CHAPTER 1. INTRODUCTION

"We may say most aptly that the Analytical Engine weaves algebraic patterns just as the Jacquard loom weaves flowers and leaves.

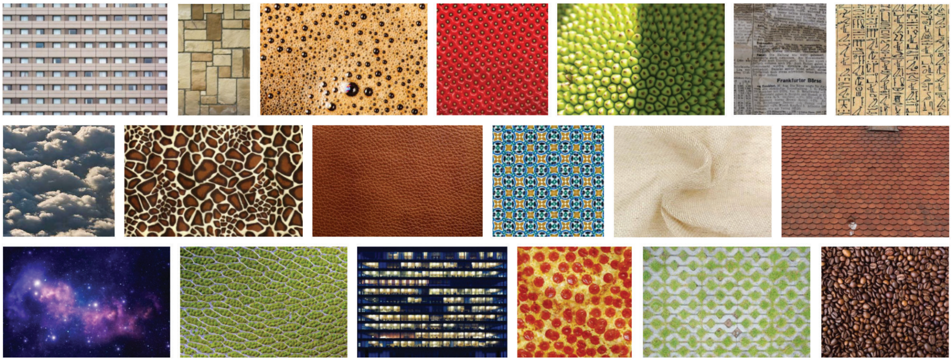
— ***Ada Lovelace***

Figure 1.1: Renders of virtual scenes in different use cases. (a) A frame of Pixar's *Soul* (2020), directed by Peter Docter. (b) A frame of *The Last of Us Part II* (2020), Naughty Dog. (c) Garments designed with *SEDDI Author*. (d) A render of *Ceramic House*, courtesy of *Studio RAP*.

Generating lifelike virtual scenes has been a long-standing goal in computer graphics. This objective is now closer than ever before, thanks to the development of efficient methods for creating such scenes. Realistic virtual scenes have a range of applications, including media and entertainment, such as art, video games, augmented and virtual reality, and films. Furthermore, they are also important in industrial and commercial settings, such as computer-aided design, architecture, and fashion. We illustrate some use cases in Figure 1.1.

Achieving photo-realism in virtual scenes requires accurate and efficient representations of the optical behavior of materials, object geometry, and illumination. Although many methods have been developed for generating realistic imagery using these representations, producing photo-realistic images often requires a significant amount of manual labor from skilled artists, as well as substantial computational resources. One of the key bottlenecks in this process is the generation of digital representations of materials. Materials, depicted in Figure 1.2, are ubiquitous in the real world and exhibit an astonishing variety of structures, regularities, and colors. Realistically rendering them also requires modeling how they interact with light, taking into account their levels of glossiness, transparency, roughness, and microgeometric structures, all of which impact their appearance. This wide range of material properties poses challenges to the development of efficient representations that can capture all the variations required to render the real world in a visually compelling way.

Figure 1.2: Examples of real-world materials and textures.



Furthermore, several design and manufacturing processes involve the manual design of materials, the outcome of which remains uncertain until different fabrication steps have been completed. This work methodology is ubiquitous in industries such as garment design or fashion, and it is slow and wastes a significant amount of valuable resources. The fashion industry alone is projected to generate more than 3 million metric tons of CO_2 emissions (10% of the annual worldwide emissions) and waste 170 billion cubic meters of water in 2025 [RSS16], significantly exacerbating the climate emergency. Advanced techniques in engineering and computing have opened up the possibility of digitally creating and editing simple materials, such as plastics or metals. However, complex deformable materials, notably fabrics, have not been extensively studied due to their intricate optical and mechanical properties. This is problematic, as billions of garments are manufactured every year (a staggering 80 billion in 2015 alone [Mor15], a number that has only increased since then). Besides, establishing a relationship between physical parameters that are relevant to the manufacturing process and perceptual parameters that closely align with how designers perceive materials is a challenging task. Achieving this mapping could be instrumental in enabling designers to modify materials according to their creative vision and effectively communicate their intentions to manufacturers.

Although plenty of solutions for material digitization exist, they are often limited in terms of accuracy, scalability, or cost. Accurately generating digital materials can be expensive and challenging, as it requires the use of inaccessible capture devices such as *gonioreflectometers*, or significant manual effort from skilled artists. These resource requirements not only lead to inefficiencies but also restrict the number of individuals who can benefit from material digitization. While cost and time constraints are major obstacles to material digitization, data-driven algorithms and low-cost devices like smartphones offer promising solutions.

Offering digital solutions to the challenges of material digitization has the potential to bring significant benefits to various industries and society as a whole. These benefits include more efficient manufacturing processes, reduced waste, a more dynamic economy, and more immersive virtual experiences for art, media, and entertainment. In this thesis, we present innovative methods that address these challenges by leveraging ideas from computer vision, physically-based rendering, and recent advances in machine learning.

In the following sections, we provide a theoretical background of physically-based rendering (Section 1.1), the open problems in the literature (Section 1.2), our goals (Section 1.3) contributions to tackle these problems (Section 1.3.1) and a list of measurable outcomes of the contributions of this thesis (Section 1.4).

1.1 Theoretical Background

In this section, we introduce important theoretical background, which is relevant to the work executed in this thesis. *This background is derived from a survey on deep intrinsic image decomposition* [Gar+22]*, which was written during the development of this thesis.*

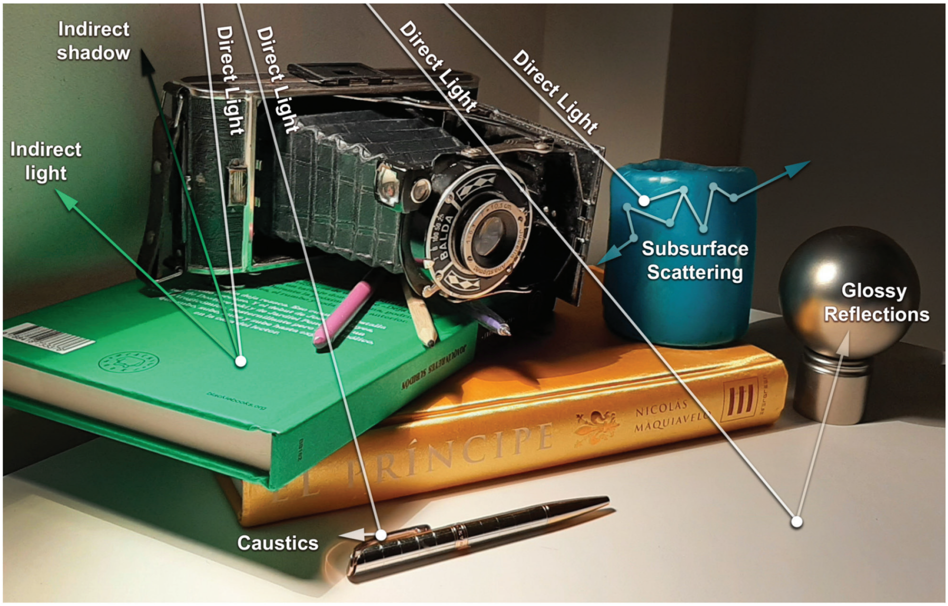
1.1.1 Physically-Based Rendering

If we look at any simple scene surrounding us, such as the photography in Figure 1.3, we can find a plethora of optical interactions: indirect lighting (color bleeding), internal scattering in translucent objects, caustics, anisotropic and glossy reflections, etc. Far from the traditional assumptions in intrinsic imaging of diffuse (lambertian) shading, direct lighting, and diffuse albedo materials.

In the following, we provide an overview of the theoretical background behind the image formation model, its derivation for non-diffuse materials, and the link with physically-based and inverse rendering. For a deeper dive into any of the concepts quite briefly described below, we recommend reading the book on physically based rendering by Pharr et al. [PJH20] and the survey by Guarnera et al. [Gua+19].

The color and luminosity at any point of an image, the incoming *irradiance*, is proportional to the sum of the outgoing *radiance* from all the visible points of the scene towards the camera sensor at the corresponding pixel, I, resulting from multiple interactions between light and matter in the scene. Naturally, this is a simplification: even if we consider the camera lenses and color filters as part of the scene, the interaction of irradiance and the sensor point affects the result, and both electronic and film cameras have specific additional image formation steps which can be simulated. In physically-based rendering, the general approach to compute this value is to use Monte Carlo estimators of the pixel and shading integral [Kaj86], which has the general form of:

Figure 1.3: Example of light transport and material interactions. Secondary bounces of light produce reflection caustics (chrome pen) and color bleeding from the green book. The wax candle exhibits multiple internal (subsurface) scattering of photons. The yellow silk fabric of the book cover shows specular anisotropic reflections due to yarn orientation. Figure from [Gar+22].



$$I = \int_{\chi} f\left(x, \Theta\right) dx \tag{1.1}$$

where f is a function that defines the radiance towards pixel I and is defined on a domain χ, generally a unit sphere, or the set of all the surfaces (A) in the scene, and depends on the scene parameters Θ, which might include the definition of the geometry (normals, z-depth, vertices), material (albedo, BRDF), or illumination sources ś far-field environment lighting, or 3D light emitters (point, area, objects).

The value of I is defined for a given λ, which is the spectral band of the camera sensor. We could define it for as many bands as desired (including non-visible ones) but in practice, the majority of camera sensors mimic the human visual system and are commonly three-band: $\lambda \in \{R,G,B\}$. In subsequent rendering equations, we will simplify the notation by assuming a single-band, and omitting the term λ.

Equation 1.1 is an integral of integrals (see Figure 1.4): to account for all the light arriving at a surface point p_1 to a sensor pixel at p_0, we have to estimate all the contributions of light from all the surfaces of the scene, recursively tracing paths bouncing in surfaces ($p_2, p_3 \dots p_n$) until we reach the light emitted by a source $L_e(p_n \to p_{n-1})$. This is referred as the Light Transport Equation (LTE) in rendering and it is another way of seeing the equation 1.1. To compute the radiance reaching the pixel I, that is from p_1 to p_0 in Figure 1.4, we would need to solve:

$$L\left(p_1 \to p_0\right) = \sum_{k=1}^{\infty} P\left(\overline{p}_k\right) \tag{1.2}$$

with $P\left(\overline{p}_k\right)$ being the radiance scattered over a path $\overline{p}_k$ with $n+1$ vertices ($p_0, p_1, p_2, p_3 \dots p_n$) and computed as:

$$P\left(\overline{p}_k\right) = \underbrace{\int_A \int_A \cdots \int_A}_{n-1} L_e\left(p_n \to p_{n-1}\right) T\left(\overline{p}_k\right) dA\left(p_2\right) \dots dA\left(p_n\right) \tag{1.3}$$

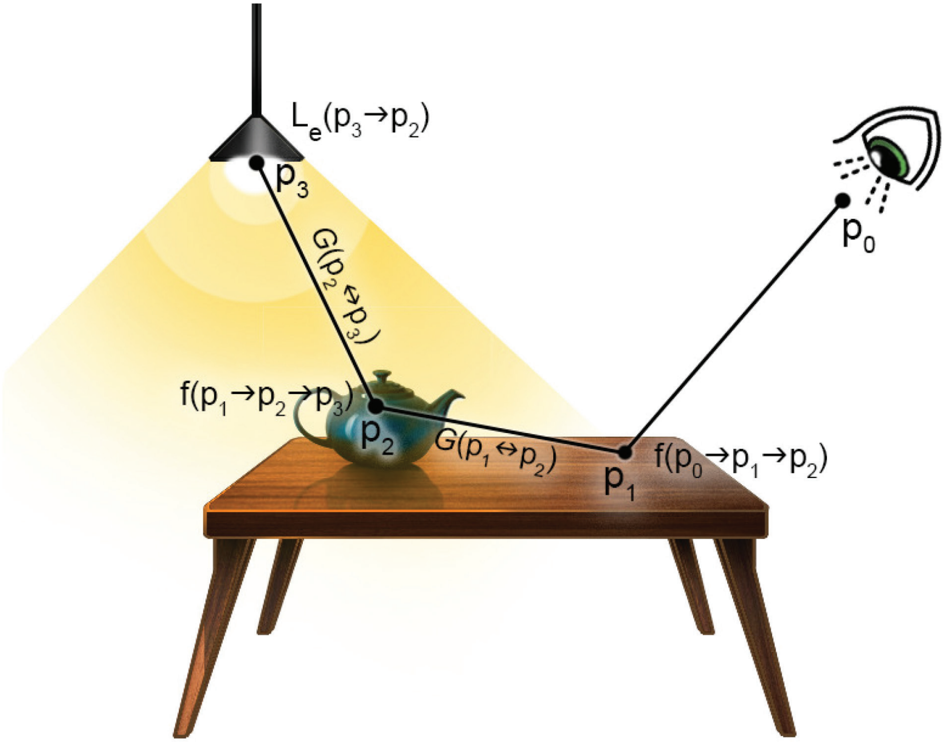
We can integrate over solid angles in the unit sphere, or surfaces (A) of the scene, being $dA(p_i)$ the differential area at point p_i. Note the term $T\left(\overline{p}_k\right)$, named *throughput* of the path: the fraction of radiance from the light source that arrives at the camera after all of the scattering at vertices between them. The total transmitted energy will be reduced at each interaction event, as some wavelengths (colors) are absorbed or scattered away from the observer. This is a common trade-off decision in many Monte Carlo rendering engines; longer paths per sample are costly to compute, while contributing less and less energy with each additional vertex, but in some scenes, they might be very relevant to reduce variance (noise) and converge with fewer samples to an accurate image.

If there are interactions with non-opaque materials (human skin, cloth, marble) or participat-ing media, such as liquids or smoke, we have to integrate volumetric scattering interactions of photons along the path between the light sources and the camera sensor pixel, requiring a more complex mathematical model such as the Radiative Transport Equation (RTE).

1.1.2 Geometry

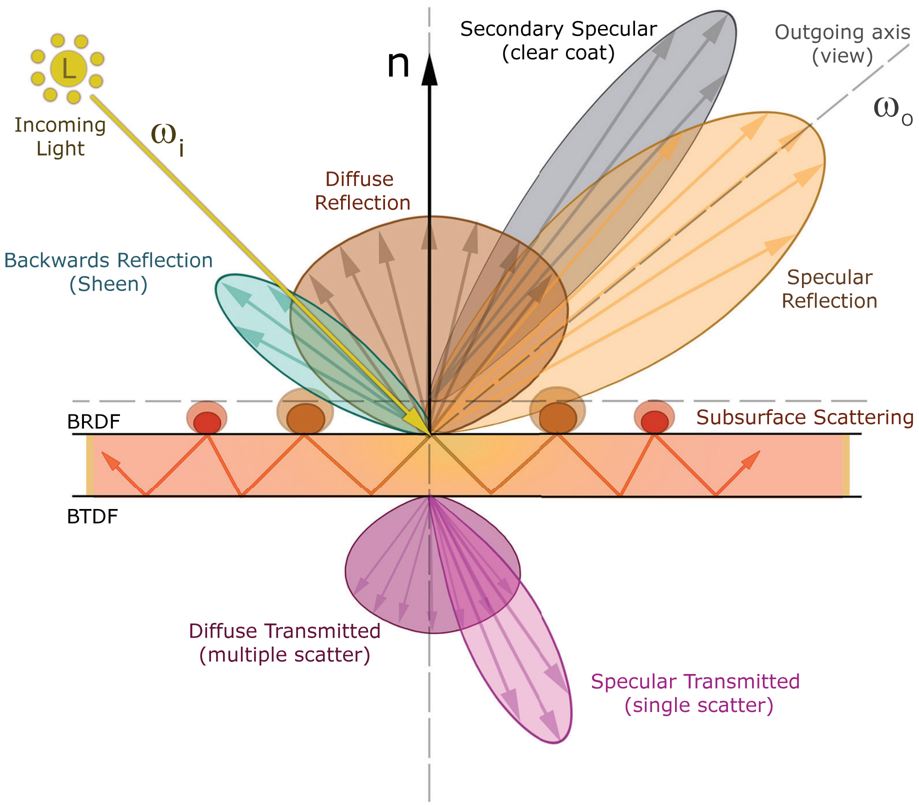
If we take a look at Figure 1.4, we can observe that materials are distributed on discrete 3D objects: the table and the cup, and even the light source if we consider it as an emissive material. We can assume that one of the most important Θ parameters is geometry, usually in the form of 3D vertices and edges, normals (N), or depth maps (D). Please note that, in contrast to a a full 3D mesh, a single camera-view depth image (*a.k.a.* Z-buffer) is an incomplete definition because the non-visible surfaces are undefined, and the reflected light paths cannot be traced behind the visible objects. For example, Nimier et al.[Nim+19] require multiple views of a smoke volume in order to reconstruct its 3D density distribution by inverse rendering.

Figure 1.4: Example of light transport, connecting a light source p_3 to a pixel I at p_0. Multiple paths like this one will need to be explored to provide a good statistical estimate of the radiance from p_1 to p_0. Figure from [Gar+22].



There are additional ways of defining this geometry, such as implicit surfaces, but the material distribution is more complex when there are no clear surface boundaries. That is the case of heterogeneous participating materials: human skin, airlight, mixed liquids, smoke, etc., where light transport between two points has to be in turn evaluated along the path to account for all the possible scattering effects. For instance, imagine a small cloud of vapor between points p_1 and p_2 in Figure 1.4, at each infinitesimal step along the path between those points, there will be a possibility of absorption (collision with a water particle) that will reduce the energy, but there is also a possibility of receiving incoming energy (not emitted by p_2), as the cloud itself is receiving direct illumination from the light source and multiple scattering events distribute the light across its volume.

Figure 1.5: 2D depiction of a physically-principled BSDF theorical model. In most standard representations, the continuous reflectance 4D function is discretized into a combination of analytic lobes (Cosine, GGX) which can be easily computed sampled. Note that subsurface scattering (photons traveling through the medium) is not composed by multiple lobes. They are depicted for descriptive purposes, but it is rather approximated by a constant value or a diffusion profile, if not explicitly computed by simulating multiple scattering events with path tracing or photon mapping. Figure from [Gar+22].



In order to compute the rendering equation in volumetric media, a distribution of parameters is required at any point of the space, not only on the 3D surfaces. The usual representations include solid 3D implicit functions (for instance 3D Perlin noise, used in cloud procedural generation), meshes and distance fields, or discrete volumetric grids which store the scattering probability function at any point of the space (also known as *phase function*) by means of voxels.

1.1.3 Materials

Each time the light interacts with a material, there is a loss of energy and a transformation of the original wavelength reflected towards the observed direction. In rendering, the result depends on the intrinsic material response for those two angles: incident light and viewing direction (*e.g.*: the camera, or another element of the scene). For *surfaces*, this response is modeled with a Bidirectional Reflectance Distribution Function (BRDF) $f_r(x,\omega_i,\omega_o)$ which yields at each particular 3D surface point x, and for each incident direction ω_i, the fraction of reflected radiance observed from a direction ω_o.

The total reflected radiance L at any point x can be obtained by integrating with Equation 1.4 over the positive hemisphere Ω^+, to sample the whole incident light attenuated by the cosine term (dot product between the incident light direction L_i and the normal of the surface N) [Kaj86]:

$$L(x, \omega_o) = \int_{\Omega^+} f_r(x, \omega_i, \omega_o) L_i(x, \omega_i)(\omega_i \cdot N)\, d\omega_i \tag{1.4}$$

Note that by integrating the computed radiance $L(x,\omega_o)$ of the points sampled from p_0 at the camera sensor ($x = p_1, \omega_o = p_1 - p_0$, see Figure 1.4), we are obtaining the irradiance at the sensor and the corresponding image pixel values (I in equation 1.1). Naturally, even the pixels themselves can be sampled several times and integrated over the camera sensor with another Monte Carlo estimator to minimize aliasing effects.

The BRDF can be extended with a Bidirectional Transmittance Distribution Function (BTDF) to conform a full Bidirectional Scattering Function (BSDF), defined in the full sphere Ω. The model can be further extended to account for Surface Scattering phenomena (BSSDF).

These functions have a minimum of four dimensions (input-output pair directions in polar coordinates) and usually three RGB values as output. It is thus technically possible to choose a discrete set of (ω_i,ω_o) orientations and create a lookup table to interpolate the response of the material, which is captured with multiple light and camera positions (*e.g.*with a gonioreflectometer). Storage becomes a major drawback for tabulated data, which can only be reduced through a significant reduction of quality. Moreover, we are considering only homogeneous surface materials, which is not often the case in actual scenes (*e.g.*, a printed paper). Those spatially-varying values

(SVBRDF) can be stored in a stack of textures; or a Bidirectional Distribution Texture Function (BTF), increasing the dimensions and size of the table.

Beyond direct compression techniques, the most successful approach in graphics has been the use of analytic N-dimensional functions to approximate the reflectance and scattering distributions. The simplest of them, Lambertian diffuse, and Phong specular shading are also well known in the computer vision community. These functions leverage symmetry (isotropy) to model the material with a few parameters. For instance, a Lambertian material only requires knowing the intrinsic albedo, a sort of base color, while the Phong model requires three additional parameters for the specular component (*e.g.*, shininess).

In the following subsections, we review the most common and simple material assumptions used in recent papers, finalizing with the most sophisticated inverse material models which are starting to be studied in computer graphics.

Lambertian Assumption

The Lambertian assumption is the most common material reflectance simplification used to tackle the problem of intrinsic image decomposition. It consists of assuming that the BRDF of a surface is constant in all directions (diffuse) and, consequently, the observed light radiance does not depend on the viewpoint. Therefore, we can omit ω_o in the surface reflectance model fr used in Equation 1.4.

If the surface is diffuse, then $f_r(\omega_i) = \frac{\rho d}{2\pi}$, with ρd denoting the diffuse albedo: the constant ratio of incident light which is reflected in any direction, independently of the viewpoint ω_o. The image pixel value I is then given by:

$$I = \underbrace{\frac{\rho d}{\pi}}_{A} \underbrace{\int_{\Omega^+} L_i(\omega_i)(\omega_i \cdot N)\, \mathrm{d}\omega_i}_{S} \tag{1.5}$$

The intrinsic model then can be defined as,

$$I = A \cdot S \tag{1.6}$$

where S contains all the shading variations due to the geometry of the local surface w.r.t. light direction. In some cases, the shading image should contain contributions of all the lights in the scene ($S_1 + S_2 + \ldots + S_k$), which for discrete directional lights, can be deterministically estimated with a linear summation:

$$S = \sum_{1}^{K} L_i(\omega_i)(\omega_i \cdot N) \tag{1.7}$$

In the case of a more realistic illumination representation, such as environment lighting, or indirect light, the shading computation requires sampling the whole hemisphere Ω^+, often recursively sampling other surfaces to approximate the integral of the incoming light. The shading component S within Equation 1.5 in the integral form, is difficult to compute and not so easily invertible and differentiable, so until recently, most intrinsic decomposition methods assumed the simpler formula described in Equations 1.6 and 1.7. Note that the illumination visibility is not considered in most cases (E.g.: cast shadows).

Non-Lambertian Assumption

There are two possible sources producing a Lambertian shading, either a surface which has an extremely rough micro-geometry, and thus reflects light equally in multiple random directions at any differential patch of the surface, or a very diffuse light source, coming from any direction with equal intensity (*e.g.*, a foggy day). Both scenarios can be combined (Equation 1.4): the shiniest object on a foggy day will look quite diffuse, while even the most diffuse materials tend to project specular reflections under focused lighting from certain view angles. However, the majority of materials in the world are not Lambertian: even the most diffuse surface will exhibit Fresnel reflections when observed at grazing angles. Therefore, most surfaces will show the view-dependent effects that are classified as specular reflections. This separation between specular and Lambertian is rather pragmatic, but arbitrary, as even a simple microfacet model (shown in Figure 1.5) requires multiple analytic 3D lobes to approximate the 3D reflectance response for an infinitesimal incoming light ray (ω_i)). The term *specular* is usually applied to narrow lobes with high probability of scattering radiance, producing high luminance values at pixels (highlights). This family of materials are of the general form:

$$L(\omega_o) = \int_{\Omega^+} \underbrace{f_r(\omega_i, \omega_0)(\omega_i \cdot N)}_{fNL} L_i(\omega_i)\,\mathrm{d}\omega_i \tag{1.8}$$

$$f_{NL}(\omega_i, \omega_o, N) = f_d(\omega_i, N) + f_s(\omega_i, \omega_o, N) \tag{1.9}$$

where f_{NL} is a non-lambertian BRDF composed by two components: f_d, a diffuse isotropic lobe, and f_s, a specular lobe which depends on the camera viewpoint (ω_o).

Dichromatic Reflection Model This particular Non-Lambertian model [MFS08; Tom94] separates the object in two reflection components (S_d, S_s), but considers that the specular component S_s might have a color α_s which could differ from the color of the reflected light:

$$L(\omega_o) = \alpha_d \underbrace{\int_{\Omega^+} f_d(\omega_i, N) L_i(\omega_i)\,\mathrm{d}\omega_i}_{S_d} \tag{1.10}$$

$$+\alpha_s \underbrace{\int_{\Omega^+} f_s\left(\omega_i, \omega_o, N\right) L_i\left(\omega_i\right) \mathrm{d}\omega_i}_{S_s} \tag{1.11}$$

$$I = \alpha_d S_d + \alpha_s S_s \tag{1.12}$$

This is the case of metallic materials which, unlike dielectric ones, will show specular reflections with a change in wavelength. Some additional effects such as colored interreflections might be also captured in all layers.

Phong and Blinn-Phong. The dichromatic model can be extended with one of the most adopted analytic approximations, either Phong $\left(f_{NL}^{P}\right)$ or Blinn-Phong $\left(f_{NL}^{P}\right)$, which could be estimated with Monte Carlo integration and arbitrary lighting, or analytically computed with directional light sources:

$$f_{NL}^{P}\left(\omega_i, \omega_o, N\right) = \alpha_d\left(\omega_i \cdot N\right) + \alpha_s\left(h \cdot N\right)^k \tag{1.13}$$

$$f_{NL}^{BP}\left(\omega_i, \omega_o, N\right) = \alpha_s\left(\omega_i \cdot N\right) + \alpha_s\left(\mathbf{r} \cdot \mathbf{v}\right)^k \tag{1.14}$$

where α_d and α_s are the colors of the diffuse and the specular reflections, the halfway vector $h = \frac{\omega_i + \omega_0}{\left\|\omega_i + \omega_0\right\|}$ depends on the light direction w_i and the view direction ω_o. The size of the specular lobe is determined by the scalar term k E R.

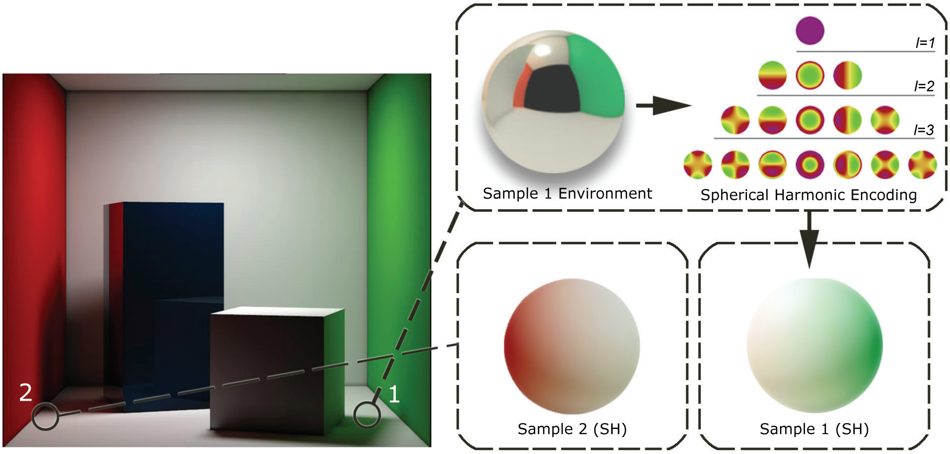
Beyond Dichromatic Models: Physically-based Materials

Naturally, the breadth of materials that can be synthesized with the previous models is very limited and not quite realistic in most cases. The advent of physically-based materials has introduced many variations [Hil+15] of the original microfacets models [TS67], which assume that a surface is composed of many very tiny facets that reflect light perfectly. By controlling the statistical distribution of their orientations, the roughness of the surface varies from mirror-like to almost diffuse. Additional optical properties are introduced in these models: Fresnel view-dependent reflectivity, multiple specular lobes, metalness (conductive materials such as gold, change the color of the highlights), multiple reflection and refraction lobes, etc.

The separation of the lobes described in the multi-lobed physically-based model shown in Figure 1.5 is not arbitrary, grouping reflections and refractions which share the same orientation and energy level. For instance, the main diffuse lobe is grouping multiple different orientations which are not view-dependent and share the same intensity and color. If it covers the full hemisphere with a cosine-like ratio, it is often referred as *Lambertian*. Likewise, the specular transmitted lobe is grouping a view-dependent peak that the observer would only see when the translucent surface is between the light source and the camera.

Even if it is often called a *single scatter* lobe, likely multiple internal scattering bounces of light are also included in this group.

Figure 1.6: Example of spatially varying illumination encoding with Spherical Harmonics (SH). The incoming lighting can be computed globally for the whole scene (far field), or locally, at multiple points (near field) as we show for the two samples near each colored wall. If we project the irradiance (top row) into an SH basis we obtain a diffuse low-frequency representation (examples in bottom row). Figure from [Gar+22].



1.1.4 Illumination

The illumination is a significant contributor to the shading term (S) in most decompositions. From a rendering perspective, as shown in Equation 1.15, the computation of the pixel radiance, L, requires considering both the emitters L_e and the irradiance: the integral of all the incoming lighting at the observed point. This incoming illumination is often neglected, considering only point light or directional analytic emitters, which simplify the shading computation by removing L_i from the integral. If f_r is also assumed to be Lambertian, only the form factor given by the cosine of the surface normal, and the light direction remains.

$$
\begin{aligned}
L(\mathrm{x}, \omega_0) = {} & L_e(\mathrm{x}, \omega_0) \\
& + \int_{\Omega^+} f_r(\mathrm{x}, \omega_i, \omega_0) L_i(\mathrm{x}, \omega_i) \max(\omega_i \cdot N, 0) \, \mathrm{d}\omega_i
\end{aligned}
\tag{1.15}
$$

In actual scenes, the incoming lighting is a combination of emitted or reflected illumination from distant objects (far field) and local surfaces close to the observed area (near field). The former is usually approximated in computer graphics with environment lighting, often based in High-Dynamic-Range (HDR) images mapped into an infinite sphere or cube surrounding the scene, while the latter can be derived from the far field illumination, by simulating the local secondary light bounces. If the geometry does not change, an environment map can be stored at multiple scene locations and distances, to include near field effects more accurately (Spatially Varying Environment Maps), although at a great memory cost, and only for static scenes.

To reduce the sampling and size of environment maps and simplify the computation, Ra-mamoorthi and Hanrahan [RH01a] proposed their compression with Spherical Harmonics (SH), a set of orthonormal basis functions defined on the spherical domain (elevation θ and azimuth ϕ angles). Thus the equation 1.16 describes the irradiance E as the sum of bases weighted by the cosine decay term A_i and the illumination coefficient $L_{l,m}$. By changing the representation of the bases $\hat{Y}$ to polynomial coordinates of a unit normal $n=(x,y,z)^T$, this becomes an efficient vector dot product operation (Equation 1.17). With the required modifications, this strategy is feasible with other orthogonal basis functions on the sphere.

$$E(\theta, \phi) = \sum_{l,m} \hat{A}_l L_{l,m} Y_{l,m}(\theta, \phi) \tag{1.16}$$

$$E = \hat{Y}^T L \tag{1.17}$$

$$E = T^T L \tag{1.18}$$

If we want to account for near field occlusion and interreflection, it is possible to precompute those local interactions, because they depend on the object geometry and materials, and not on far field illumination. This family of techniques is known as *precomputed radiance transfer* (PRT) [SKS02]: they precompute multiple events of light transport (see Figure 1.4) into the T term in Equation 1.18 with Monte Carlo pathtracing. In this fashion, each pixel will have secondary light bounces stored in a light transport map. If only the visibility term $V(\omega_i)$ is considered for, the method will be storing the ambient occlusion shadows, but not colored interreflections.

In Figure 1.6, we can see a pyramid of spherical harmonics bases $Y_{l,m,}$ with different coefficients. It is important to know that, although usually nine bases are considered enough to account for 99% of the far-field irradiance at diffuse surfaces, this percentage is significantly smaller in glossy surfaces (requiring many more coefficients). Moreover, a small number of coefficients will never account for high frequency effects, such as cast shadows from high-frequency light sources (E.g: a point light representing the sun), even producing ringing artifacts if we try to increase the accuracy by adding more bases. There are other popular basis in rendering such as Haar Wavelets or Spherical Gaussians (SG) [Wan+09], which also have very interesting properties.

1.2 Open Problems

In this section, we aim to address the challenges inherent in creating precise and efficient virtual representations of materials and scene radiance. We begin by discussing the current difficulties in acquiring and editing high-quality digital materials that are suitable for photorealistic physically based rendering. Specifically, we will explore the challenges associated with tileable texture synthesis and image-based material propagation.

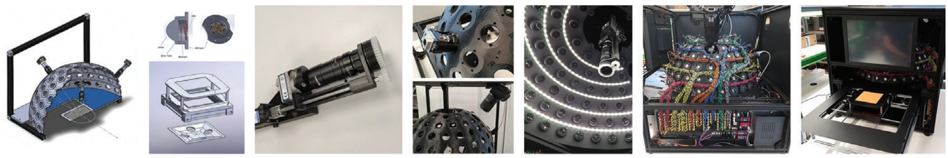
We then turn our attention to the major roadblocks that hinder low-cost digitization of the mechanical and optical parameters of real-world materials. We will examine the various obstacles to overcome when attempting to accurately capture and reproduce the physical characteristics of materials in a virtual environment, using scalable and affordable pipelines.

Lastly, we will review the limitations of current research on radiance encoding using neural networks, with a particular focus on representing material reflectance and scene illumination. By identifying these challenges, we hope to provide insight into the current state-of-the-art in material and radiance modeling for physically based rendering.

Digital Material Generation

Accurate and high resolution material reflectance digitization is a crucial problem for generating compelling and realistic virtual environments. As we have mentioned in Section 1.1.3, this digitization process is typically done using a *gonioreflectometer*, a highly complex machine designed to measure the photometric response of a material using light sources and cameras placed at different positions around the hemisphere. Existing devices for measuring material appearance in spatially-varying samples are limited to a single scale, either micro or mesoscopic. This is a practical limitation when the material has a complex multi-scale structure, like is common on many real-world materials. In many materials, accurately measuring their spatially-varying reflectance requires high resolution microscopic images. For example, the optical behavior of textiles is very dependent on small fibers or their yarn twist [CLA19]. Moreover, many materials also show large-scale *mesoscopic* variations, like prints, plaid, or tartan patterns, which cannot be captured using microscopic photographs. Further, many of such devices cannot capture important optical properties, like the material transmittance, which is key for the realistic rendering of objects like foliage or fabrics. There are no available capture devices or algorithms for accurate and high resolution SVBSDF digitization of heterogeneous materials. In order to introduce such materials, which are very common in many industries like fashion, textile, or leather manufacturing, into rendering pipelines, these machines need to be created and accurate algorithms need to be developed to handle the data they capture. Finally, such a device may serve as a ground truth data generation source of training models that operate on more limited data scenarios, like in single image digitization

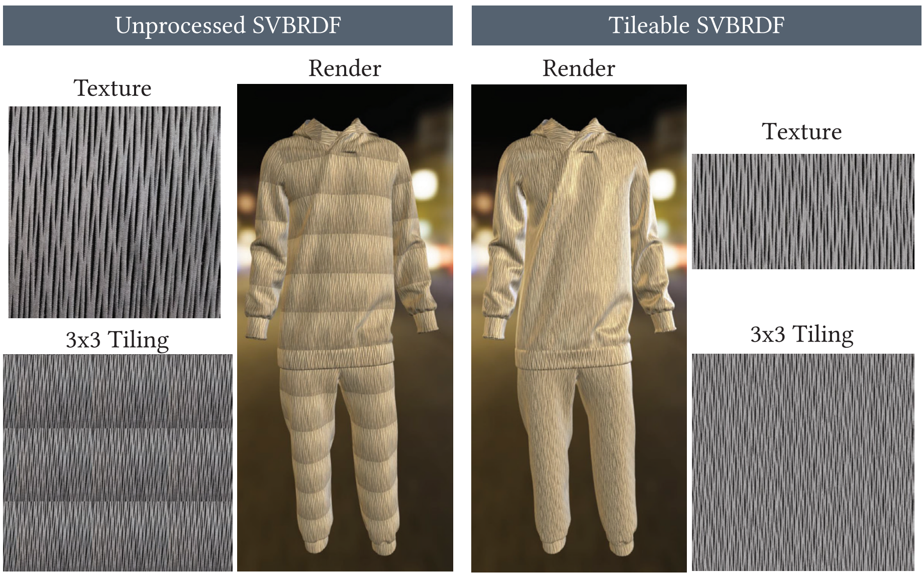
Figure 1.7: Schemes and photos of SEDDI's optical device, which can capture highly accurate and detailed SVBSDFs. (a) Schema of a cross-section of the hemisphere including micro camera and one polar camera. (b) (Top) Schema of the collimated LED design to account for the polarizer and collimating lens; (Bottom) Schema of the main holder and backlight support. (c) Microscopic optical setup. (d) (Top) A Polar Camera; (Bottom) Mid-distance Camera. (e) Interior of the dome with the diffuse LED strip activated. (f) Exterior of the dome and wiring. (g) Holder and exterior cover. Figure from [Gar+23].



Tileable SVBRDFs Even if reflectance is accurately measured for a large portion of a material, in rendering settings, it is also important that the digital material is *tileable*. A tileable texture, also known as a seamless texture, is a texture image that can be repeated infinitely in all directions without any visible seams or discontinuities. In other words, a tileable texture is designed in a way that it seamlessly connects with identical copies of itself, which can be arranged side-by-side or stacked on top of one another to create a larger, continuous pattern without any visible repetition. This is often used in computer graphics, especially for creating textures on 3D models, or even for tiling backgrounds on websites or applications. We illustrate the importance of *tileable SVBRDFs* in Figure 1.8. Recent progress has been made in generating tileable textures from a single input example [Mor+17; Rod+19; Mor+20], however, these works present several shortcomings for tileable SVBRDF generation. First, they either assume a particular level of regularity, or the generated textures lose a significant amount of visual fidelity with respect to the input exemplars. Further, these methods are limited to generating a single texture map (i.e. the albedo of the material). This is problematic for tileable SVBRDF synthesis, which requires transforming every texture map in the SVBRDF into a tileable map, all while preserving pixel-wise coherence between maps. All in all, generating high-quality tileable SVBRDFs from a single input example is not currently possible with the methods proposed in previous work.

Material Attribute Transfer Effective material representations require understanding the properties that uniquely define them, which we refer to as *visual* *material attributes*. These attributes are spatially-varying parameters that maintain spatial coherency with respect to the material structure, while remaining invariant to changes in the scene illumination or the geometry of the underlying object. For example, they may represent optical properties of a SVBRDF, artistic stylizations or higher-level properties, as in semantic segmentation masks. Obtaining these attributes for large samples of the material is very challenging. One way to address this problem is to obtain them at a small exemplar of the material, and *transferring* them into larger portions of it, using a large input image which serves as *guidance*. Different visual attribute transfer methods have been proposed in the past, however, they are limited in their efficiency, robustness, generalization or overall performance. Ideally, such a transfer method should be robust to new material illumination conditions, camera degradations or input distortions; as well as being efficient (providing interactive transfer times even in low computational budgets), controllable and predictable, and capable of generating high resolution outputs, which are required for realistic SVBRDFs. Developing a method with these characteristics may prove useful for many downstream tasks, like efficient data generation or material digitization pipelines.

Figure 1.8: On the left, an input, unprocessed SVBRDF, which is not tileable. On the right, a tileable version of the same material. Renders generated using SEDDI's Textura.ai.



Scalable Digitization

Material Reflectance Estimation As mentioned, accurate material digitization typically requires expensive devices which take significant amounts of time and manual input to operate. One such case is the capture machine illustrated in Figure 1.7, which is presented in [Gar+23]. These provide highly realistic digital representations of material reflectance at the cost of scalability, hindering their applicability in many real world settings. For instance, many manufacturing processes, like those in the textile industry, generate a massive variety of different materials, which cannot be easily nor cheaply digitized with the required cadence that characterizes these industries. Further, the limited availability of these devices introduces additional inefficiencies and economic and ecological costs, as physical samples of the materials must be shipped from the client to the institution holding the capture device.

Data-driven material reflectance estimation from low-cost devices is a long-standing problem in the literature [AWL13; AWL15; Des+18; Hen+21; Guo+20b; VPS21]. These methods promise to provide scalable, efficient, inexpensive on-site material digitization. They typically rely on one or more flash-lit images of a material taken with a smartphone camera, and a deep learning model trained on synthetic materials, which outputs plausible estimations. However, these approaches present several drawbacks that make them unsuitable for practical digitization workflows: First, many of such methods rely on generative models which tend to produce artifacts, or use large neural networks which puts limits on the output resolution, hindering the quality of the estimated materials. Besides, smartphone flash-lit images introduce significant calibration challenges; while most existing datasets used to train these models are purely synthetic, further limiting the quality of these estimations. Moreover, many of these methods use perceptual losses for training or in test-time optimization, which, while they help digitize stochastic materials, present additional challenges for guaranteeing the repeatability and consistency required for building a digital inventory. On top of these limitations, few-shot material reflectance estimation is still an ill-posed problem, as such, many different outputs may be plausible given the same input. So far, there is no way to quantify uncertainty for this problem, which hinders the applicability of these systems in real-world scenarios, and limits their trustworthiness, efficiency, and reliability.

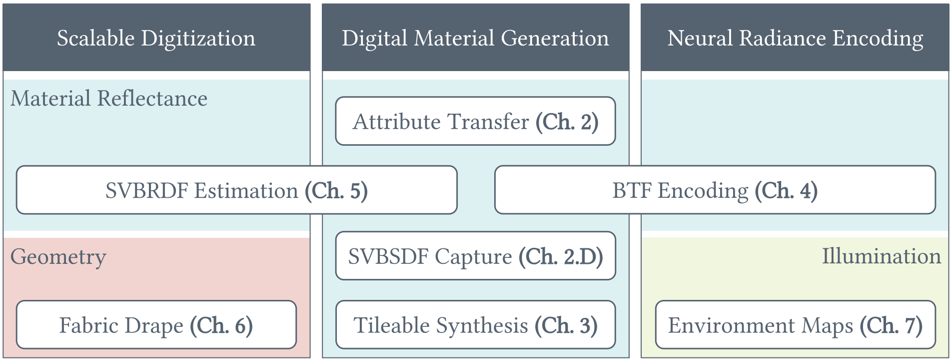
Fabric Mechanical Behaviour Besides material reflectance, another important component of scenes is object geometry. While many real-world objects have static shapes, which can be scanned with relatively accessible devices like the LiDAR sensor in *iPhones*, there are materials that change shape when external forces are applied to them, like gravity. In this sense, *deformable materials* require particular computational models and capture pipelines. One of the most common of such materials is fabrics, which are ubiquitous in the real world, and used in our clothing, furniture, or vehicles. For them, accurately capturing their static shape is not enough: We need to be able to simulate their behavior on new scenarios, like new garments or dynamic forces, like wind [Ber+23] or human movement [SOC22]. Furthermore, textiles are incredibly varied in their fabrication patterns (eg different weave patterns or knitted structures), compositions, and finishing, which not only determine their optical appearance [CLA19] but also their mechanical behavior, as shown in Figure 1.9. The problem of capturing fabric mechanical properties so they can be re-simulated in virtual settings has been amply studied in the literature, however, current solutions require tedious human intervention or expensive capture devices, hindering their scalability. Further, there is a lack of understanding of the perceptual similarity of fabric mechanical behavior, which creates additional challenges for designing accurate fabric mechanical parameter estimation pipelines.

Figure 1.9: Two real garments and their *digital twins*, for two fabrics with different mechanical behavior, as seen on their final drapes. Figure from [Rod+23b].

Neural Radiance Encoding

Reflectance In Section 1.1.3, we have mentioned several limitations of current representations of material reflectance. In particular, Bidirectional Texture Functions (BTFs) provide dense and accurate material reflectance measurements, however, they are typically pro-hibitive in terms of computational and memory cost. Recent work [Rai+19; Rai+20; Kuz+21] introduce different neural compression algorithms for BTFs, achieving remarkable realism and efficiency. Neural reflectance encoding methods trained on either synthetic or measured BTFs have thus shown promising results for increased realism of rendered materials. However, existing neural material encodings are immutable, meaning that their output for a certain query of UVs, camera, and light vector is fixed once they are trained. This can become very limiting when the fragment of the material used for training is too small or not tileable, which frequently happens when the material has been scanned with a capture device. Therefore, current neural reflectance encodings, while accurate and efficient, have severe limitations in terms of editing capabilities, which hinder their applicability and usefulness in real world rendering settings.

Figure 1.10: Overview of the structure of this thesis. We propose different algorithms to solve problems (each column) in different components (each color) of virtual scenes.



Illumination Finally, as we mention in Section 1.1.4, there are many available approximations for representing real-world illumination in virtual scenes, including Spherical Harmonics, Spherical Gaussians, or Haar Wavelets. These, however, struggle to efficiently and accurately represent non-diffuse illumination conditions, like high-frequency light sources, as the sun or light bulbs. Recent work on neural representations of natural illumination [GES22] provides better approximations for such cases, but struggle with indoor lighting or night scenes with multiple light sources. Furthermore, these neural approximations do not provide sampling and PDF evaluation capabilities, which are essential for Monte Carlo rendering in path tracing. To the best of our knowledge, there are no lighting approximation methods that provide these sampling capabilities, while working accurately on any type of input environment map. Recent work on invertible generative models and implicit neural representations may provide a promising pathway to achieve these objectives.

1.3 Goals & Contributions

The primary objective of this thesis is to create innovative algorithms for computational representation and capture of materials and scene radiance. These are important for several reasons. These digital representations may be used in a wide variety of applications and problems, including computer graphics, video games, film production, virtual and augmented reality, or computer-aided design. They have the potential to significantly increase the visual realism and quality of the virtual environments created in these settings. Furthermore, the accuracy and efficiency of these methods directly impact the quality and computational resources required for these tasks. As such, obtaining accurate and efficient representations and digitization algorithms for materials and other scene components can help reduce the computational cost of rendering, design, capture, and modeling, which can lead to faster, more efficient, and scalable pipelines. Moreover, they may enable efficient procedural dataset creation for training models to solve downstream tasks, like dense segmentation, image or video generation, or embodied vision. Finally, besides efficiency and accuracy, additional features like differentiability or uncertainty quantification capabilities create opportunities for inverse rendering applications or active learning.

To this end, we introduce new learning-based methods to address the open problems described before. Our proposed solutions strive to achieve maximum computational and data efficiency, as well as accuracy, control and reliability. Furthermore, our algorithms are designed to be beneficial to end-users by incorporating human requirements such as perceptual components, edition, predictability, robustness, and uncertainty quantification. All of our proposed methods leverage neural networks, incorporating recent advances in implicit representations, architecture and loss function design, and training procedures. They are all fully differentiable, and may be incorporated into optimization pipelines to solve inverse rendering problems. Our solutions for radiance encoding may be seamlessly incorporated into path-tracer engines for increased render efficiency, or into neural scene representations for increased realism. Further, our high quality casual digitization methods allow for scalable and low cost material capture, which can help end users create their own realistic virtual materials, or create large datasets at a lower cost. We provide an overview of this thesis in Figure 1.10.

Our goals in this thesis tackle the aforementioned challenges in digital material generation, scalable digitization, and radiance encoding, and can be summarized as follows:

- Design new learning-based methods for predictable, controllable and high quality digital material propagation, generation, edition, and synthesis.
- Introduce novel efficient and differentiable representations for scene radiance.
- Democratize high-quality material capture by presenting new methods which are low cost, scalable, accurate, reliable, and perceptually-validated.

1.3.1 Contributions

The work developed in this thesis has led to the following contributions to achieve the goals presented above:

- A deep-learning based method for propagating spatially-varying attributes of a material to larger samples of the same or similar materials. At the core of this method is a lightweight fully-convolutional neural network trained using a *photometric dataset* and an extensive data augmentation policy. These contributions allow our models to generalize to new illumination, color, and geometric conditions. We show the effectiveness of our system on transferring attributes of different types, including surface normals, semantic segmentation, and artistic editions (Chapter 2).

- An extension to the previous method, for enabling a dual-scale capture system capable of digitizing a single material at high-resolution and accuracy levels. To do so, we extend the method presented in Chapter 2 for allowing to transfer multiple property maps at the same time, and with a higher degree of accuracy (Chapter 2.C).

- A generative model capable of creating *tileable textures* using a single image as input. By leveraging state-of-the-art generative adversarial networks (GANs) for texture synthesis, a novel latent space manipulation algorithm, and using the discriminator as a quality metric, our proposed model can generate tileable textures of higher perceptual quality, and at a lower cost than previous work. Further, we propose an extension for generating tileable SVBRDFs from a single input, and show the effectiveness of our method across textures of different levels of regularity. (Chapter 3).

- A neural field representation capable of *BTF encoding and transfer*. Building upon previous work on neural material representations, we propose a lightweight latent representation that can be decoded by an implicit neural network into per-texel reflectance values. This latent representation is estimated by an autoencoder-like network, which can be used to propagate BTF values to novel structures, allowing for neural material edition and transfer, and providing new capabilities for these types of representations. (Chapter 4).

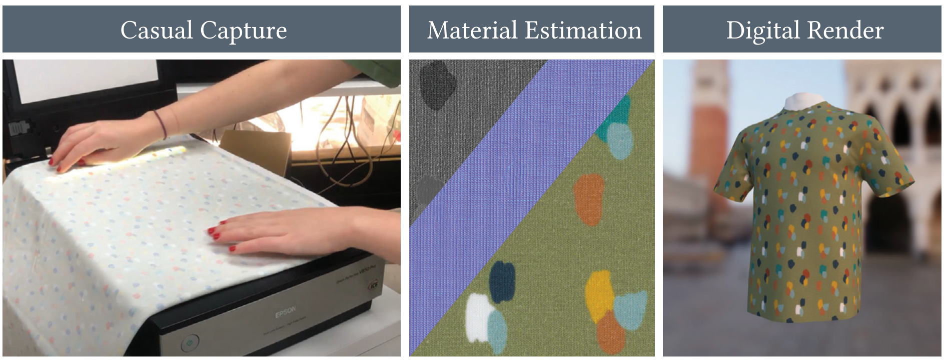
- A novel single-image capture system capable of generating high-resolution SVBRDFs of materials captured from commodity scanners. Leveraging a custom-built dataset and a novel attention-enhanced image-to-image translation generative model trained with a variety of loss functions, our model provides artifact-free, highly accurate digitizations, using solely microgeometry patterns as cues. We further propose the first uncertainty quantification algorithm for SVBRDF estimation methods, building upon Bayesian deep learning approximations using Monte Carlo dropout during test time, and perceptual BRDF metrics. We show the effectiveness of our uncertainty metric for predicting digitization error at test time and for an active learning experiment which shows that uncertainty sampling helps build more data-efficient capture systems. (Chapter 5). We illustrate the capabilities introduced by this model in Figure 1.11.

- A learning-based system for digitizing mechanical properties of fabrics using a casual capture setup and a commodity depth camera. Training solely on synthetic data and leveraging an attention-enhanced multi-image neural network, transfer learning, and an extensive data augmentation policy, our proposed method can accurately estimate mechanical properties of fabric samples, without the need for expensive capture equipment. Further, we propose a novel image-based perceptual metric to measure mechanical similarity between fabrics, which we validate with a user study. (Chapter 6).

- A novel neural method for environment maps, which provides an efficient global illumination representation. At the core of our method lie two neural networks: a *normalizing flow*, capable of learning the *pdf* of an input environment map, and to efficiently sample directions from it; and a *implicit neural network* with sinusoidal activations capable of mapping from directions to linear RGB radiance values. We show the effectiveness of our method on *Multiple Importance Sampling* applications in rendering, obtaining accurate illumination at a significantly lower cost than traditional methods; as well as better quality environment maps than previous work on illumination representations. (Chapter 7).

- A comprehensive study of the capabilities and limitations of deep neural networks for *intrinsic image decomposition*. With a careful review of previous work, approaches, assumptions, datasets, losses, and neural network architectures, we provide a new categorization of these works. We further propose future research directions, building upon recent work on differentiable and inverse rendering and advances In machine learning systems. This contribution has been published in [Gar+22], a part of it is used in Section 1.1.

Figure 1.11: An illustration of the material digitization process enabled by some methods presented in this thesis. We show a picture of a user casually scanning a piece of fabric (left), the estimated SVBRDF for that material (middle), and a digital render of this SVBRDF (right). Assets generated using SEDDI Textura.ai.



Each of these contributions has been developed in collaboration with other researchers and engineers, many of whom are listed as authors of the publications listed below. However, most of this thesis would not have been possible without the outstanding work of the whole team at SEDDI, which has played roles in important tasks, including administrative work, dataset creation, financial support, help with software and infrastructure, as well as valuable scientific discussions. My individual contributions to each project can be inferred from my relative position in the author list.

1.4 Measurable Outcomes

1.4.1 Publications & Patents

Peer Reviewed Publications
The contributions in this thesis have led to the following publications, listed in chronological order:

- *"Neural Photometry-guided Visual Attribute Transfer"*
  Carlos Rodriguez-Pardo, Elena Garces; *IEEE Transactions on Visualization and Computer Graphics (TVCG), 2021* [RG21].
  This journal has an impact factor of 5.226, and its position in the JCR index is 13 out of 110 (Q1) in the category Computer Science, SOFTWARE ENGINEERING (data from 2021).

- *"A Survey on Intrinsic Images: Delving Deep into Lambert and Beyond"*
  Elena Garces, Carlos Rodriguez-Pardo, Dan Casas, Jorge Lopez-Moreno; *International Journal in Computer Vision (IJCV), 2022* [Gar+22].
  This journal has an impact factor of 13.369, and its position in the JCR index is 10 out of 145 (Q1) in the category COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE (data from 2021).

- *"SeamlessGAN: Self-Supervised Synthesis of Tileable Texture Maps"*
  Carlos Rodriguez-Pardo, Elena Garces; *IEEE Transactions on Visualization and Computer Graphics (TVCG), 2022* [RG22].
  This journal has an impact factor of 5.226, and its position in the JCR index is 13 out of 110 (Q1) in the category COMPUTER SCIENCE, SOFTWARE ENGINEERING (data from 2021)

- *"How Will It Drape Like? Capturing Fabric Mechanics From Depth Images"*
  Carlos Rodriguez-Pardo, Melania Prieto-Martin, Dan Casas, Elena Garces; *Computer Graphics Forum (Proceedings of Eurographics 2023)* [Rod+23b].
  This journal has an impact factor of 2.363, and its position in the JCR index is 53 out of 110 (Q2) in the category COMPUTER SCIENCE, SOFTWARE ENGINEERING (data from 2021).

- *"UMat: Uncertainty-Aware Single Image High Resolution Material Capture"* Carlos Rodriguez-Pardo, Henar Dominguez, David Pascual, Elena Garces; *Proceed-ings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023* [Rod+23a].
  This conference is the highest impact venue in computer science according to *Google Scholar*, and is rated as a CORE *A++* conference (data from 2021).

- *"Towards Material Digitization with a Dual-scale Optical System"*
  Elena Garces, Victor Arellano, Carlos Rodriguez-Pardo, David Pascual, Sergio Suja, Jorge Lopez-Moreno; *ACM Transactions on Graphics (Proceedings of SIGGRAPH), 2023* [Gar+23]. This journal has an impact factor of 7.403, and its position in the JCR index is 9 out of 110 (Q1) in the category COMPUTER SCIENCE, SOFTWARE ENGINEERING (data from 2021).

- *"NeuBTF: Neural Fields for BTF Encoding and Transfer"*
  Carlos Rodriguez-Pardo, Konstantinos Kazatzis, Jorge Lopez-Moreno, Elena Garces; 2023
  This paper is currently under peer review.

- *"NEnv: Neural Environment Maps for Global Illumination"*
  Carlos Rodriguez-Pardo, Javier Fabre, Elena Garces, Jorge Lopez-Moreno; 2023
  This paper is currently under peer review.

Patents
Some of the work developed during my PhD has resulted in two patent applications:

- *"Generation of macro-scale material property maps from images and micro-scale proper-ties"*
  Carlos Rodriguez-Pardo, Elena Garces; *(SEDDI 2020, Filed Patent Status)*
- *"Neural synthesis of tileable textures"*
  Carlos Rodriguez-Pardo, Elena Garces; *(SEDDI 2020, Filed Patent Status)*

1.4.2 Awards & Other Merits

We include here a list of awards and other merits received throughout the development of this thesis.

- *Top Reviewer* recognition at the 2022 Conference on Neural Information Processing Systems (NeurIPS).
- *Outstanding Reviewer* recognition at the 2022 European Conference on Computer Vision (ECCV).
- Our work on *Neural Visual Attribute Transfer* [RG21] (Chapter 2) was invited to the *ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*.

- Our work on Generative Models for Tileable Texture Synthesis [RG22] (Chapter 3) was presented at *Machine Learning Tokyo* as an invited talk, and as a poster presentation at the AI for Content Creation Workshop (AI4CC) at CVPR 2022.

1.4.3 Student Supervision

During the development of this thesis, I have co-supervised the bachelor's final thesis of four computer science students:

- *"Multimodal Exploration of Textile Databases Based on Perceptual Parameters"*
  María Pilar Alcarria Peinado, 2021. *Final grade: 94/100*
- *"A Fashion Recommender System Based on Multimodal Attributes"*
  Gonzalo Llosa, 2023. *(Ongoing)*
- *"Image Generative Models for Garment Design"*
  Alfonso Chiclana, 2023. *(Ongoing)*
- *"Web Interfaces for the Exploration of Text-to-Image Generative Models"*
  Carlos Cepeda, 2023. *(Ongoing)*

1.4.4 Teaching, Talks, and Seminars

During the first half of 2023, I worked as an adjunct lecturer at Universidad Carlos III in Madrid, where I taught and led a mandatory graduate course on *Intelligent Data Analysis*, at the MSc in Computer Engineering.

Throughout my Ph.D., I had the opportunity to disseminate my work across different formal and informal events. I presented some of the work in this Ph.D. thesis on different venues, including:

- *ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D), 2022*: I had the chance to present the work in Chapter 2 on propagating spatially-varying attributes of materials, as an invited talk of our TVCG paper.

- *Machine Learning Tokyo init, May 2022*: I presented the work in Chapter 3 on tileable texture synthesis as an invited seminar.

- *AI4CC workshop at CVPR, June 2022; & ELLIS Doctoral Consortium, September 2022*: I presented the work in Chapter 3 on tileable texture synthesis as poster presentations.

- *EUROGRAPHICS, May 2023*: We will present the work in Chapter 6 on casual and perceptually-validate mechanical parameter capture of fabrics.

- *CVPR, June 2023*: We will present the work in Chapter 5 on uncertainty-aware high resolution SVBRDF digitization.

1.4.5 Reviewing Activity

During the development of this thesis, I have served as a reviewer at multiple conferences, workshops, and journals on computer vision and machine learning:

- *British Machine Vision Conference (BMVC)*: 2020
- *European Conference on Computer Vision (ECCV)*: 2022, Outstanding Reviewer
- *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 2020, 2022, 2023
- *IEEE International Conference on Computer Vision (ICCV)*: 2021, 2023
- *International Conference on Learning Representations (ICLR)*: 2022
- *International Conference on Machine Learning (ICML)*: 2023
- *LatinX in Computer Vision Research Workshop*: 2023
- *Neural Information Processing Systems (NeurIPS)*: 2022 Top Reviewer, 2023
- *The Visual Computer Journal (TVCJ)*: 2020-2022
- *Women in Computer Vision Workshop (WiCV)* : 2022, 2023

1.4.6 Industrial Research Projects and Commercial Products

This thesis has been developed within an industrial Ph.D. in SEDDI, a startup based in Madrid, Spain, which has the goal of providing deep tech solutions to the textile and fashion industries. A significant part of the work presented in this thesis is part of commercially-available products and industry research projects. Notably, some of the algorithms (or variations thereof) presented in Chapters 2, 2.C and 5 are part of Textura.ai, an AI-powered commercially available material digitization product. Further, the *drape similarity metric* presented in Chapter 6 has also been a part of a project developed for a textile manufacturer.