Towards Material Digitization with a Dual-scale Optical System Supplementary Material

ELENA GARCES, Universidad Rey Juan Carlos & SEDDI, Spain
VICTOR ARELLANO, SEDDI, Spain
CARLOS RODRIGUEZ-PARDO, Universidad Rey Juan Carlos & SEDDI, Spain
DAVID PASCUAL-HERNANDEZ, SEDDI, Spain
SERGIO SUJA, SEDDI, Spain
JORGE LOPEZ-MORENO, Universidad Rey Juan Carlos & SEDDI, Spain

ACM Reference Format:
Elena Garces, Victor Arellano, Carlos Rodriguez-Pardo, David Pascual-Hernandez, Sergio Suja, and Jorge Lopez-Moreno. 2023. Towards Material Digitization with a Dual-scale Optical System Supplementary Material. *ACM Trans. Graph.* 42, 4, Article 1 (August 2023), 9 pages. https://doi.org/10.1145/3592147

1 SUMMARY OF CONTENTS

In this supplementary material, we include a detailed description of the calibration process (Section 2), an evaluation of the precision of the device (Section 3), further implementation details (Section 4), the formulation of the SVBSDF (Section 5), an analysis of the optimal number of lights for the micro-fitting process (Section 6), and extensive implementation details of the mesoscale propagation (Section 7). We also provide a webpage where all the materials can be explored and downloaded.

2 WHITE BALANCING AND RADIOMETRIC CALIBRATION

The goal of this process is to obtain relative and white-balanced radiance values for the images we take with the device using all the available cameras and lights. This process typically follows standard procedures using calibration targets such as colorcheckers, greycards, or spectralon-coated materials; to provide correct light reflection measurements these materials are mostly opaque. Our setup, besides reflection, also captures transmission with a set of lights located at the back of the holder, making these conventional calibration methods unsuccessful to obtain global radiance measurements. Therefore, we have implemented a custom calibration process using a white diffuse transmissive sheet, Zenith Polymer® White Diffuser, to jointly calibrate reflectance and transmission. According to the specs, the sheet has a transmission of 38%, 40% and 42% in the RGB bands, respectively, and pure Lambertian material reflection. However, as this value is not provided by the manufacturer, measuring it is one of the first steps of our calibration process. Note that all the computations are done using RGB linearized HDR images [Debevec and Malik 2008].

Anchor Values. As anchor measured reflectance, r_{wp}, we use the value provided by the manufacturer for the white patch of an X-Rite ColorChecker® Classic calibration target. We capture both this patch and the white sheet with the mid-range camera under diffuse illumination, obtaining i_{wp} and i_{ws}, respectively. Given these values and the known reflectance r_{wp}, we can derive the reflectance of the white sheet as:

$$r_{ws} = i_{ws} i_{wp}^{-1} r_{wp} \tag{1}$$

Radiometric Calibration Process. To calibrate the whole system, we follow the typical white balancing approach, however, using the measured reflectance of the white sheet, r_{ws}, instead of the value of the standard greycards. As mentioned before, this is necessary to calibrate reflectance and transmittance jointly. During calibration, we obtain one irradiance value for each pair of camera and directional LED r_c using the white sheet. Then, we calibrate each new input image i_c as: $\hat{i}_c = i_c r_c^{-1} r_{ws}$ In the case of the micro camera, we obtain that value by averaging every pixel in the captured image because at that scale even Spectralon exhibits specularities [Nam et al. 2016]. For the mid-range and polar cameras, we only average the pixels in the images that fall within the area captured by the micro camera. In that fashion, we avoid taking into account regions that are not evenly illuminated by the directional LEDs. For calibrating the micro camera, which has a polarizing filter, we use as reference value the sum of both modes of polarization, which are a better approximation of the irradiance that reach the other cameras that lack polarization filters. All the calibration values are pre-computed and stored excepting the diffuse images taken with the mid-range camera, which are calibrated on-the-fly using two markers placed in the holder frame, due to the bad repeatability of captures with DSLR cameras [Schwartz et al. 2014]. In order to avoid inconsistencies, these markers have been built using the same material as the white sheet used in the rest of the process.

Color accuracy is further improved by optimizing a Color Correction Matrix (CCM) that transforms the measured irradiance values into calibrated RGB color values, using as a reference the

Authors' addresses: Elena Garces, elena.garces@urjc.es, Universidad Rey Juan Carlos & SEDDI, Spain; Victor Arellano, victor.arellano@gmail.com, SEDDI, Spain; Carlos Rodriguez-Pardo, carlos.rodriguezpardo.jimenez@gmail.com, Universidad Rey Juan Carlos & SEDDI, Spain; David Pascual-Hernandez, david.pascual@seddi.com, SEDDI, Spain; Sergio Suja, sergio.suja@seddi.com, SEDDI, Spain; Jorge Lopez-Moreno, jorge@jorg3.com, Universidad Rey Juan Carlos & SEDDI, Spain.

0730-0301/2023/8-ART1 $15.00
https://doi.org/10.1145/3592147

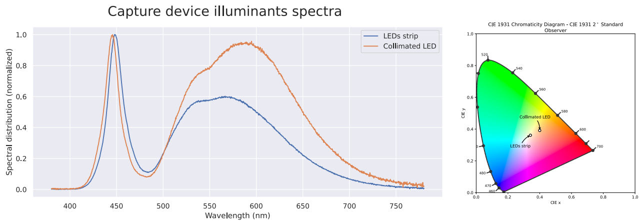

Fig. 1. Measured spectrum (left) and color (right) of the two different lighting setups.

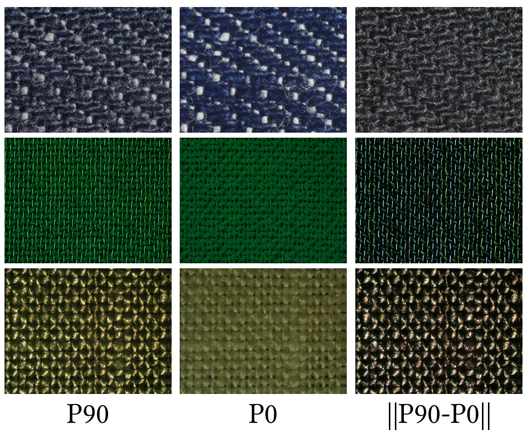

Fig. 2. From left to right: Image captured with polarization mode P90 with a directional light, corresponding image captured with polarization mode P0 and the absolute difference between the two polarization configurations. We show a two-colored DENIM material (top), a GREEN SATIN (middle), with highly specular vertical yarns, and a LINEN (bottom).

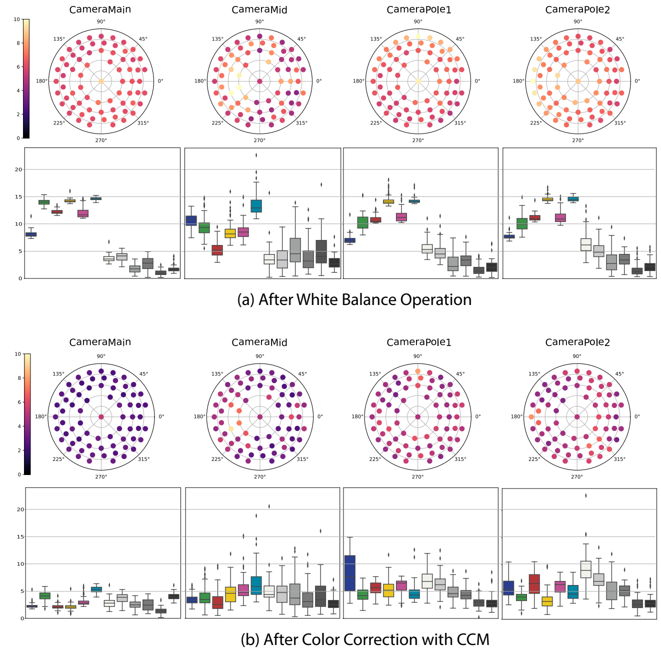

Fig. 3. (a) Results after white balance operation and before color correction using Color Correction Matrix (CCM). (b) Results after color correction using CCM. (a-b) Top Row: DeltaE2000 per camera and lighting projected in the hemisphere. The central circle is the error with diffuse lighting. Note that we do not plot grazing angles with azimuth greater than 72 degrees singe the images taken at those position contain self-occlussions. (a-b) Bottom row: Box plot containing aggregated color values for all the directional LEDs per color patch. Note that while the main camera has pretty uniform calibrated values, the other cameras present higher variability due to their non-orthogonal positions in the dome. Although the grayscale patches present accurate results after the white balancing operation (a), to obtain accurate color it is necessary to perform this extra color calibration step using CCM, as show in (b).

X-Rite ColorChecker Classic calibration target. Following previous work [Varghese et al. 2014], we find a $3x3$ CCM by solving a non-linear optimization problem in which the perceptual error ($CIE2000\ \Delta E$) between measured and reference color values is minimized using the Nelder-Mead algorithm [Gao and Han 2012]. In Section 3, we show the color accuracy after each process.

3 EVALUATION OF THE OPTICAL DEVICE

We empirically characterize the precision of the most critical pieces of the system: the color accuracy of the multi-camera setup, the sharpness of the microscopic camera, the accuracy of the polarized measurements, and the uniformity of the collimation of the directional lights.

Color Calibration. We report the color accuracy of our device and cameras using the X-Rite ColorChecker™ Classic calibration target. Figure 3 shows the average error for the 24 patches of the calibration target for each pair of cameras and lights, before and after the color correction. We observe that, although the white balance operation provides reasonable results for the grey-ish shades, the colored patches greatly benefit from this step. The average precision is below perception (less than three) for the micro camera, while the other cameras suffer from more imprecision. Measurements for the micro camera are done using the sum of both polarization modes. As can be observed in the polar cameras plot of the figure, most of the error is concentrated in the opposite position to where the camera is located, which indicates that the specular reflection might be the cause of the higher error.

Sharpness. Building a microscopic optical system is challenging and very precise optics and sensors are needed to avoid common artifacts such as vignetting, out of focus blur, or lens aberrations. For characterizing the sharpness of our microscopic camera, we use as reference a 1.5" x 1.5" grid distortion target [1]. More specifically, the target features a grid of dots with fixed frequency in which the distance between dot centers is $125\mu m$. Given a micro scale capture of this target, as the one in Figure 4 (a) (bottom right), we detect each dot center and crop a squared region of 64x64 pixel size around it. Then, we get a 1D projection of this region by averaging every row and column as it can be seen in Figure 4 (a) (top right). Finally,

[1] https://www.thorlabs.com/thorproduct.cfm?partnumber=R2L2S3P1

we estimate sharpness as the ratio between the maximum gradient in the 1D projection of the captured dot and the maximum gradient of its *ideal* counterpart. These gradients are estimated using second order central differences. In order to avoid a measurement biased by image contrast, the binary *ideal dot* minimum and maximum values are defined by the minimum and maximum values of the captured average dot image. Results are shown in Figure 4 (b), image sharpness is similar between regions, but some areas in the corners are less defined. These differences might be caused by both lens distortion and slight misalignment between sensor and sample planes.

Polarization. We empirically measure if our theoretical polarization angles match the real ones. For each *directional* LED, we vary the camera polarization angle and record the average intensity. We then compute the maximum and minimum angles values of the curve and compare them with their theoretical counterparts. We observe an average error of 4.43 degrees and always below 12 degrees. Figure 6 (a) shows and example of the curve as well as images obtained at minimum (P0) and maximum (P90) peaks of intensity.

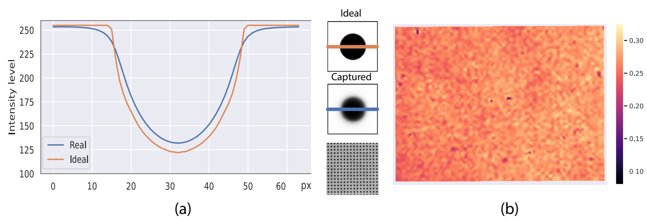

Fig. 4. Sharpness evaluation. (a) Comparison between an ideal dot and a captured one, this difference is used to evaluate the sharpness. (b) Our sharpness measurement results when capturing at microscale level using diffuse illumination. Lighter values correspond to higher sharpness levels.

Collimation. Our system requires uniform and parallel directional illumination reaching the microscale area to ensure an accurate fit. To measure this, we capture an image of the *white sheet* and compute the standard deviation of the luminance channel for each directional LED. As shown in Figure 6 (b), this deviation is below 0.15 for all the LEDs, suggesting that the collimation is accurate.

4 IMPLEMENTATION DETAILS

Our system takes around 55 minutes to fully capture a photometric dataset of a material. For the micro optimization, our device takes HDR images with dynamically computed exposure values. However, the dynamic range may be underestimated at this level for highly specular fabric materials. Thus, to prevent issues during the optimization, we limit the maximum value to two in the images and renders without affecting the final appearance. The normals are re-projected from a circular 2D projection to a square space using elliptical grid mapping [Fong 2015], giving more area to places of the circle closest to the border. We set minimum or maximum bounds to some of our maps during some iterations, which facilitates convergence to a better minima. Concretely, we constraint roughness to a minimum of 0.3 during 60% of the iterations of the fit, IOR to

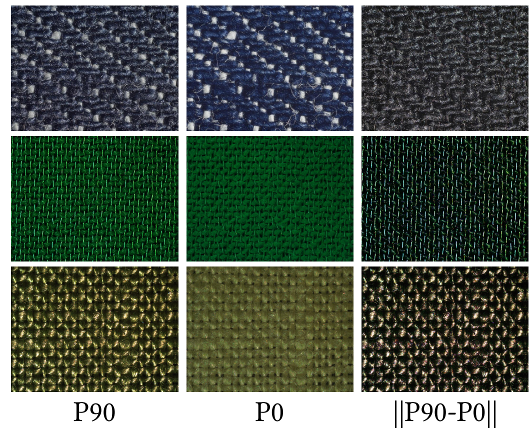

Fig. 5. From left to right: Image captured with polarization mode P90 with a directional light, corresponding image captured with polarization mode P0 and the absolute difference between the two polarization configurations. We show a two-colored DENIM material (top), a GREEN SATIN (middle), with highly specular vertical yarns, and a LINEN (bottom).

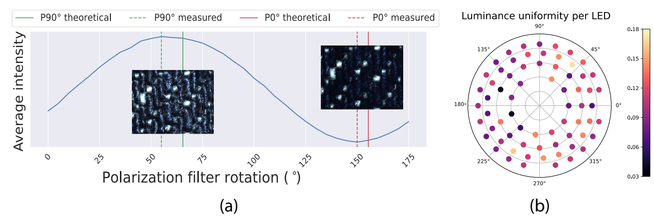

Fig. 6. (a) Polarization Example. Example of the impact of polarization for a directional LED. Average intensity of the captured image (y-axis) varying the angle of polarization of the micro camera (x-axis). The insets are captured images at those peak angles. (b) Collimation Test. Illumination uniformity for each directional light in our dome, measured as the standard deviation of the normalized luminance channel on the diffuse *white sheet*.

a minimum of 0.1 during 80% and anisotropy to both a minimum of 0.1 (during 90% of the fit) and a maximum of 0.9 (during 80%). We run 700, 500, and 600 iterations for the first, second, and third steps of the fit, respectively. On average, it takes around 317 seconds. The optimization process is done in GPU, through a differentiable renderer implemented in PyTorch [Paszke et al. 2017]. Training and evaluation of the mesoscale propagation algorithm takes approximately 15 minutes on a NVIDIA 1080Ti GPU for the single microscale capture configuration, and 25 when multiple captures are available. Futher implementation details regarding the mesoscale propagation step are described in the supplementary material.

4.1 Flyouts

We propose an algorithm for image-based fly-out detection that leverages our polarized directional LEDs. First, we build a fly-out confidence map as the difference between the most grazing and the most orthogonal illuminations in our dome. With grazing illuminations, fly-outs are well lit, while the rest of the fabric is shadowed. We use polarized images at P90 to maximize the visibility of these

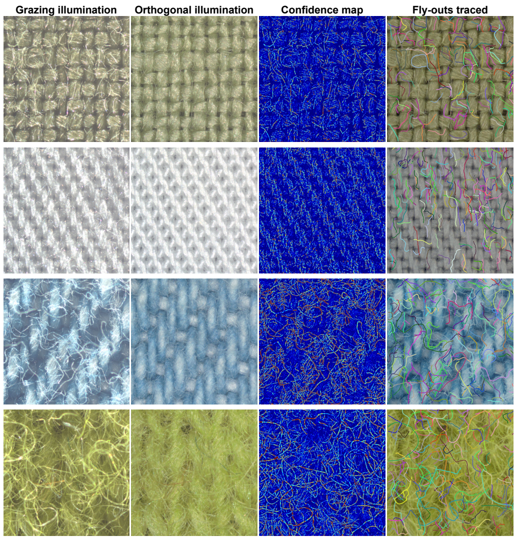

Fig. 7. Overview of our fly-out fibers detection algorithm. From left to right, examples of grazing and orthogonal illuminations, confidence maps, and traced fly-outs for different materials.

fibers in the images, dominated by specular reflections. In order to convert pixel values to probabilities, we divide our confidence map by its total value, i.e. the sum of every pixel value. Once we have the confidence map, we initialize random seeds across the image using as probabilities the values of said map. It is important to note that the number of required random seeds is dependent on the image size. In our particular application, a fixed number of 10000 seeds has shown good results. For each random seed, we trace a fly-out following the path of minimum gradient. When all the fly-outs have been traced, we filter those that are not long enough. More specifically, we filter fly-out paths shorter than 150 pixels, which given our resolution would be equal to 0.27mm, i.e. fly-outs smaller than the width of an average yarn. Furthermore, if the information is available from the fiber twist estimation, we also filter the fibers that follow the same direction as the yarns underneath. Finally, we look for overlap between fly-outs and remove possible duplicates to get our final result. Results for each step of the process for different materials are shown in Figure 7.

5 FULL SVBSDF MODEL EXPLAINED

We use the dielectric BSDF material model of Disney [Burley 2015] that allow easy manipulation for artists. The differences with respect to most common models based on Disney 2012 [Burley and Studios 2012; Karis and Games 2013] are: First, it uses real IOR instead of Schlick approximation. Second, we use a single specular tinted lobe for all the angles. Finally, for transmittance, we use a basic model with a single diffuse lobe.

Full Equation.

$$f = (\text{diffuse} + \text{specular})\ (\mathbf{n} \cdot \mathbf{l}) + \text{transmittance}\ (\mathbf{m} \cdot \mathbf{l}) \qquad (2)$$

where $\mathbf{v}$, and *Light*, $\mathbf{n}$, and $\mathbf{m}$ are the view direction, light direction, surface micro normal, and surface macro normal. We assume a local frame in this equation, thus, $\mathbf{m} = [0, 0, 1]$. t is the diffuse color of the transmittance.

$\mathbf{l}$	Direction of incident light
$\mathbf{v}$	Direction of view
$\mathbf{h}$	Half angle for reflection, $\mathbf{h} = (\mathbf{l}+\mathbf{v})/\lvert\lvert\mathbf{l}+\mathbf{v}\rvert\rvert$
$\mathbf{n}$	Micro-surface normal
$\mathbf{m}$	Macro-surface normal
$\theta_{\mathbf{l}}, \theta_{\mathbf{v}}$	Angle with the normal $\mathbf{n}$ of $\mathbf{l}$ and $\mathbf{v}$
η_1, η_2	Index of refraction for two mediums
$\mathbf{r}$	Angle of refracted light
$\mathbf{t}_g^x, \mathbf{t}_g^y$	Major axis of anisotropy and its perpendicular
b	basecolor or diffuse albedo
σ_r	roughness
α	degree of anisotropy
ρ_s	specular tint coefficient

Table 1. Common notation

5.1 Diffuse Term.

$$\begin{aligned}
\text{diffuse} &= \frac{\mathrm{b}}{\pi}(1-0.5F_{\mathbf{l}})(1-0.5F_{\mathbf{v}}) + f_{\text{retro-reflection}}\\
f_{\text{retro-reflection}} &= \frac{\mathrm{b}}{\pi}R_R(F_{\mathbf{l}} + F_{\mathbf{v}} + F_{\mathbf{l}}F_{\mathbf{v}}(R_R-1))\\
F_{\mathbf{l}} &= (1-cos(\theta_{\mathbf{l}}))^5\\
F_{\mathbf{v}} &= (1-cos(\theta_{\mathbf{v}}))^5\\
R_R &= 2\sigma_r \cos^2\theta_{\mathbf{h}}
\end{aligned}$$

5.2 Specular Term.

$$\text{specular} = \text{specular-tint}\frac{D(\mathbf{h})G(\mathbf{l},\mathbf{v})F(\mathbf{l})}{4(\mathbf{l}\cdot\mathbf{n})(\mathbf{v}\cdot\mathbf{n})}, \qquad (3)$$

Specular Tint

$$\begin{aligned}
\text{specular-tint} &= \text{chr}(\mathrm{b})\rho_s + 1 - \rho_s,\\
\text{chr}(\mathrm{b}) &= \frac{\mathrm{b}}{\max(\mathrm{b}_r, \mathrm{b}_g, \mathrm{b}_b)}
\end{aligned} \qquad (4)$$

Distribution GTR2aniso D as an anisotropic GGX microfacet model , but with $a_x \neq a_y$ as our model does consider anisotropy:

$$\begin{aligned}
D_{\text{GTR}_2\text{aniso}}(\mathbf{h}) &= \frac{1}{\pi\alpha_x\alpha_y}\frac{1}{\left((\mathbf{h}\cdot\mathbf{t}_g^x)^2/\alpha_x^2 + (\mathbf{h}\cdot\mathbf{t}_g^y)^2/\alpha_y^2 + (\mathbf{h}\cdot\mathbf{n})^2\right)^2}\\
\alpha_x &= \max\{0.001, \sigma_r^2(1-0.9\alpha)^{-1/2}\}\\
\alpha_y &= \max\{0.001, \sigma_r^2(1-0.9\alpha)^{1/2}\}
\end{aligned}$$

Geometric Term G as the separable Smith shadowing function for GGX microfacets, derived by Heitz et al. [2014].

$$G(\mathbf{l}, \mathbf{v}) = G(\mathbf{l}) * G(\mathbf{v}) \tag{5}$$

$$G(\mathbf{v}) = \frac{1}{1 + 0.5\Lambda(\mathbf{v})}$$

$$\Lambda(\mathbf{v}) = -1.0 + \sqrt{1.0 + \left(\frac{\cos\phi_{\mathbf{v}}\alpha_x + \sin\phi_{\mathbf{v}}\alpha_y}{\sin\theta_{\mathbf{v}}} \tan\theta_{\mathbf{v}}\right)^2}$$

Fresnel F as the full Snell Fresnel equations:

$$F(\mathbf{l}) = \frac{1}{2}(R_{\mathrm{p}} + R_{\mathrm{s}}) \quad \text{(for unpolarized illumination, used when rendering)}$$

$$F(\mathbf{l}) = R_{\mathrm{s}} \quad \text{(for S-polarized illumination, used when fitting)}$$

$$R_{\mathrm{s}} = \left(\frac{\eta_1 \cos\theta_{\mathbf{l}} - \eta_2 \cos\theta_{\mathbf{r}}}{\eta_1 \cos\theta_{\mathbf{l}} + \eta_2 \cos\theta_{\mathbf{r}}}\right)^2 \tag{6}$$

$$R_{\mathrm{p}} = \left(\frac{\eta_2 \cos\theta_{\mathbf{r}} - \eta_1 \cos\theta_{\mathbf{l}}}{\eta_2 \cos\theta_{\mathbf{r}} + \eta_1 \cos\theta_{\mathbf{l}}}\right)^2 \tag{7}$$

where $\mathbf{r}$ is the refracted angle that can be ignored, as by Snell laws $\eta_1 \sin\theta_{\mathbf{l}} = \eta_2 \sin\theta_{\mathbf{r}}$.

5.3 Transmittance Term.

Importantly, for this term we use the normals of the macro surface $\mathbf{m}$ instead of the micro ones. Transmittance is defined as the diffuse transmittance image as coming from capture: transmittance = t.

6 NUMBER OF LIGHTS

We have explored the effect in micro-fit accuracy produced by the number and orientation of directional LED captures used as input. Our goal is to minimize the number of images to reduce computational cost, but within reasonable error boundaries. In Figure 8 we show some of the configurations tested, inspired by [Nielsen et al. 2015]. We departed from a densely sampled space using All views and reduced the amount by two Mod 2 and by three Mod 3. We futher refined the selection by densely sampling azimuth angles Minimal and exploring equal-elevation bands Row 2, 4 and Row 2, 5 (Figures 8). The best results were obtained with All views followed by Mod 2 and Row 2, 4. Note that some materials like Velour-Green and Leather-Rhombus necessitate uniform sampling, while others like Denim or Interlock show similar performance reducing the sampling ratio. The Minimal sampling strategy is in no case recommended, as it introduced little useful information due to inter-yarn cast shadows dominating the images.

7 MESOSCALE PROPAGATION DETAILS

7.1 Problem Formulation

As in [Rodriguez-Pardo and Garces 2021], our goal is to transfer spatially-varying attributes of a material captured at microscale to larger samples of the same material. We thus formulate this problem using an image-to-image translation framework. For training, our methods takes as input a *set of D photometric datasets* of a material $\mathcal{I}_L^D$, and their pixel-wise corresponding SVBSDFs M_P^D, each comprised of a list of P maps: $p \in$ {albedo, roughness, transmittance, IOR, anisotropy, tangents, normals, specularTint, opacity}. Each photometric dataset is comprised of a set of RGB images of the material captured under different illumination conditions: $\mathcal{I}_L^d = \{i_l^d | i_l^d \in \Re^{n\times m\times 3}\}$, $|\mathcal{I}_L^d| \geq 1, \forall d \in D$. In contrast to [Rodriguez-Pardo and Garces 2021], which only allowed to learn from a single photometric dataset and to transfer a single attribute, our extended approach allows for learning from multiple datasets $|D| \geq 1$ and multiple property maps $|P| \geq 1$ using a single model. We train a model T which learns to transfer from each image in each photometric dataset to its correspoding SVBSDF: $T(i_l^d) \approx M_P^d, \forall d \in D, l \in L$. During evaluation, we use a guidance image of the material X, which represents a larger sample of it, and estimate its corresponding property maps: $M^{\mathrm{X}} \leftarrow T(\mathrm{X})$. At evaluation time, this guidance image X may be captured with different conditions to those in the training dataset, including different camera or illumination conditions.

7.2 Datasets

We use a more comprehensive photometric dataset to what is proposed in [Rodriguez-Pardo and Garces 2021]. In particular, for a single microscale capture, we use all the light sources used in the fitting algorithm, as well as diffuse-lit microscale images. Additionally, we use both polarization modes separately and, for the same light source, we construct an extra image by averaging the images taken under both polarization configurations. The total size of each photometric dataset is of $|\mathcal{I}_L^d| = 108, \forall d \in D$, four times more data compared with the more limited 27 images proposed in [Rodriguez-Pardo and Garces 2021]. For evaluation, our guidance image is taken with the mid-range camera and *diffuse* illumination, at a resolution of 4072×4072 pixels and a surface area of 10×10 centimeters. We typically use a single microscale capture ($|D| = 1$). However, for multi-colored or highly heterogeneous materials in which a single capture is not enough to fully represent the material variability, we capture as many datasets as needed.

7.3 Loss Function

Our loss function is the combination of two terms: weighted pixel-wise losses for each target map, and a multi-map style loss:

$$\mathcal{L} = \sum_p \lambda_p \mathcal{L}_{pixel_p} + \lambda_{style}\mathcal{L}_{style} \tag{8}$$

$\mathcal{L}_{pixel}$ is the $\mathcal{L}_1$ norm weighted per map in the SVBSDF, λ_p. $\mathcal{L}_1$ produces sharper results than higher-order alternatives [Rodriguez-Pardo and Garces 2021], such as $\mathcal{L}_2$. For the $\mathcal{L}_{style}$ loss term, we follow recent work on multi-channel texture synthesis, which extend deep perceptual losses for style transfer to the SVBRDF synthesis problem [Chambon et al. 2021]. This component of the loss function acts as a regularizer and allows the model to generate maps which better preserve the apperance of the ground truth training data.

7.4 Implementation Details

Model Design. We use a lightweight U-Net [Ronneberger et al. 2015] architecture, with a few modifications based on recent work to maximize its efficiency and the quality of its outputs. We specify the full model architecture and model sizes on Figure 9. In every

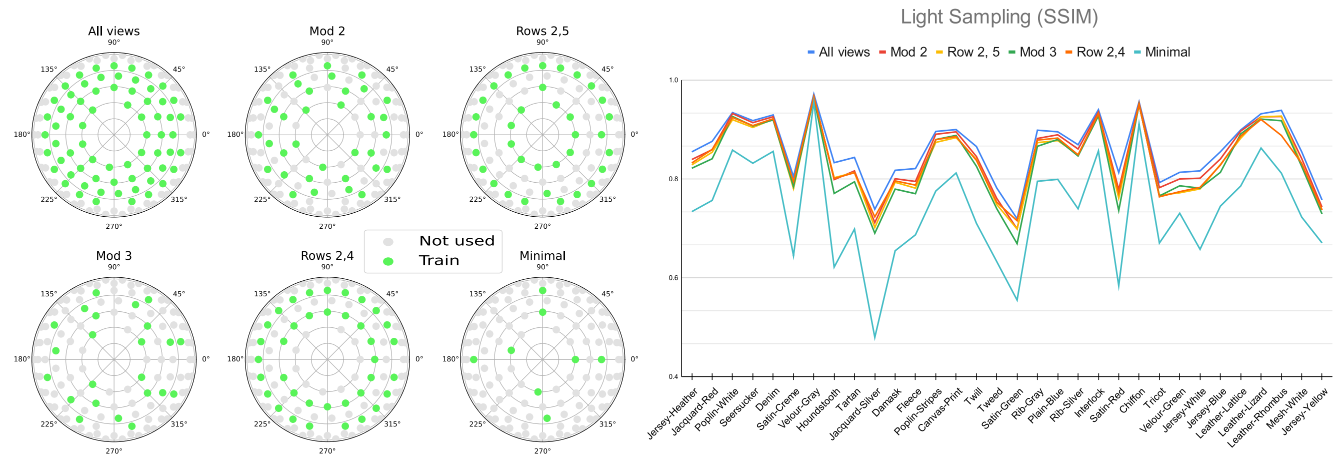

Fig. 8. Micro fit input evaluation. Left: Sampling strategies showcasing in Green the LEDs used. Right: SSIM error per fabric in the dataset. In (a) All-views, 127 lights were used. In (b) Mod 2, one out of two images were evenly sampled (34 images). In (c) Row 2, 5 half of the elevation rows 2 and 5 are used (19 images). In (d) Mod 3, one out of three images were evenly sampled (23 images). In (e) Row 2 and 4 half of the elevation rows 2 and 4 are used (19 images). In (f) Minimal an example of minimal configuration with a very reduced set of 8 images.

convolutional block of the model, we use residual connections [Diakogiannis et al. 2020; He et al. 2016], for better training convergence and preservation of details in the input images. We use 1×1 convolutions on these residual connections. We use a single decoder for each map, so as to maximally preserve their individual appearance and statistics. This has been proposed on recent work on texture synthesis [Rodriguez-Pardo and Garces 2022], material capture [Deschaintre et al. 2021] or intrinsic decomposition [Janner et al. 2017]. On the convolutional blocks, we leverage Group Normalization [Wu and He 2018]. Upsampling is done using transposed convolutions. As shown on Figure 9, the model has four hidden layers on the encoder and each decoder, however, we vary the size of those layers depending on whether the model is trained with a single microscale capture (for which we use a width factor of $W = 16$) or multiple ($W = 32$). Every other implementation detail in the model (stride, bias, pooling) follows [Rodriguez-Pardo and Garces 2021; Ronneberger et al. 2015].

Training and Evaluation. We use PyTorch [Paszke et al. 2017], Torchvision [Marcel and Rodriguez 2010], and Kornia [Riba et al. 2020] for training. We leverage mixed precision training and automatic gradient scaling [Micikevicius et al. 2017], to accelerate the training process and regularize the models. Optimization is done using Adam [Kingma and Ba 2014] for 5000 iterations, with a learning rate of 0.002, a batch size of 40 and no weight decay. This process takes around 15 minutes when training with a single microscale capture, and 25 when using multiple captures as input. We evaluate the guidances images using half precision. We measure these times on a NVIDIA 1080Ti GPU.

Data Augmentation. We follow the same training procedure defined in [Rodriguez-Pardo and Garces 2021]. However, we do not use random rotations or shears, and remove the color invariance data augmentation whenever multiple microscale captures are available. Further, we introduce random Gaussian Blurs for data augmentation, using a $p = 0.5$, a kernel size of 5 and sigma selected uniformely at random for each element in each batch: $\sigma \sim \mathcal{U}(0.1, 11)$. As in [Rodriguez-Pardo and Garces 2021], we also perform random cropping, using crops of 128×128 pixels. For each element in each batch during training, we randomly choose the photometric dataset $d \in D$, and, from it, a random light source $l \in L$.

Loss Function. We weight the loss function as follows: For the pixel wise loss: $\lambda_{normals} = 3, \lambda_{IOR} = 1, \lambda_{roughness} = 1, \lambda_{albedo} = 3, \lambda_{tangent} = 3, \lambda_{anisotropy} = 1, \lambda_{specularTint} = 1, \lambda_{transmittance} = 3, \lambda_{opacity} = 1$. The perceptual component [Chambon et al. 2021] is weighted with $\lambda_{style} = 0.25$, and we use the *AlexNet* variant of *LPIPS* [Zhang et al. 2018] as our backbone, which provides a powerful yet lightweight loss for texture transfer, as shown in [Rodriguez-Pardo and Garces 2021].

7.5 Results

In Figure 10, we show an ablation study on the improvements of the maps propagation method proposed in [Rodriguez-Pardo and Garces 2021]. We build upon their proposed implementation, removing their random shifts and rotations, and make progressive changes, in order, to the dataset size, model architecture, loss function, and data augmentation. We observe high-quality propagations, but the baseline sometimes produces overly smooth outputs (Fleece, Leather-Brown, Rib-Silver, or Leather-Lizard). Our proposed increased dataset and model architecture modifications enhance the maps sharpness, and our perceptual loss function allows for the propagation of additional details. Our final model removes the color augmentation proposed in [Rodriguez-Pardo and Garces 2021] whenever multiple microscale captures are available, and introduces a random blurs during training, allowing for achieving the highest quality results and for accurate base color propagations in challenging cases, as in TARTAN.

REFERENCES

Brent Burley. 2015. Extending the Disney BRDF to a BSDF with integrated subsurface scattering. *Physically Based Shading in Theory and Practice SIGGRAPH Course* (2015).

Brent Burley and Walt Disney Animation Studios. 2012. Physically-based shading at disney. In *ACM SIGGRAPH*, Vol. 2012. vol. 2012, 1–7.

Thomas Chambon, Eric Heitz, and Laurent Belcour. 2021. Passing Multi-Channel Material Textures to a 3-Channel Loss. In *ACM SIGGRAPH 2021 Talks*. 1–2.

Paul E Debevec and Jitendra Malik. 2008. Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH 2008*. 1–10.

Valentin Deschaintre, Yiming Lin, and Abhijeet Ghosh. 2021. Deep polarization imaging for 3D shape and SVBRDF acquisition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 15567–15576.

Foivos I Diakogiannis, François Waldner, Peter Caccetta, and Chen Wu. 2020. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing* 162 (2020), 94–114.

Chamberlain Fong. 2015. Analytical methods for squaring the disc. *arXiv preprint arXiv:1509.06344* (2015).

Fuchang Gao and Lixing Han. 2012. Implementing the Nelder-Mead simplex algorithm with adaptive parameters. *Computational Optimization and Applications* 51, 1 (2012), 259–277.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 770–778.

Eric Heitz. 2014. Understanding the masking-shadowing function in microfacet-based BRDFs. *Journal of Computer Graphics Techniques* 3, 2 (2014), 32–91.

Michael Janner, Jiajun Wu, Tejas D Kulkarni, Ilker Yildirim, and Josh Tenenbaum. 2017. Self-supervised intrinsic image decomposition. *Advances in Neural Information Processing Systems* 30 (2017).

Brian Karis and Epic Games. 2013. Real shading in unreal engine 4. *Proc. Physically Based Shading Theory Practice* 4, 3 (2013), 1.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

Sébastien Marcel and Yann Rodriguez. 2010. Torchvision the machine-vision package of torch. In *Proceedings of the 18th ACM international conference on Multimedia*. 1485–1488.

Paulius Micikevicius, Sharan Narang, Jonah Alben, Gregory Diamos, Erich Elsen, David Garcia, Boris Ginsburg, Michael Houston, Oleksii Kuchaiev, Ganesh Venkatesh, et al. 2017. Mixed precision training. *arXiv preprint arXiv:1710.03740* (2017).

Giljoo Nam, Joo Ho Lee, Hongzhi Wu, Diego Gutierrez, and Min H Kim. 2016. Simultaneous acquisition of microscale reflectance and normals. *ACM Transactions on Graphics (TOG)* 35, 6 (2016), 185–1.

Jannik Boll Nielsen, Henrik Wann Jensen, and Ravi Ramamoorthi. 2015. On optimal, minimal BRDF sampling for reflectance acquisition. *ACM Transactions on Graphics (TOG)* 34, 6 (2015), 1–11.

Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic differentiation in pytorch. (2017).

Edgar Riba, Dmytro Mishkin, Daniel Ponsa, Ethan Rublee, and Gary Bradski. 2020. Kornia: an open source differentiable computer vision library for pytorch. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 3674–3683.

Carlos Rodriguez-Pardo and Elena Garces. 2021. Neural Photometry-guided Visual Attribute Transfer. *IEEE Transactions on Visualization and Computer Graphics* (2021), 1–1.

Carlos Rodriguez-Pardo and Elena Garces. 2022. SeamlessGAN: Self-Supervised Synthesis of Tileable Texture Maps. *IEEE Transactions on Visualization and Computer Graphics* (2022).

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 234–241.

Christopher Schwartz, Ralf Sarlette, Michael Weinmann, Martin Rump, and Reinhard Klein. 2014. Design and implementation of practical bidirectional texture function measurement devices focusing on the developments at the University of Bonn. *Sensors* 14, 5 (2014), 7753–7819.

D Varghese, R Wanat, and RK Mantiuk. 2014. Colorimetric calibration of high dynamic range images with a ColorChecker chart. *Proceedings of the HDRi* (2014).

Yuxin Wu and Kaiming He. 2018. Group normalization. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 3–19.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 586–595.

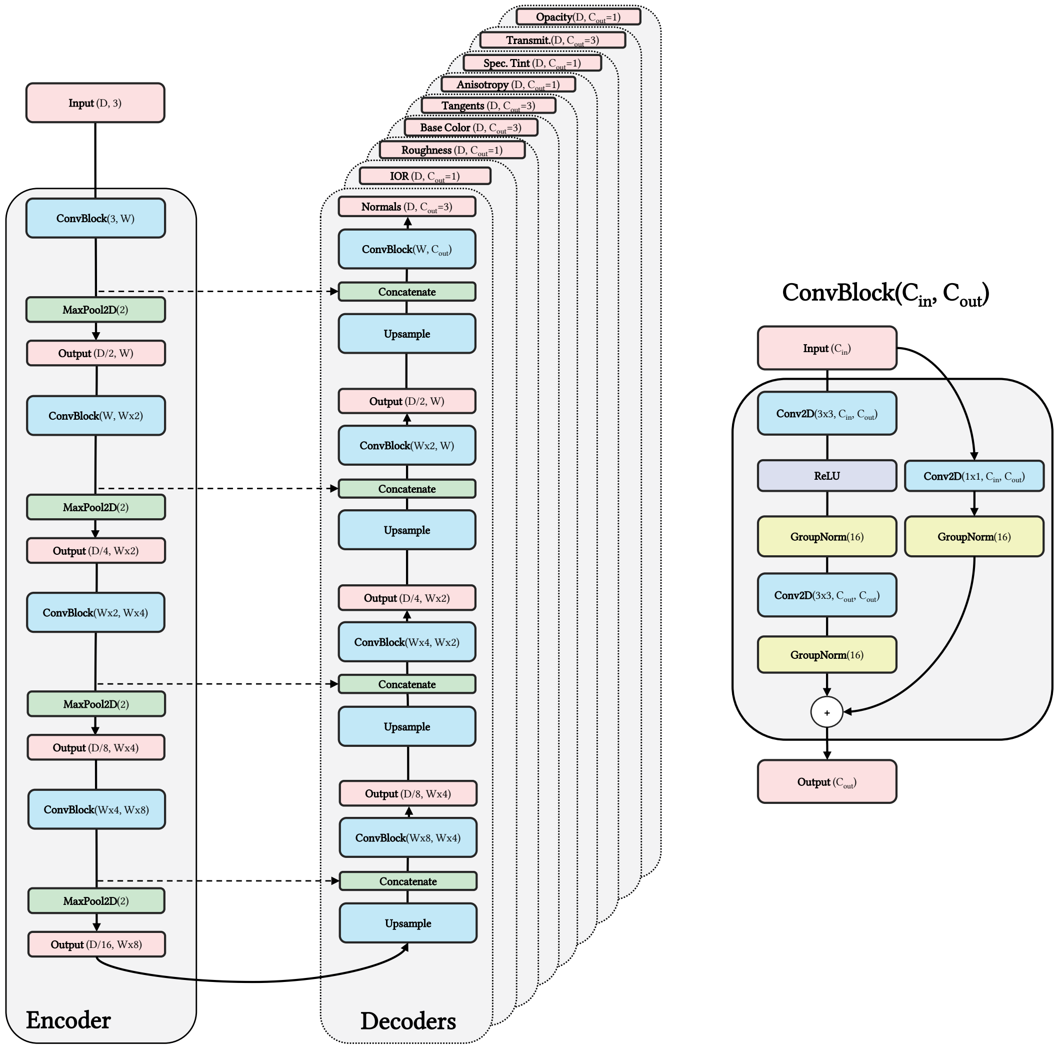

Fig. 9. A full diagram of the model architecture we use for our mesoscale propagation problem. Building upon [Rodriguez-Pardo and Garces 2021], use a lightweight U-Net [Ronneberger et al. 2015] architecture. Following previous work [Deschaintre et al. 2021; Janner et al. 2017; Rodriguez-Pardo and Garces 2022], we use a different decoder for every map we aim to transfer. Each of our convolutional blocks, illustrated on the right, contain residual connections [Diakogiannis et al. 2020; He et al. 2016] and Group Normalization [Wu and He 2018]. In red, we show the input/output dimensions (spatial, channels) of each layer; in blue, convolutional blocks and layers; in green, upsampling and concatenating operations; in yellow, normalization layers; and in purple, regularizations and non-linearities. We set W, which controls layer width, to $W = 16$ when a single microscale capture is available as training data, $W = 32$ otherwise.

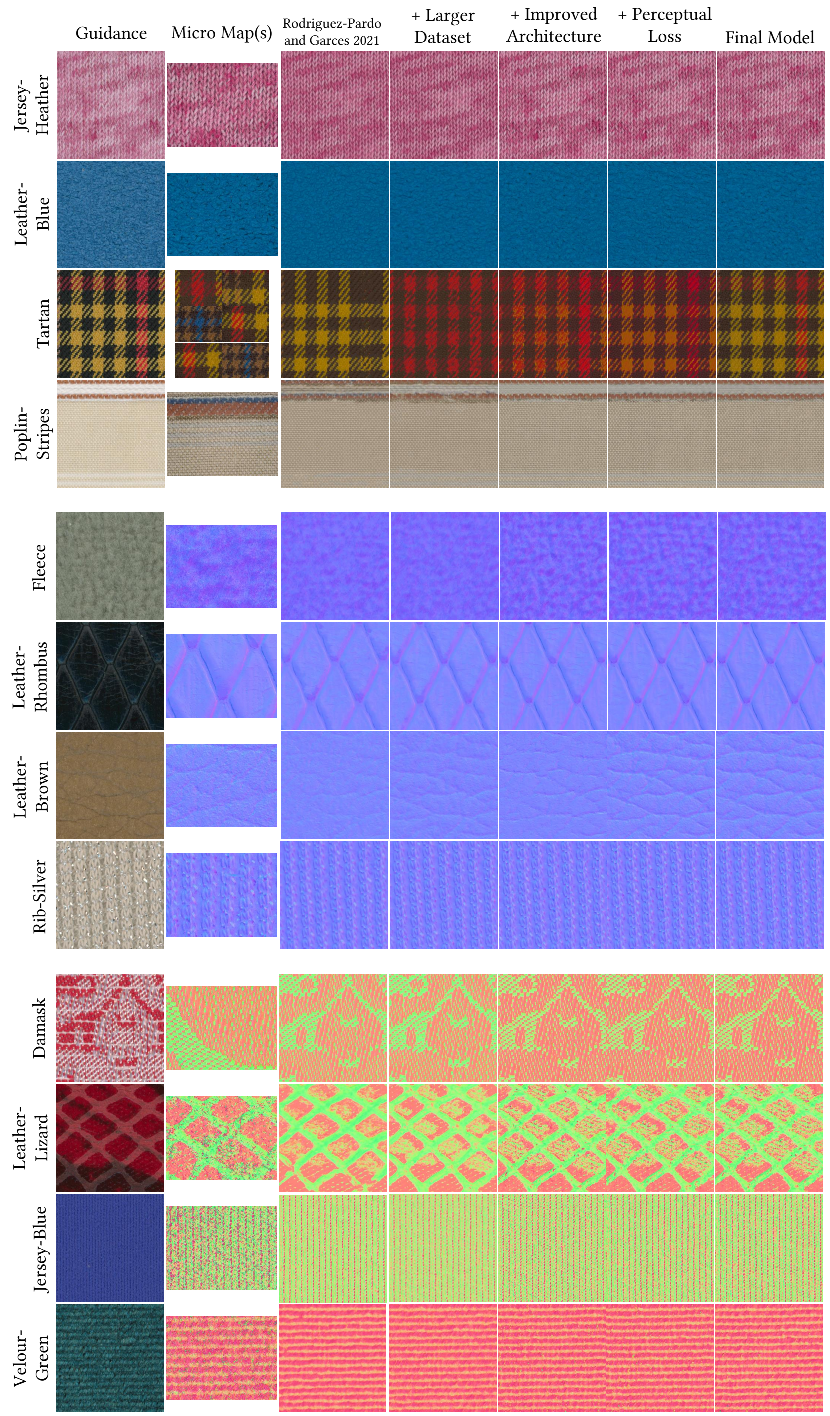

Fig. 10. Ablation study of our proposed improvements with respect to the maps propagation method proposed in [Rodriguez-Pardo and Garces 2021], on base color (top rows), normals (mid) and tangent map (bottom) transfer. We start from the implementation in [Rodriguez-Pardo and Garces 2021], which shows adequate but smooth results. On its right, we show the impact of increasing the dataset size, which tends to generate higher-quality estimations (Leather-Brown). With our improved architecture (fifth column), we achieve sharper maps (see Leather-Lizard, Rib-Silver, or Fleece). Using a perceptual loss (sixth column), we further improve the maps quality (Leather-Brown). Our final model (last column), which removes the color augmentation in [Rodriguez-Pardo and Garces 2021] and introduces random blurs to the model training, achieves the highest quality, and allows for accurate base color transfers, as shown in Tartan. Input guidance image and training micro maps are shown on the first and second columns, respectively. Best viewed in color on a screen.