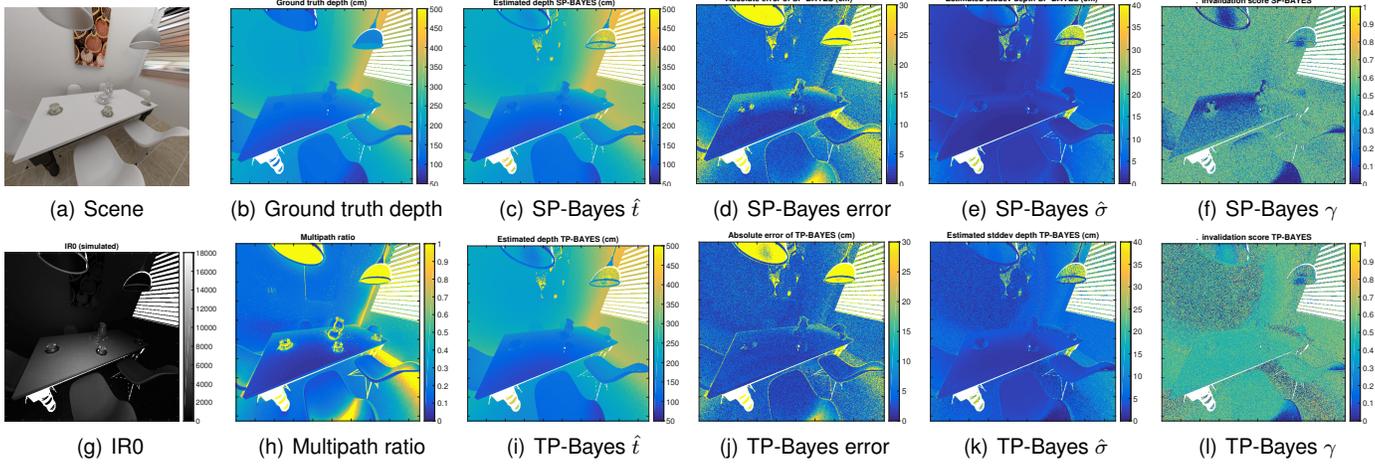


Fig. 9: Rendered simulation (scene adapted from “The Breakfast Room” by Wig42, licensed CC-BY from blendswap.com). Significant multipath error reductions due to the two-path model are visible (wall, table, floor, chairs). Specular surfaces (lamp shade) remain problematic.

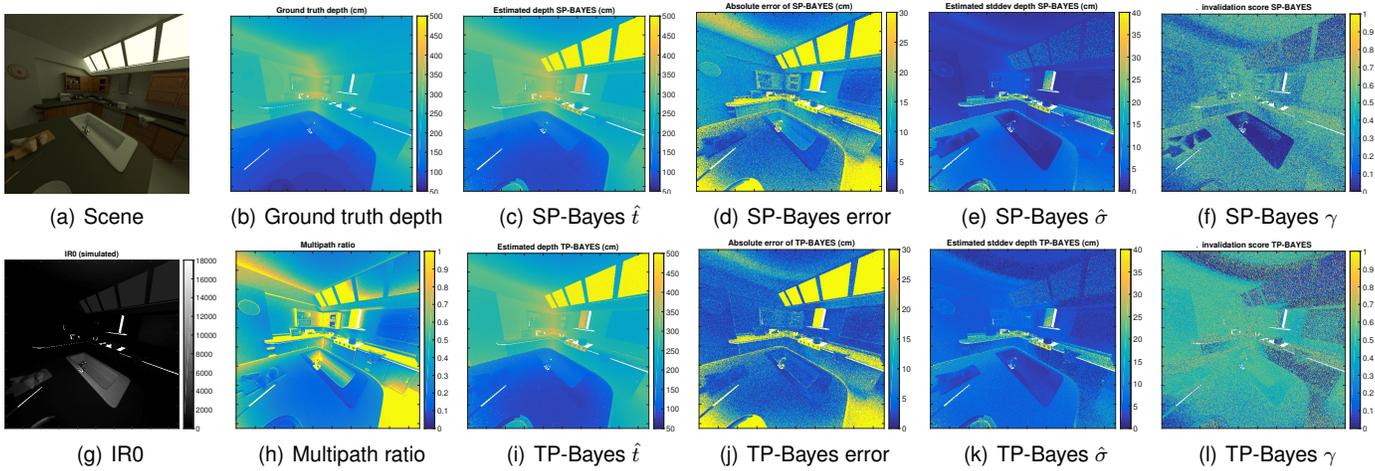


Fig. 10: Rendered simulation (scene adapted from “Kitchen Nr 2” by oldtime, licensed CC-BY from blendswap.com).

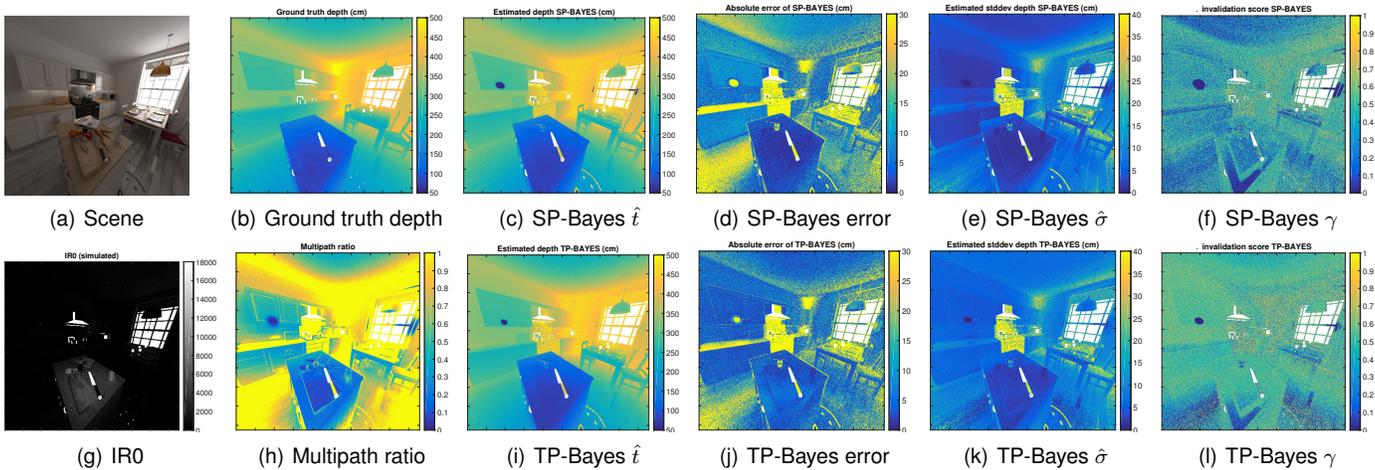


Fig. 11: Rendered simulation (scene adapted from “Country-Kitchen Cycles” by Jay-Artist, licensed CC-BY from blendswap.com). Overall strong multipath error reduction across the scene but higher overall single-frame jitter due to the strong ambient lighting.

[16] E. Veach and L. Guibas, “Bidirectional estimators for light transport,” in *Fifth Eurographics Workshop on Rendering*, 1994.
 [17] E. Veach and L. J. Guibas, “Metropolis light transport,”

in *Proceedings of the ACM SIGGRAPH Conference*, 1997, pp. 65–76.
 [18] W. Jakob, “Mitsuba renderer,” 2010, <http://www.mitsuba-renderer.org>.

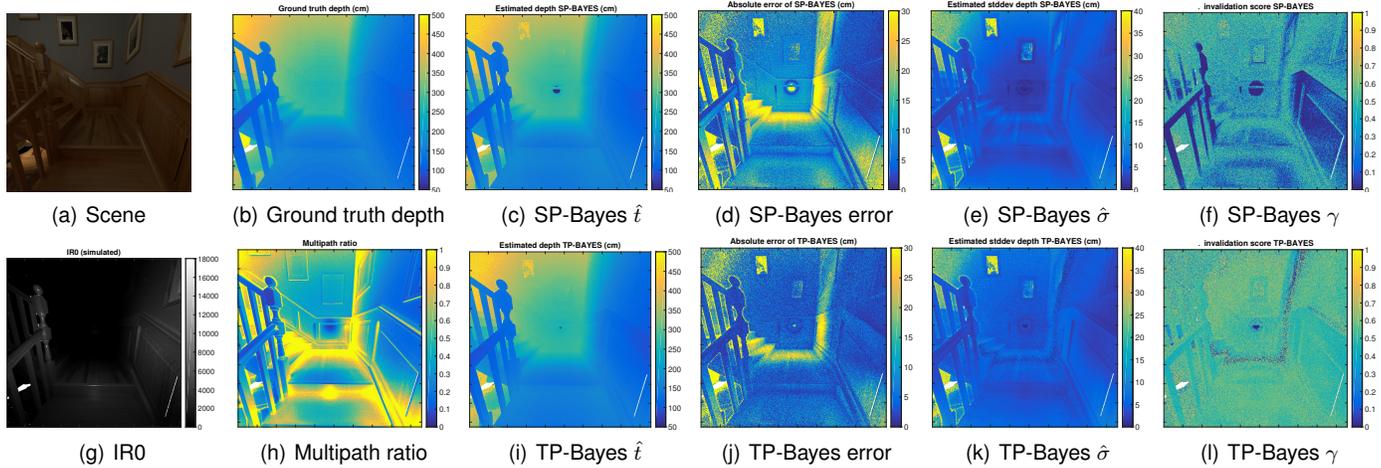
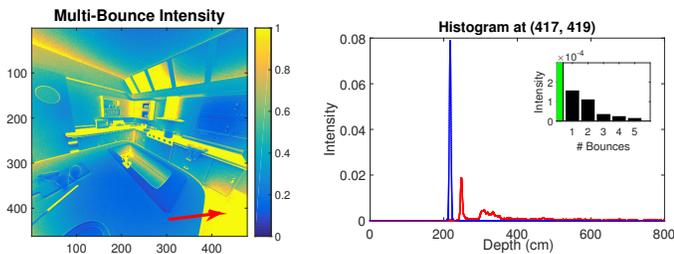


Fig. 12: Rendered simulation (scene adapted from “The Wooden Staircase” by Wig42, licensed CC-BY from blendswap.com).



(a) Problematic area with high depth error (see Fig. 10(j)) that is not recognized by $\hat{\sigma}$ (see Fig. 10(k)), nor by γ (see Fig. 10(l)).
 (b) Intensity measure at the high-lighted pixel showing complex multipath phenomena with more than two-path responses and large diffuse multipath.

Fig. 13: Explaining a failure of $\hat{\sigma}$ and γ to recognize areas of large depth errors.

[19] N. Duffield, C. Lund, and M. Thorup, “Priority sampling for estimation of arbitrary subset sums,” *Journal of the ACM (JACM)*, vol. 54, no. 6, p. 32, 2007.
 [20] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, “Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data,” *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1737–1754, 2008.

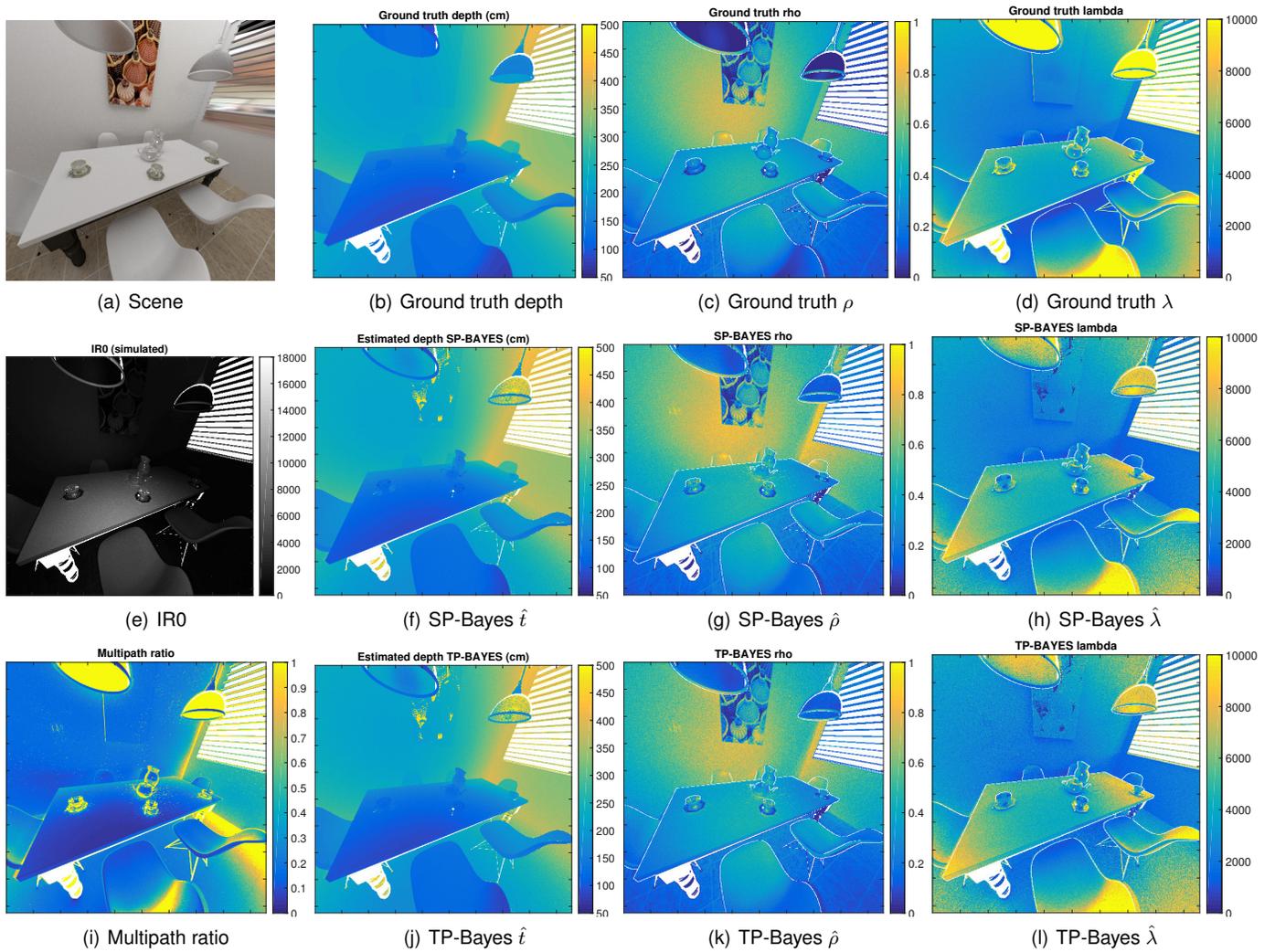


Fig. 14: Intrinsic image decomposition, rendered simulation (scene adapted from “The Breakfast Room” by Wig42, licensed CC-BY from blendswap.com).

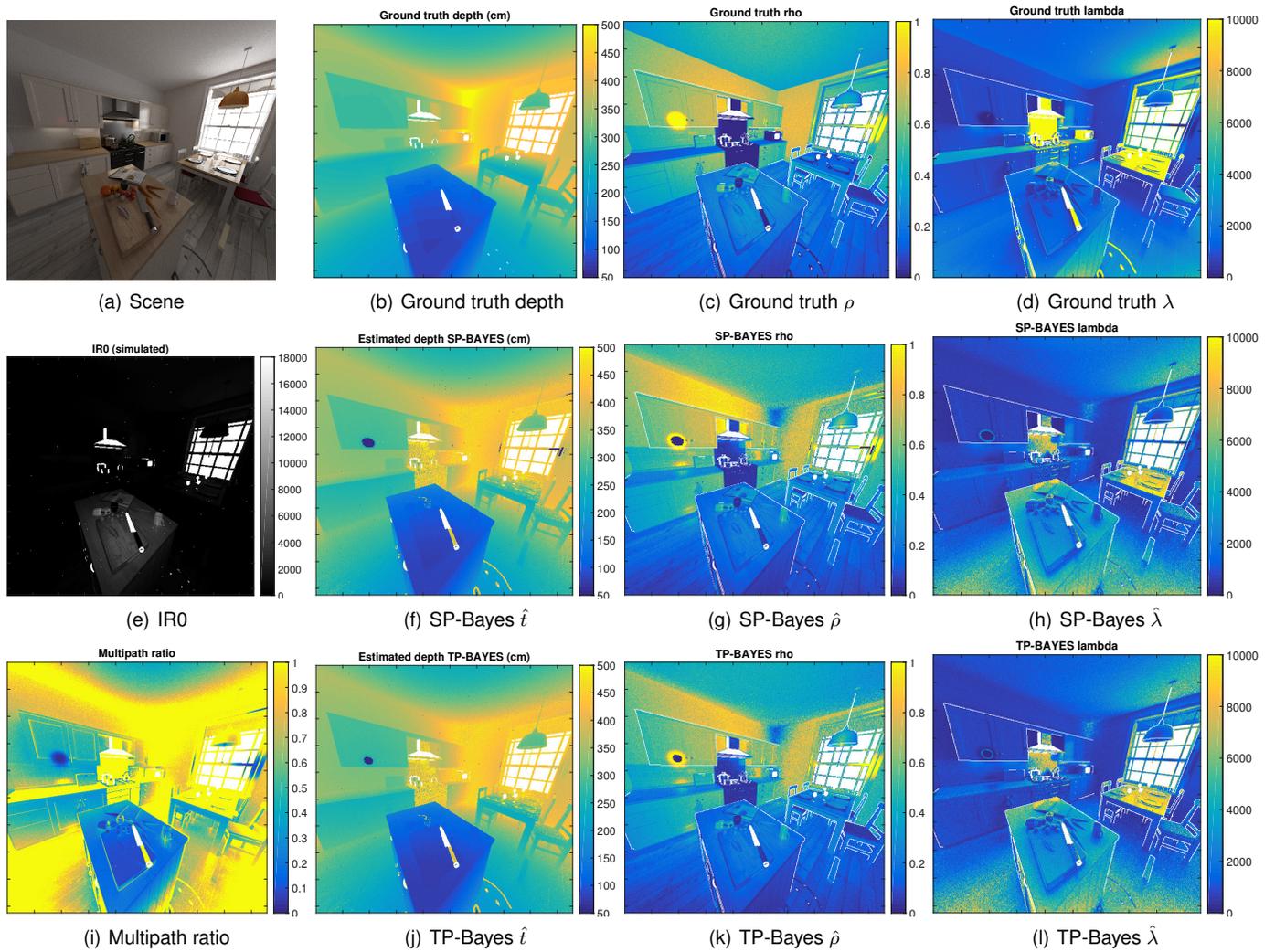


Fig. 15: Intrinsic image decomposition, rendered simulation (scene adapted from “Country-Kitchen Cycles” by Jay-Artist, licensed CC-BY from blendswap.com).