Separable Subsurface Scattering: Expanded Technical Report

Technical Memo ATVI-TR-19-02

Jorge Jimenez[1,2], Károly Zsolnai[3], Adrian Jarabo[1], Christian Freude[3], Thomas Auzinger[3], Xian-Chun Wu[2], Javier von der Pahlen[2], Michael Wimmer[3] and Diego Gutierrez[1]

[1]Universidad de Zaragoza [2]Activision-Blizzard [3]Vienna University of Technology

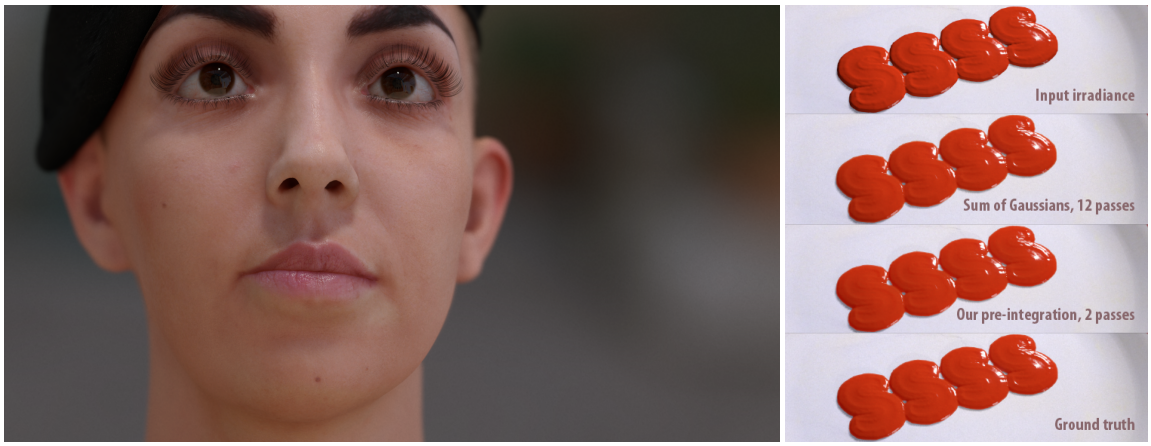


Figure 1: Real-time results of our method for simulating translucent materials (skin on the left, ketchup on the right). Our separable subsurface-scattering method enables the generation of these images using only two convolutions (versus 12 in the sum-of-Gaussians approach [dLE07, JSG09]) and seven samples per pixel, while featuring quality comparable with the current state of the art, at a fraction of its cost. It can be implemented as a post-processing step and takes only 0.39 ms on the PS4 at 1080p for typical portrait shots, which makes it highly suitable for challenging real-time scenarios.

This technical report extends the content of [JZJ*15], by adding some additional explanations and details, and including new tests with more complex multilayered materials. In case of use of the content of this document, please cite the original publication.

Abstract

In this paper we propose two real-time models for simulating subsurface scattering for a large variety of translucent materials, which need under 0.4 milliseconds per frame to execute (for typical portrait shots). This makes them a practical option for real-time production scenarios. Current state-of-the-art, real-time approaches simulate subsurface light transport by approximating the radially symmetric non-separable diffusion kernel with a sum of separable Gaussians, which requires multiple (up to twelve) 1D convolutions. In this work we relax the requirement of radial symmetry to approximate a 2D diffuse reflectance profile by a single separable kernel.

We first show that low-rank approximations based on matrix factorization outperform previous approaches, but they still need several passes to get good results. To solve this, we present two different separable models: a) kernel pre-integration, which yields a high-quality diffusion simulation; and b) an artist friendly model, which offers an attractive trade-off between physical accuracy and artistic control. Both reduce rendering subsurface scattering to just two 1D convolutions, while delivering results comparable to techniques with higher cost. Using importance-sampling and our stochastic sampling strategies, that rotate part of the kernel, only seven samples per pixel are required, even for high resolution imaging or close ups where our separable kernels might yield artifacts. We additionally show that separable profiles built with only two 1D Gaussians are able to perceptually match Monte

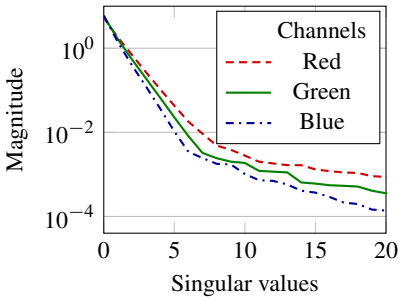


Figure 2: Decay of the singular values in the singular value decomposition of a diffuse reflectance profile used to simulate subsurface scattering in skin. Only the components associated to the first few singular values contribute appreciably to the reconstruction of the profile, making a low-rank approximation feasible.

Carlo simulation and multipole diffusion references. Our methods can be implemented as simple post-processing steps without intrusive changes to existing rendering pipelines.

1. Introduction

The accurate depiction of translucent materials is an important but challenging topic in the motion picture and video game industries. Rendering realistic subsurface scattering (SSS) implies simulating how light travels and scatters inside translucent media, which is an expensive process. While offline rendering scenarios can afford longer computation times, real-time applications, such as video games, impose severe time constraints, often leading to the exclusion of subsurface scattering and translucency effects. This in turn hinders the level of realism that can be achieved.

One of the most common approaches to compute subsurface scattering efficiently exploits the fact that it blurs high-frequency details and illumination. This means that simulating subsurface scattering can be approximated as a convolution with a diffusion kernel that mimics the diffuse reflectance profile for a given translucent medium. While the exact reconstruction normally requires an expensive two-dimensional convolution, d'Eon et al. [dLE07] showed that it can be approximated by a sum of radially symmetric Gaussians. Thus, due to the separability of Gaussians, the 2D convolution can be computed using a set of cheaper 1D passes, which allows high-quality skin rendering in real time. This approach was later extended to screen space, modulating the width of the kernel according to per-pixel depth information [JSG09].

However, in order to get adequate results, several Gaussians are needed to model the diffuse reflectance profile. This translates into multiple 1D convolutions per frame, which is still costly. In this paper, we make the key observation that exact simulated diffusion kernels, which are in general mathematically non-separable, can be closely reconstructed

Model	Visual quality	1D convolutions	Separable
1 Gaussian	low	2	✓
2+ Gaussians	high	2 per Gaussian	✗
Kernel pre-integration	high	2	✓
Artist-friendly model	controllable	2	✓

Table 1: Compared to the state of the art [dLE07, JSG09] (above the separation line), our proposed techniques (below) offer solutions for a variety of trade-off choices. Our techniques are able to provide high-quality results by using a separable approximation of only two 1D convolutions.

by a low-rank factorization for a wide range of materials (see Figure 2). Based on this, we present two different separable models that allow simulating subsurface scattering with just two 1D convolutions: the first one, kernel pre-integration, allows reconstructing a high-quality diffusion profile based on the observation that the irradiance is close to be additively separable, while the second is an artist-friendly model that, following the previous observation, provides an attractive trade-off between physical accuracy and ease of use for artistic editing of the scattering profiles (Table 1). Coupled with importance-sampling and our stochastic sampling strategies, our methods only require seven samples per pixel (see Figures 1 and 15).

Our methods can be implemented as simple post-processing steps and do not rely on complex alpha-blending pipelines or Gaussian levels of detail [JG10], and work with dynamic objects without any additional cost. Moreover, all our rank-1[†] approximations typically execute in less than 0.4 ms per frame on modern commodity hardware and exhibit negligible fixed costs, as regions of the scene with no visible scattering can be quickly culled using stencil buffering. Our separable approximation of subsurface scattering fills the gap between physically based subsurface-scattering rendering and highly time-constrained environments such as games, and it is currently being used in game engines and production pipelines.

The main contributions of this work are the following:

- A pre-integrated rank-1 model that can optimally approximate 2D convolution kernels when applied to additively separable signals. These type of signals include shadow boundaries, which are among the most important cues of subsurface scattering.
- An artist-friendly model that allows to intuitively edit subsurface scattering parameters. We show how this model is able to faithfully represent (and edit) Monte Carlo simulations with only two Gaussians, including multi-layer skin, which is among the most difficult materials showing this phenomena.
- A stochastic sampling technique that allows to properly

[†] Rank-1 2D kernels are those that can be analytically separated into two successive 1D kernels.

filter high-frequency non-additively separable signals like pore detail in faces.
- Overall, a practical real-time technique useful in watertight production scenarios.

2. Previous Work

Offline techniques. The simulation of scattering inside translucent materials dates back to the radiative transfer equation [Cha60], which can be solved by traditional path-sampling techniques. The solution of this integral is a very demanding process in terms of computation time, especially if solved for a high number of bounces. Optimization techniques to reduce the computation times include using a dipole model [JMLH01, JB02] and modeling multiple scattering as a diffusion process [Sta95]. Donner and Jensen [DJ05] extended the dipole into a multipole model that allows modeling multi-layered translucent materials, such as skin. The same authors later introduced a photon diffusion technique to combine photon tracing and the diffusion approximation [DJ07]. These works yield impressive results with computation times in the order of seconds per image. A more recent technique improves upon the inherent inaccuracies of the diffusion theory by using a more refined diffusion model to separate single and multiple scattering terms, alongside with the quantization of the Green's function of the diffusion equation to obtain realistic all-frequency results [DI11]. Further improvements revolve around better importance sampling [KF12], while a recent class of techniques rely on solving the searchlight problem by means of Monte Carlo integration imbued by multiple importance sampling [HCJ13].

Real-time techniques. Borshukov and Lewis [BL03] approximate subsurface scattering by blurring a 2D diffuse irradiance texture using a Gaussian filter. While this technique is efficient and maps well to the GPU, it neglects the more subtle details of subsurface scattering. This idea is later extended by d'Eon et al. [dLE07, dL07] to develop a high-quality real-time skin shader. They approximate the multipole model with a sum of Gaussians, and use them to blur the irradiance signal in texture space. Since the Gaussians are separable, this allows transforming the expensive 2D convolutions into a cheaper set of 1D convolutions. This technique enables real-time frame rates, while giving results that are comparable to offline simulations. In follow-up work, additional optimizations are introduced, based on computing a single 2D convolution at 13 jittered sample points, which account for direct reflection and two levels of scattering [HBH09]; unfortunately, 13 samples are not enough for a 2D convolution, which leads to poor results.

Although these techniques provide real-time frame rates, they scale poorly with the number of translucent objects in the scene, since the subsurface-scattering simulation needs to be performed on a per-object basis. To overcome this, Jimenez et al. [JSG09, JG10] propose to translate the simulation from texture to screen space. The diffuse reflection of the translucent object is blurred as a post-processing step employing the sum-of-Gaussians formulation, thereby limiting subsurface scattering computations to the visible parts of the objects. Other techniques operating in screen space include the work of Mertens et al. [MKB*05], using importance sampling of the BSSRDF, Shah et al. [SKP09], who use a splatting process for integration instead of a gathering step, and Mikkelsen [Mik10], who shows how convolution with a Gaussian can be expressed as a cross bilateral filter. Penner and Borshukov [PB11] pre-integrate the illumination effects of subsurface scattering due to curvature and shadowing into textures, assuming that surface normals can be pre-blurred, and that soft shadows are used. Recent real-time techniques use the diffusion approximation to render optically thick materials [WZT*08], including using finite elements and finite differences to support arbitrary, non-locally flat geometry [WWH*10, LSR*13]. In contrast to these works, our technique defers the blurring until the shading has been computed, thus retaining all the geometric detail. Furthermore, it is simpler to implement and yields very high frame rates, suitable for the most challenging real-time scenarios.

3. Separable Subsurface Scattering

The diffuse reflectance of a homogeneous translucent material due to subsurface light scattering is characterized by its 2D diffuse reflectance profile $R_d(x,y)$, which describes the light reflected around a normally incident pencil beam on the origin of a surface of an infinite half-space [JMLH01]. For a homogeneous material, R_d is radially symmetric, and can be characterized by a 1D diffusion profile $R_d(r)$ such that $R_d(x,y) = R_d(\|(x,y)\|)$[‡]. It can be used to calculate the radiant exitance $M_e(x,y)$ at an arbitrary surface point (x,y) according to:

$$M_e(x,y) = \int_{\mathbb{R}^2} E(x',y')\, R_d(x-x', y-y')\, dx'dy', \qquad (1)$$

where $E(x,y)$ is the irradiance at point (x,y), and both $M_e(x,y)$ and $E(x,y)$ are measured in Wm^{-2}. Note that Equation (1) has the form of a 2D convolution with the 2D reflectance profile: $M_e(x,y) = (E * R_d)(x,y)$.

Approximation of 2D diffusion profiles. For real-time applications, carrying out the 2D convolution in Equation (1) is prohibitively expensive. However, if we can express the profile R_d as an approximation A consisting of a sum of *separable* functions, it is possible to approximate this operation

‡ Note that the radial symmetry of the diffusion assumption follows from assuming a normally incident incoming light, which in general does not hold in real applications. For a description of SSS with directional-dependent diffusion we refer the reader to e.g. [DLR*09, HCJ13, FHK15].

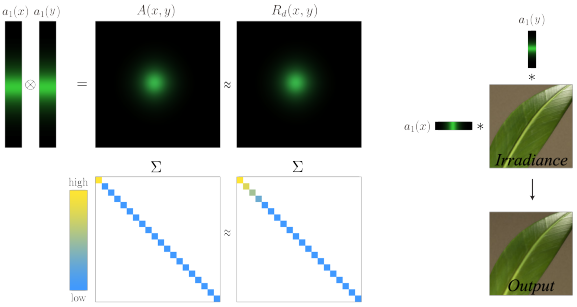


Figure 3: Overview of our approach: based on the low-rank nature of the diffusion kernel $R_d(x,y)$, shown by the Σ matrix below storing the singular values of the kernel (magnitude in grayscale), we approximate $R_d(x,y)$ with $A(x,y) = a_1(x)a_1(y)$. This simplifies the simulation of subsurface scattering (right) to just two 1D convolutions per summand with the irradiance signal.

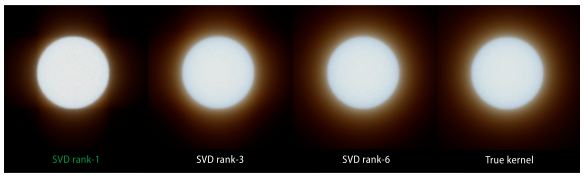


Figure 4: Results of rank-N approximations obtained using the SVD of the discrete diffusion profile for skin, for $N = \{1,3,6\}$. Using the SVD's rank-1 (separable) approximation leads to poor results, since most of the kernel's energy is stored in the center of the kernel. Increasing the rank of the approximation leads to a more faithful approximation of the diffusion kernel (effectively reducing radial asymmetry), but at the cost of introducing several passes, which makes it inefficient for time-constrained applications. We refer to the supplementary material for more results using the SVD low-rank approximation.

by a sequence of $2N$ 1D convolutions, which exhibit significantly smaller computational complexity:

$$(E * R_d)(x,y) \approx (E * A)(x,y) = \sum_{i=1}^{N} ((E * a_i) * a_i)(x,y)$$
$$\text{with } A(x,y) = \sum_{i=1}^{N} a_i(x)a_i(y), \tag{2}$$

where the approximation A is defined by 1D functions a_i. From the radial symmetry of R_d it follows that the same functions a_i can be employed in both coordinate directions. d'Eon et al. [dLE07] observed that zero-mean Gaussians G are suitable functions for approximation:

$$R_d(x,y) \approx A_g(x,y) = \sum_{i=1}^{N} w_i G(x,y;\sigma_i), \tag{3}$$

where σ_i denote the standard deviation of the respective Gaussians. Due to the separability of the Gaussian kernel, the convolution with A_g can be realized as $2N$ 1D convolutions. Unfortunately, for the most demanding real-time scenarios these $2N$ convolutions are still too expensive.

In practice, however, we only work with discretized diffusion kernels, which can be interpreted and analyzed as 2D matrices. This allows us to make the observation that most of the energy of typical diffusion kernels is stored in the first few singular values, and in particular in the first one (see Figure 2). This means that the diffusion profile can be approximated with a low-rank separable approximation, and that a single separable kernel can reproduce most of the kernel perceptual qualities, if chosen appropriately. The application of such a separable kernel for rendering is illustrated in Figure 3.

An obvious choice for a discretized separable kernel would be using just the first component of the *singular value decomposition* (SVD), which – according to the Eckhart-Young theorem [EY36] – gives the best low-rank approximation with respect to the Frobenius norm. Unfortunately, the rank-1 approximation of the kernel using SVD exhibits a rather large energy loss and produces unsatisfactory results even if energy conservation is enforced by normalizing the kernel. This is due to the fact that a pure kernel-space factorization does not take into account that in image-space, some parts of the kernel are more relevant than others (i.e. perceptually, for SSS the kernel fall-off is more important than the center of the kernel, specially at longer viewing distances). Higher-rank SVD-based approximations (i.e., $N \approx 2-6$) converge very rapidly to the original kernel (Figures 4 and 7). However, the increased computation times make it a less attractive option for real-time applications.

In the following we show that, under certain assumptions, even a rank-1 approximation can be used to reconstruct the diffusion kernel with high accuracy (Section 4), and then propose an artist-friendly separable model that allows intuitive editing of the appearance of translucent materials (Section 5).

4. Pre-integrated Separable Kernel

Both the sum-of-Gaussians approximation and our SVD-based method produce unsatisfactory results for a single summand, i.e. $N = 1$. Due to the non-separability of discretized representations of realistic diffusion profiles, it is not possible to fully reconstruct the effect of their convolution with 2D signals by a single separable kernel. Additionally, separable approximation kernels are in general not radially symmetric, as illustrated by an example in Figure 5. It is, however, possible to completely reproduce a profile's behavior on a special class of signals: assuming that the irradiance is additively separable, i.e., $E(x,y) = E_1(x) + E_2(y)$ or, equivalently,

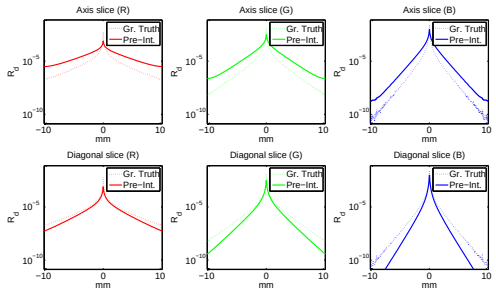


Figure 5: Plot of our pre-integrated kernel compared to the ground truth for human skin, in both the axial and diagonal directions; it can be seen that due to the loss of radial symmetry, our kernel gives different results for the axial and diagonal directions, as opposed to the ground-truth kernel. For this figure, the pre-integrated kernel is using the 2D kernel $A_p(x,y)$, which is obtained from the 1D kernels $a_p(x)$ and $a_p(y)$ (see Equation 5). Note that our method does not try to mimic the kernel to be close to the actual 2D diffusion kernel, but it tries to match the final result of the convolution (Figure 6 and 7). In particular, the pre-integrated kernel method optimizes for additively separable inputs, such as shadow boundaries. Additional comparisons can be found in the supplementary material.

$\frac{\partial E}{\partial x \partial y} = \frac{\partial E}{\partial y \partial x} = 0$, the radiant exitance M_e is given as follows:

$$\begin{aligned}
M_e(x,y) &= \iint E(x',y')\, R_d(x-x',y-y')\, dx'dy' \\
&= \iint \big(E_1(x') + E_2(y')\big)\, R_d(x-x',y-y')\, dx'dy' \\
&= \int E_1(x') \underbrace{\int R_d(x-x',y-y')\, dy'}_{a_p(x-x')}\, dx' \\
&\quad + \int E_2(y') \underbrace{\int R_d(x-x',y-y')\, dx'}_{a_p(y-y')}\, dy' \\
&= \int E_1(x')\, a_P(x-x') \underbrace{\frac{1}{\|a_p\|_1} \int a_p(y-y')\, dy'}_{=1} dx' \\
&\quad + \int E_2(y')\, a_P(y-y') \overbrace{\frac{1}{\|a_p\|_1} \int a_p(x-x')\, dx'}^{=1} dy' \\
&= \iint E(x',y') \frac{1}{\|a_p\|_1} a_p(x-x') a_p(y-y')\, dx'dy'
\end{aligned} \tag{4}$$

where a_p denotes the pre-integrated 1D kernel of R_d along a coordinate axis. Due to the radial symmetry of R_d, we have $a_p(x) = a_p(y)$, where $\|a_p\|_1 = \|R_d\|_1$ by definition. Hence, we define the pre-integrated[§] kernel A_p of the diffusion profile as:

$$A_p(x,y) = \frac{1}{\|R_d\|_1} a_p(x) a_p(y). \tag{5}$$

Note that A_p reproduces the *exact* 2D convolution with R_d in the presence of additively separable irradiance signals, such as straight shadow boundaries of arbitrary orientation or general axis-aligned 1D functions (see Figure 7 for an example of a vertical shadow boundary).Thus, applying the pre-integrated kernel consists on convolving the irradiance vertically then horizontally using the 1D kernel $a_p(x)$ when $A_p(x,y)$ is normalized to one.

Even for general (i.e., non-additively separable) signals, this approximation yields good results for a wide range of materials and scenarios (see Figure 6 and 15), since most real-world signals $E(x,y)$ can be locally approximated with additively separable functions. However, this formulation offers limited control to an artist and needs to be discretized to be used in practical applications (which can be a disadvantage as Section 7 shows). In the following section we describe a separable, *artist-friendly* model that overcomes these two limitations (Section 5), and then describe our solution to improve the quality in presence of non additively separable irradiance signals (Section 6).

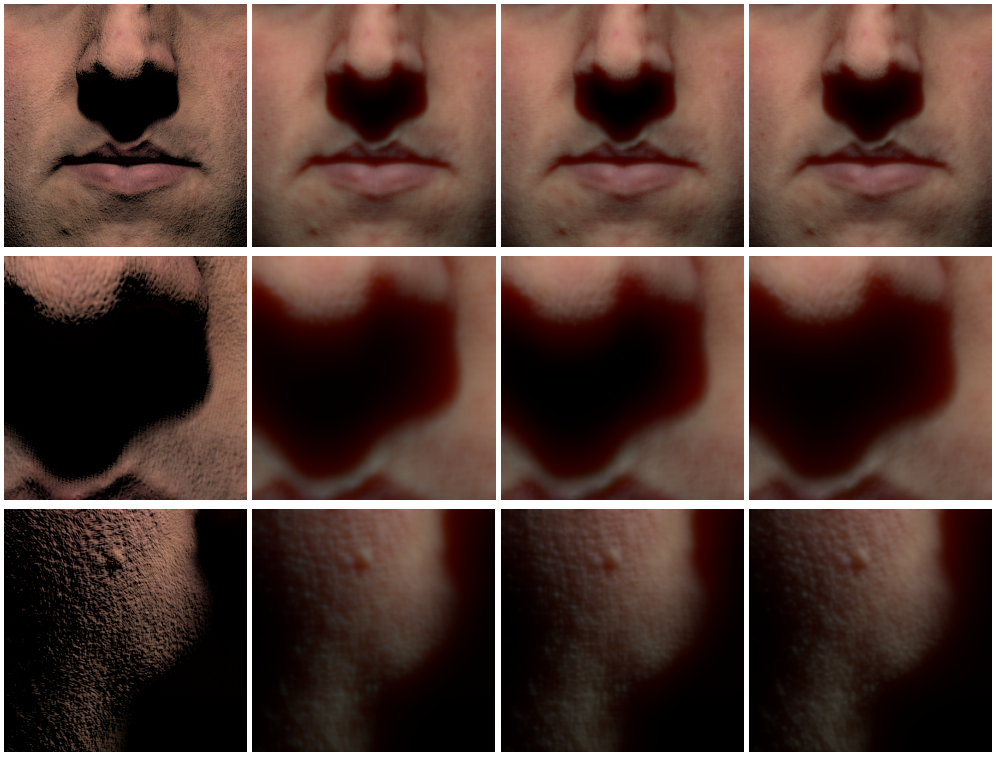
5. An Artist-Friendly Separable Model

Our pre-integrated kernel (Section 4) is able to reconstruct a wide range of materials given its diffusion kernel, which can be obtained from measured data [JMLH01, MES*11] or from simulations. That the energy of typical diffusion kernels is stored in the first few singular values, and that additively separable inputs (which often happen in practice) can be exactly represented by separable kernels evidence that we can find mathematically a separable kernel very close to the ground truth for the general case. However, in a production environment these profiles might not be the optimal ones, since they might not match the assets (e.g., color of the albedo maps) being captured or computed for different types of skins. To solve this issue, we propose an approach suitable for *artistic editing* of subsurface scattering, based on physically meaningful parameters.

A general approximation adding separable kernels (see Equation 2) exhibits a vast amount of degrees of freedom if general functions a_i are used. Even a sum of N Gaussians (see Equation 3) requires the manipulation of $2N$ parameters per wavelength by the artist. Moreover, changing the parameters of one Gaussian in the presence of many others may lead to unexpected results. Jimenez et al. [JJG12] proposed ad-hoc transformations to a base profile in a separable approximation to overcome this limitation. However, this model lacked

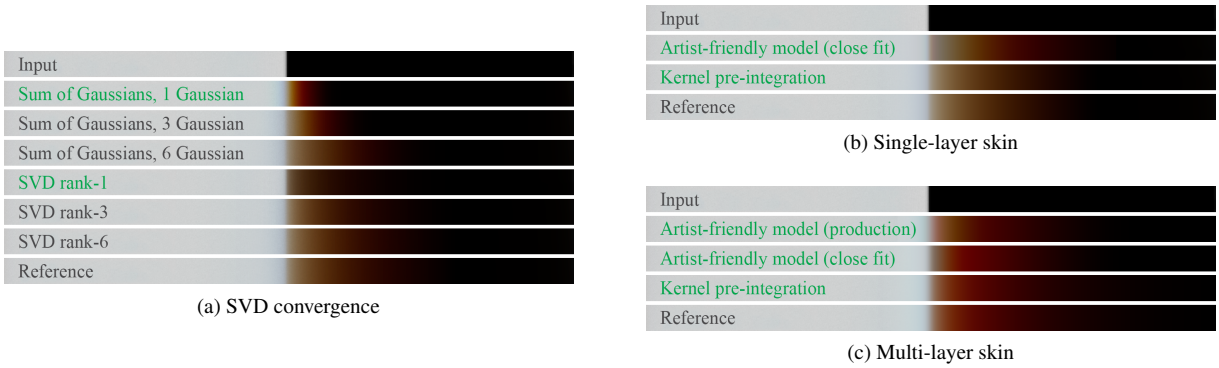
[§] Note that our pre-integrated formulation is fundamentally different from the one proposed by Penner and Borshukov [PB11]: while they pre-integrate the colored gradients found in the shading and shadow boundaries due to subsurface scattering, we pre-integrate the *kernel*, which is later applied to simulate SSS.

(a) Input (b) Kernel pre-integration (c) Artist-friendly model (d) Single-layer reference



(e) Input (f) Kernel pre-integration (g) Artist-friendly model (h) Multi-layer reference

Figure 6: Comparison of the different techniques proposed with single- and multi-layer skin references. The single-layer reference (d) is a Monte Carlo simulation of the *Skin1* parameters in [JMLH01], whereas the multi-layer reference (h) is the six-Gaussian fit for skin [dLE07, JSG09]. Note that our separable approximations (kernel pre-integration and the artist-friendly model) lead to similar quality results with just two 1D convolutions, as opposed to the twelve needed by the sum-of-Gaussians approach (h). Artist-friendly model images are using the close fit parameters (see Table 2). For the case of the single-layer reference, our artist-friendly model yield slightly blurrier results under significant magnification, but is still a plausible approach at regular viewing distances. For illustration purposes, the comparisons are done with pure convolutions over the diffuse lighting (with no depth awareness).



(a) SVD convergence

(b) Single-layer skin

(c) Multi-layer skin

Figure 7: Comparison of the results for our techniques applied to a step-like irradiance. Left: SVD converge very rapidly to the ground truth kernel. In particular, the rank-3 approximation closely resembles the true kernel, favorably comparing with the Sum of Gaussians approximation for the same number of convolutions, at the expense of losing some radial symmetry (as shown in Figure 4). However, the sum-of-Gaussians [dLE07, JSG09] and a low-rank SVD-based decomposition need several convolutions to match the true kernel. Right: our proposed *pre-integrated* kernel, is exact for axis-aligned functions, such as this example, and only requires two 1D convolutions. Our *artist-friendly* model provides a rich design space: this is illustrated by two approximations, where for the first (close fit in (b) and (c)), perceptual similarity to the ground truth was the modeling objective, while the second (production in (c)) was tweaked by an artist for production purposes, giving a pink tint to the near scattering Gaussian for a more fleshy skin appearance (see Figure 10 for actual render comparisons). References in (a) and (b) use a Monte Carlo simulation of the *Skin1* parameters in [JMLH01], while (c) uses the six-Gaussian fit for skin [dLE07, JSG09]. The style of presentation was inspired by previous work [DI11, HCJ13].

an intuitive mapping to the underlying physics of diffusion. Instead, inspired by previous work allowing modeling and editing of subsurface profiles with simple low-dimensional functions [KKCF13], our approach is based on the more intuitive concept of splitting subsurface scattering into near- and far-range scattering, encoded in two Gaussians. These form the basis of our separable kernel a_m as:

$$\begin{aligned} a_m(x) &= w\,G(x,\sigma_n) + (1-w)\,G(x,\sigma_f) \\ A_m(x,y) &= a_m(x)a_m(y) \\ G(x,\sigma) &= \frac{1}{2\pi\sigma}e^{-r^2/2\sigma^2}, \end{aligned} \tag{6}$$

where σ_n and σ_f represent the standard deviation of the near and far scattering Gaussians respectively, and w is the weighting factor. For simplicity of use, we model $\sigma_n = \sigma'_n s$ and $\sigma_f = \sigma'_f s$ respectively, where σ_n and σ_f are in range $[0,1]$, and s is a global scaling parameter of the translucency specified in world space units (e.g. inches or millimeters). Note that a_m represents a mixture of two 1D Gaussians $G(x,\sigma)$, which result in a separable (rank-1) kernel. This is different from the 2D Gaussian mixture approach of d'Eon et al. [dLE07], where two Gaussians would not yield a separable (rank-1) kernel, but a solution of rank two. Note that normalization factor of the Gaussian above is $1/2\pi\sigma_n$ instead of the two-dimensional's $1/2\pi\sigma_n^2$.

Our model takes the benefits of the separable approximation described in Section 3, which is able to match the ground truth under different light configurations (Section 4), while offering a rich design space that allows for intuitive editing of the *appearance* of translucent materials (Figures 8 and 10), including easy integration in production pipelines, where this appearance control is usually required. Our approximation features another important property: since it is based on Gaussians, it is a *continuous* parametric representation of the profile; this representation allows computing the diffusion profile analytically at run-time, as will be shown in Section 7. Moreover, the Appendix introduces an alternative *guided* deviation from the actual diffusion kernel, to generate viable rank-1 approximations that emphasize certain translucency effects.

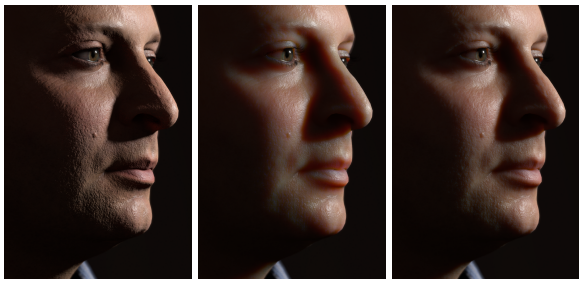


Figure 8: An example of the intuitive editing capabilities of our artist-friendly model. Left: Input irradiance map, without subsurface scattering. Middle: Adjusting the far scattering. Right: Final result after adjusting the near scattering and the balance between the two. Shifting more energy to near scattering allows preserving the bump details.

6. Stochastic Kernel Rotation

While our separable approximations leads to high-quality results under natural lighting, it may lead to some artifacts under high-frequency illumination, when the spatial footprint of the signal becomes smaller than the bandwidth of the kernel (see Figure 11). This is common also in high-resolution, close-up shots (e.g. the pores of the skin), producing an asymmetric *star-like* pattern (see Figure 9a).

To alleviate this problem, we apply a randomized per-pixel rotation of the filtering axes, similar to Huang et al. [HBR*11]. Using a randomized rotation, as opposed to other alternatives such as kernel jittering along its sampling axis, has two key advantages: *(i)* it breaks the visible cross pattern; and *(ii)* due to the radial symmetry of the diffusion kernel, it is not necessary to reintegrate the kernel, since distances are preserved. To avoid GPU cache thrashing and given that we only want to apply the rotation on high frequency signals, we only apply this to samples close to the pixel being evaluated (closer than 10% of the kernel size in our implementation, although it is dependent on the zoom and the used kernel). This solves the artifacts in small-scale features such as skin pores, although for higher-scale features, such as the light dot in Figure 11, there may still be visible artifacts (this is however a pathological case, not common in real-world applications). Figure 9 demonstrates the effects of randomizing the sample positions by rotating the filtering axes. Note that, in addition to masking the artifacts, it also reduces banding problems due to under-sampling (see Figure 13).

7. Rendering

Our approximation of the diffusion profile, represented as just one separable kernel, can be applied both in texture- and in screen space. For efficiency, we use the screen-space approach of Jimenez et al. [JSG09], including translucency [JWSG10], separating into different buffers the albedo, diffuse and specular components, simulating subsurface transport only in the diffuse layer, and compositing for final rendering. In the following, we highlight additional improvements to the rendering pipeline, resulting in an optimized code of just 16 instructions per sample.

Kernel footprint and evaluation. As our technique works in screen space, the size of the kernel is a function of the projected surface area in the pixel, which depends on depth and surface orientation. This area is typically specified in world-space units instead of pixels, making the definition of the kernel size more intuitive for artists.

When using a discretized kernel, the effect of the surface orientation can be taken into account by using ad-hoc correction factors [JSG09]. In contrast, using the simplified two-Gaussian artist-friendly model A_m allows us to work on real-world distances: we transform the depth of the pixel being evaluated and of each sample to world space, and then calculate the distance d between them. This distance is used to apply the profile on the fly, yielding more accurate results than using derivatives or ad-hoc correction factors. However, this approach has two problems: (1) we cannot accurately bake the kernel weights by evaluating the area covered by each sample (so we can only evaluate the kernel in the sample position), and (2) the number of instructions per sample used by the GPU (as generated by DirectX 11 `fxc`) increases from 16 in a simple ad-hoc correction technique to 28, almost halving the performance, making it useless in production scenarios. The latter problem is due to the following: a) converting from depth and pixel position to world-space requires a few additional ALU instructions; b) evaluating a 2-Gaussian RGB profile requires 6 `exp` instructions, which are costly to execute even on modern GPUs.

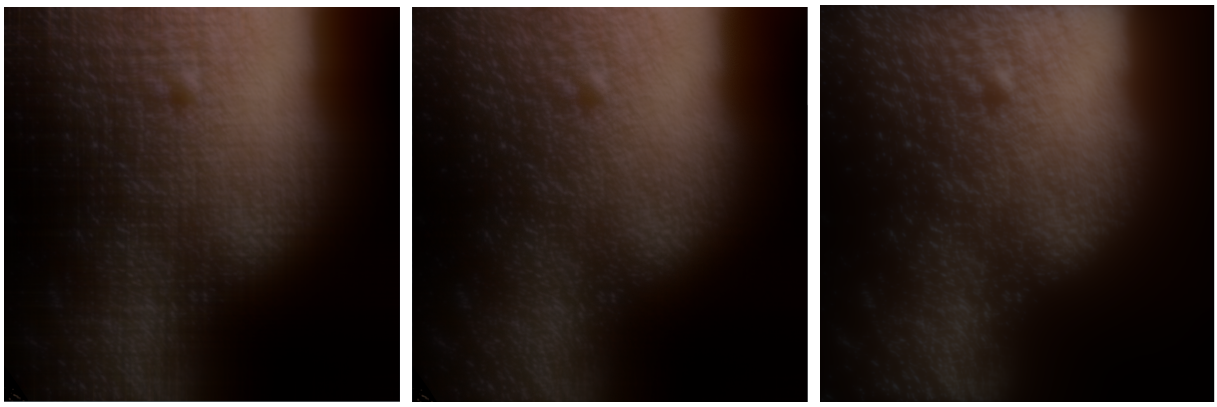
We solve these problems by following an approach similar to the one proposed by Mikkelsen [Mik10], splitting the 1D profile application on 3D distances d to 2D distances d_{xy} with an accurate depth d_z correction factor. This transforms the evaluation of the kernel as:

$$\begin{aligned} a_m(r) &= w\,G(r,\sigma_n) + (1-w)\,G\left(r,\sigma_f\right) \\ &= w\,e^{\frac{-d_z^2}{2\sigma_n}}\,G(d_{xy},\sigma_n) \;+\; (1-w)\,e^{\frac{-d_z^2}{2\sigma_f}}\,G(d_{xy},\sigma_f) \\ &\approx e^{\frac{-d_z^2}{2\sigma_{max}}}\left(w\,G(d_{xy},\sigma_n) \;+\; (1-w)\,G(d_{xy},\sigma_f)\right) \\ &= e^{\frac{-d_z^2}{2\sigma_{max}}}\,a'_m(d_{xy}), \end{aligned} \tag{7}$$

where $\sigma_{max} = max(\sigma_n,\sigma_f) = \sigma_f$ and $r = \sqrt{d_{xy}^2 + d_z^2}$. Note that we are making the approximation of taking the maximum variance of the Gaussians, which simplifies the profile application to an accurate depth correction, versus the ad-hoc corrections used in e.g. [JSG09]. This allows us to precompute accurate weights for $a'_m(d_{xy})$ using area integration, and reduce the number of instructions to 16 by: a) reducing the number of `exp` from 6 to a single one; b) avoiding the conversion to world space by directly working with depths; c) applying typical low-level optimizations [Per14].

Importance sampling. To compute the convolution with the 1D functions of the pre-integrated kernel approximation (a_p in Equation 5) during rendering, they need to be discretized. In general these functions exhibit a very uneven energy distribution, so a uniform discretization will either require a high resolution for acceptable quality, which entails a significant performance impact, or result in aliasing when a lower resolution is used. To solve this issue, we use importance sampling on the 1D function, which allows to allocate a greater amount of sample points near the center, where most of the energy of the signal is found.

For the case of the artist-friendly model and as previously explained, only the *xy* plane is discretized (a'_m in Equation 7), with the z dimension being still analytically resolved. There-



(a) Rand 10% (b) Rand 30% (c) Ground Truth

Figure 9: Although per-pixel randomizing the sampling axis angle for a radius of 10% of the kernel size is enough for usual portrait distances and resolutions of 1080p, extreme close-ups may reveal artifacts due to the non-separability of the kernel in these conditions (a). Performing a randomization on a radius of 30% completely removes these artifacts (b), visually matching the result of the ground truth kernel (c).

fore, for this model, importance sampling is only applied to the baked 2D plane of the kernel (a'_m).

In order to minimize the execution time and the number of memory accesses, we sample all channels at the same positions determined by the dominant channel (which is the red channel for the skin). For the non-dominant channels, we calculate sample weights by integrating the profile over the area covered by the sample. In Figure 12, we demonstrate the difference in image quality with importance sampling compared to the ground truth and uniform sampling. In this particular portrait shot under natural lighting, using 7 samples with importance sampling allows a faithful representation of skin subsurface scattering. This importance sampling can also be used to improve the quality of previous methods, such as the sum-of-Gaussians approach [dLE07, JSG09]. Harsh lighting scenarios require the usage of additional samples, as Figure 13 shows.

8. Results

We validate our proposed techniques with a range of different translucent materials. Figures 1 (left), 6, and 14 depict *human skin*. Figure 7 compares between the different methods also using the skin diffusion profile. Figure 15 features results for *soap* and *marble*, while Figure 1 (right) shows *ketchup*. Additional results are presented in the supplementary material.

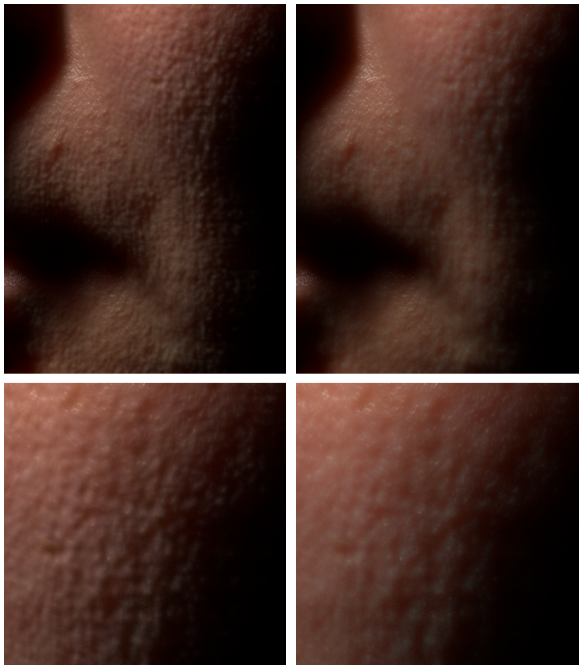
We note that the execution time of every technique described in this manuscript is dominated by the convolution computation, so their timings are roughly proportional to the number of convolutions they require (i.e., six Gaussians is approximately six times slower than our separable approximation). Figure 6 compares between our separable approximation, the current state-of-the-art [dLE07, JSG09], and the ground-truth 2D convolution.

The pre-integrated kernel gives visually convincing results and can be computed analytically from an arbitrary profile. This enables subsurface scattering to be viable for severely time-constrained real-time applications. Note that in the case of a one-dimensional irradiance signal, such as the one in Figure 7, the pre-integration technique yields the *analytic* ground truth.

While the artist-controlled approximation allows designing the diffusion arbitrarily, it can also be used to match a specific profile manually. Although not as exact as the pre-integrated technique for additively separable signals (Figure 7), the artist-friendly model: a) still captures key features in a visually convincing way, such as color bleeding into the shadowed region; b) accurately preserves bump details (Figure 6); and c) it can be seamlessly used in production scenarios. The close-fit kernels for human skin (Figures 6 and 7) shows how the two-Gaussian artistic model can be used to approximate the ground truth. We again emphasize that this two-Gaussian approach is different from the sum of Gaussians proposed by d'Eon et al. To demonstrate the editing capabilities of our artist-friendly model, we show results using different types of skin, including one actually developed by an artist for improving the results in our particular assets (our *production* subsurface scattering parameterization), and others trying to mimic the exact skin diffusion profiles obtained with single- and multi-layer models (Figure 7). Figure 10 shows how our artist-friendly model allows to match the appearance of translucency to the goal asset, by using our *production* parameterization. The parameters for these kernels are included in Table 2.

Kernel	s (mm)	σ'_n R	σ'_n G	σ'_n B	σ'_f R	σ'_f G	σ'_f B	w
Single-layer close fit	10.34	0.077	0.034	0.02	1.0	0.45	0.25	0.5
Multi-layer close fit	9.27	0.055	0.044	0.038	1.0	0.2	0.13	0.5
Production	9.9	0.22	0.07	0.07	1	0.3	0.15	0.65

Table 2: Skin rendering parameters for our artist-friendly separable model with different configurations illustrated in our results. *Single-layer fit* approximates a Monte Carlo simulation using the *Skin1* parameters in [JMLH01]. *Multi-layer fit* approximates the sum of gaussians in [dLE07], fitting a multilayered skin model [DJ05]. Parameters σ'_n, σ'_f and w in range $[0, 1]$.



(a) Multi-layer fit (b) Matched

Figure 10: A kernel designed to match this particular asset using our artist-friendly model shows a more natural and fleshy look than the multi-layer fit reference, due to the better pink-colored bump filling.

Our general convolution scheme can be implemented in both texture- and screen space. In the latter, only the visible pixels are processed and the irradiance of an object is never calculated at higher density than what is required for the final screen pixels [JSG09]. As this leads to a favorable fixed-cost property that is sought-after in real-time applications, we believe these techniques offer a desirable choice for such scenarios. Our method does not require any additional considerations for dynamic objects and moving light sources, and scales well with the area (see Figure 14). The quality of the simulation can be adjusted by setting the number of samples in the kernel, which gives a trade-off between quality and cost.

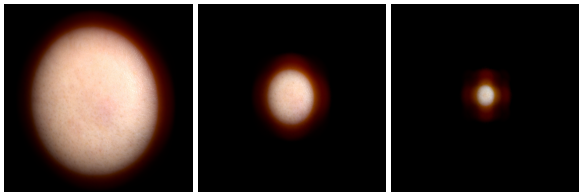
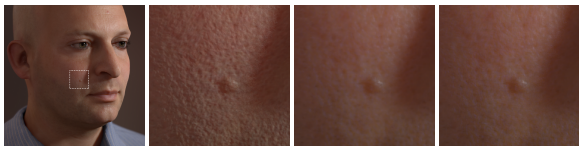


Figure 11: In harsh lighting conditions, progressively smaller features may reveal the fact that the applied separable profiles are not radially symmetric.

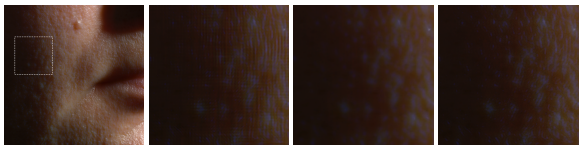


(a) (b) 7 (Uniform) (c) 513 (Uniform) (d) 7 (IS)

Figure 12: The quality of the rendered images, with and without importance sampling of the approximated kernel. (a) Initial image; (b) Uniform sampling with only seven samples leads to obvious aliasing artifacts; (c) Ground-truth created by uniform sampling with 513 samples; (d) Our importance sampling with as low as seven samples.

Limitations and future work. As discussed previously, trading off radial symmetry for separability might create some artifacts revealing the inherent shape of the kernel in extreme close-up views with high-frequency illumination. While the stochastic sampling approach proposed in Section 7 generally solves this issue for very small-scale details (e.g., skin pores), some artifacts may still remain in some situations (Figure 11). However, we have not experienced this phenomenon in any of the practical cases.

Section 3 shows how the error of our reconstruction is determined by the squared sum of the higher-order singular values. Our method relies on taking advantage of the fact that even though the diffusion profiles are of moderately high rank, the information contained therein is highly structured.



(a) (b) 13 (Non-Rand.) (c) 65 (Non-Rand.) (d) 13 (Randomized)

Figure 13: In harsh lighting conditions, extreme close-ups may reveal artifacts even in the presence of importance sampling. (a) Initial image; (b) Importance sampling with 13 samples shows banding artifacts (please zoom in in the digital version for a better view); (c) Up to 65 samples are needed to eliminate visible banding; (d) Our stochastic sampling approach also eliminates banding while keeping the sample count low.

Using our separable model, we have been able to match the target profiles well in terms of RMS error.

Since the pre-integrated approximation yields a faithful reconstruction of the diffusion profile with only one separable convolution, a low-rank approximation can be found by approximating the difference between the pre-integrated kernel and the actual diffusion profile. This could lead to better low-rank approximations. In our use cases, a low-rank SVD-based reconstruction already provided sufficient quality for ranks equal or larger than three (see supplementary material for results). This suggests that an SVD-based reconstruction of diffusion kernels might be very useful for simulating efficient high-quality subsurface scattering in off-line or interactive environments.

9. Conclusions

We have presented two techniques to generate separable approximations of diffuse reflectance profiles to simulate subsurface scattering for a variety of materials using just two 1D convolutions. Our separable models yield state-of-the-art results in less than 0.4 millisecond per frame (in typical shots), which makes high-quality subsurface scattering affordable even in the most challenging real-time contexts such as games, where every desired effect may have a budget of tenths of a millisecond.

Using axis-aligned pre-integration, we have presented a high-quality separable approximation that is probably optimal for additively separable irradiance signals. We also proposed an artist-friendly model that allows intuitive artistic control on the appearance of subsurface scattering based on only three parameters, and that allows seamless integration into our separable framework. We have additionally shown that low-rank approximations based on matrix factorization yield higher performance than the sum of Gaussians, which suggests an interesting avenue of future work for efficient subsurface scattering simulation.

Our algorithm works as a post-processing step, which makes it very efficient and simple to integrate in existing rendering pipelines, reducing the complex subsurface light transport [JMLH01] to its barebones (a seven samples blur filter with 16 assembly instructions per sample). Moreover, we have shown how combining importance sampling and stochastic sampling allows using only seven samples per pixel in many cases of practical interest.

Acknowledgments

The authors want to thank the reviewers for their insightful comments; Infinity Realities, in particular Lee Perry-Smith, for his head model and for the *Lauren* model; the Institute of Creative Technologies at USC, in particular Paul Debevec, for the *Ari* and *Bernardo* models; Bernardo Antoniazzi for letting us use his likeness; and Zbynek Kysela for the modeling and texturing work. Furthermore, we want to thank the Stanford University Computer Graphics Laboratory for the *Dragon* model, and the following contributors from Blend Swap under CC-BY licence: longrender for the *Dish* model, metalix for the *Green apple* model, betomo16 for the *Plant* model, and PickleJones for the *Grapes* model. We also thank Felícia Fehér for editing the figures. This research has been partially funded by the European Commission, 7th Framework Programme, through projects GOLEM and VERVE, the Spanish Ministry of Economy and Competitiveness through project LIGHTSLICE, and project TAMA, and the Austrian Science Fund (FWF) through project no. P23700-N23.

References

[BL03] BORSHUKOV G., LEWIS J. P.: Realistic human face rendering for 'The Matrix Reloaded'. In *ACM SIGGRAPH 2003 Sketches & Applications* (2003).

[Cha60] CHANDRASEKHAR S.: *Radiative Transfer*. Dover Books on Intermediate and Advanced Mathematics. Dover Publications, 1960.

[DI11] D'EON E., IRVING G.: A quantized-diffusion model for rendering translucent materials. *ACM Trans. Graph. 30*, 4 (2011).

[DJ05] DONNER C., JENSEN H. W.: Light diffusion in multi-layered translucent materials. *ACM Transactions on Graphics 24*, 3 (2005), 1032–1039.

[DJ07] DONNER C., JENSEN H. W.: Rendering translucent materials using photon diffusion. In *Proceedings of the Eurographics Symposium on Rendering* (2007), pp. 243–252.

[dL07] D'EON E., LUEBKE D.: Advanced techniques for realistic real-time skin rendering. In *GPU Gems 3*, Nguyen H., (Ed.). Addison Wesley, 2007, ch. 14, pp. 293–347.

[dLE07] D'EON E., LUEBKE D., ENDERTON E.: Efficient rendering of human skin. In *Proceedings of Eurographics Symposium on Rendering* (2007), pp. 147–157.

[DLR*09] DONNER C., LAWRENCE J., RAMAMOORTHI R., HACHISUKA T., JENSEN H. W., NAYAR S.: An empirical bssrdf model. *ACM Trans. Graph. 28*, 3 (2009).

[EY36] ECKART C., YOUNG G.: The approximation of one matrix by another of lower rank. *Psychometrika 1*, 3 (1936), 211–218.

Figure 14: Close-up and regular shots of Bernardo and Ari models, with runtimes of 0.9 and 0.39 ms at 1080p on the PS4. A close-up comparison before (middle right) and after (top right) applying our separable subsurface-scattering model. Note that without taking into account subsurface light transport, the realism of the image is lost.

Figure 15: Real-time results for *soap* and *marble*. The insets show (from top to bottom) input irradiance, the sum-of-Gaussians approach [dLE07] with 1 Gaussian, our analytic kernel pre-integration technique and the ground truth. Both the sum of Gaussians and ours are run with the same number of convolutions, thereby yielding similar execution times. Our method is able to faithfully capture the effect of subsurface scattering in shadowed regions (*soap*), and retains intricate texture features (*dragon*).

[FHK15] FRISVAD J. R., HACHISUKA T., KJELDSEN T. K.: Directional dipole model for subsurface scattering. *ACM Trans. Graph. To Appear* (2015).

[HBH09] HABLE J., BORSHUKOV G., HEJL J.: Fast skin shading. In *Shader X7*, Engel W., (Ed.). Charles River Media, 2009, ch. 2.4, pp. 161–173.

[HBR*11] HUANG J., BOUBEKEUR T., RITSCHEL T., HOLLÄNDER M., EISEMANN E.: Separable approximation of ambient occlusion. In *Eurographics 2011-Short papers* (2011).

[HCJ13] HABEL R., CHRISTENSEN P. H., JAROSZ W.: Photon beam diffusion: A hybrid monte carlo method for subsurface scattering. *Computer Graphics Forum (Proceedings of EGSR 2013) 32*, 4 (2013).

[JB02] JENSEN H. W., BUHLER J.: A rapid hierarchical rendering technique for translucent materials. *ACM Transactions on Graphics 21*, 3 (2002), 576–581.

[JG10] JIMENEZ J., GUTIERREZ D.: *GPU Pro: Advanced Rendering Techniques*. AK Peters Ltd., 2010, ch. Screen-Space Subsurface Scattering, pp. 335–351.

[JJG12] JIMENEZ J., JARABO A., GUTIERREZ D.: *Separable Subsurface Scattering*. Tech. rep., Universidad de Zaragoza, 2012.

[JMLH01] JENSEN H., MARSCHNER S., LEVOY M., HANRAHAN P.: A practical model for subsurface light transport. In *Proceedings of ACM SIGGRAPH 2001* (2001), pp. 511–518.

[JSG09] JIMENEZ J., SUNDSTEDT V., GUTIERREZ D.: Screen-space perceptual rendering of human skin. *ACM Transactions on Applied Perception 6*, 4 (2009), 1–15.

[JWSG10] JIMENEZ J., WHELAN D., SUNDSTEDT V., GUTIERREZ D.: Real-time realistic skin translucency. *IEEE Computer Graphics and Applications 30*, 4 (2010), 32–41.

[JZJ*15] JIMENEZ J., ZSOLNAI K., JARABO A., FREUDE C., AUZINGER T., WU X.-C., VON DER PAHLEN J., WIMMER M., GUTIERREZ D.: Separable subsurface scattering. *Computer Graphics Forum 34*, 6 (2015).

[KF12] KULLA C., FAJARDO M.: Importance sampling techniques for path tracing in participating media. *Comp. Graph. Forum 31*, 4 (June 2012), 1519–1528.

[KKCF13] KING A., KULLA C., CONTY A., FAJARDO M.: Bssrdf importance sampling. In *ACM SIGGRAPH 2013 Talks* (2013).

[LSR*13] LI D., SUN X., REN Z., LIN S., TONG Y., GUO B., ZHOU K.: Transcut: Interactive rendering of translucent cutouts.

Visualization and Computer Graphics, IEEE Transactions on 19, 3 (2013), 484–494.

[MES*11] MUNOZ A., ECHEVARRIA J. I., SERON F., LOPEZ-MORENO J., GLENCROSS M., GUTIERREZ D.: BSSRDF estimation from single images. *Computer Graphics Forum 30* (2011), 455–464.

[Mik10] MIKKELSEN M.: *Skin Rendering by Pseudo-Separable Cross Bilateral Filtering*. Tech. rep., Naughty Dog Inc., August 2010.

[MKB*05] MERTENS T., KAUTZ J., BEKAERT P., REETH F. V., SEIDEL H. P.: Efficient rendering of local subsurface scattering. *Computer Graphics Forum 24*, 1 (2005), 41–50.

[PB11] PENNER E., BORSHUKOV G.: *GPU Pro 2*. AK Peters Ltd., 2011, ch. Pre-Integrated Skin Shading, pp. 41–55.

[Per14] PERSSON E.: Low-level shader optimization for next-gen and DX11. In *Game Developers Conference* (2014).

[SKP09] SHAH M. A., KONTTINEN J., PATTANAIK S.: Image-space subsurface scattering for interactive rendering of deformable translucent objects. *IEEE Computer Graphics and Applications 29* (2009), 66–78.

[Sta95] STAM J.: Multiple scattering as a diffusion process. In *Eurographics Rendering Workshop* (1995), pp. 41–50.

[WWH*10] WANG Y., WANG J., HOLZSCHUCH N., SUBR K., YONG J.-H., GUO B.: Real-time rendering of heterogeneous translucent objects with arbitrary shapes. In *Computer Graphics Forum* (2010), vol. 29, Wiley Online Library, pp. 497–506.

[WZT*08] WANG J., ZHAO S., TONG X., LIN S., LIN Z., DONG Y., GUO B., SHUM H.-Y.: Modeling and rendering of heterogeneous translucent materials using the diffusion equation. *ACM Trans. Graph. 27*, 1 (Mar. 2008), 9:1–9:18.

Appendix A: Guided optimization

For a guided deviation from the actual diffusion kernel, we additionally present a practical optimization framework to generate viable rank-1 approximations that emphasize the translucency effects of certain features. We aim to find a separable approximation $A_s(x,y) = a_s(x)a_s(y)$ to the diffusion profile $R_d(x,y)$, defined by the solution to the minimization problem:

$$a_s = \underset{a}{\operatorname{argmin}} \int_{\mathbb{R}^2} \Gamma(x,y)\big(R_d(x,y) - a(x)a(y)\big)^2\,dx\,dy \quad \text{subject to } \|R_d\|_1 = \|a\|_1^2. \tag{8}$$

We optimize for minimal $\mathcal{L}_2$ distance while still retaining energy conservation via the 1-norm constraint (see Section 3). The *guide function* $\Gamma(x,y)$ provides the means to select the appropriate length scale of intended artistic effect, and has the form:

$$\Gamma(x,y;k) = \left(x^2+y^2\right)^{k/2}(1-e^{-bx^2})(1-e^{-by^2}), \tag{9}$$

where k denotes the *guide parameter* and b gives the *suppression term* of the center cross region (we use $b = 50$). Consult the supplementary material for a motivation of this function. Varying k between 0 and 4 provides control over the perceived sharpness of the approximation A_s. $k \approx 2$ yields the approximation of the actual diffusion kernel, while $k = 0$ provides a visually sharper variant and $k = 4$ a smoother approximation that models the far-range scattering of the diffusion profile more faithfully. This optimization-based approach approximates diffusion kernels while still allowing to emphasize either near or far scattering with the help of a single parameter. This way, the user can adapt the kernel to a given scene and its specific properties, which is not possible by using the single, fixed solution of the pre-integration scheme.