



Neural reflectance transformation imaging

Tinsae G. Dulecha¹ · Filippo A. Fanni¹ · Federico Ponchio² · Fabio Pellacini³ · Andrea Giachetti¹

Published online: 17 July 2020
© The Author(s) 2020

Abstract

Reflectance transformation imaging (RTI) is a computational photography technique widely used in the cultural heritage and material science domains to characterize relieved surfaces. It basically consists of capturing multiple images from a fixed viewpoint with varying lights. Handling the potentially huge amount of information stored in an RTI acquisition that consists typically of 50–100 RGB values per pixel, allowing data exchange, interactive visualization, and material analysis, is not easy. The solution used in practical applications consists of creating “relightable images” by approximating the pixel information with a function of the light direction, encoded with a small number of parameters. This encoding allows the estimation of images relighted from novel, arbitrary lights, with a quality that, however, is not always satisfactory. In this paper, we present NeuralRTI, a framework for pixel-based encoding and relighting of RTI data. Using a simple autoencoder architecture, we show that it is possible to obtain a highly compressed representation that better preserves the original information and provides increased quality of virtual images relighted from novel directions, especially in the case of challenging glossy materials. We also address the problem of validating the relight quality on different surfaces, proposing a specific benchmark, SynthRTI, including image collections synthetically created with physical-based rendering and featuring objects with different materials and geometric complexity. On this dataset and as well on a collection of real acquisitions performed on heterogeneous surfaces, we demonstrate the advantages of the proposed relightable image encoding.

Keywords Reflectance transformation imaging · Relighting · Neural network · Autoencoder · Benchmark

1 Introduction

Reflectance transformation imaging [4,9,11] is a popular computational photography technique, allowing to capture the rich representations of surfaces including geometric details and local reflective behavior of materials. It consists

of capturing sets of images from a fixed point of view with varying light direction (and in some cases varying light wavelengths). These sets are often referred to also as multi-light image collections (MLIC). A recent survey [14] shows that there is a large number of applications exploiting interactive relighting and feature extraction from this kind of data, in different fields like cultural heritage, material science, archaeology, quality control, and natural sciences.

Typical RTI acquisitions consist of 50–100 images that are encoded into compressed “relightable images.” This is usually done by fitting the per-pixel sampled image intensity with a function of the light direction with a small number of parameters, stored then as the novel pixel data. This encoding is used in interactive relighting tools, to estimate normals and material properties or to derive enhanced visualizations [12]. A good RTI encoding should be compact and allow interactive relighting from an arbitrary direction, rendering the correct diffuse and specular behaviors of the imaged materials and limiting interpolation artifacts. These requirements are not always satisfied with the methods currently employed in practical applications [13].

✉ Tinsae G. Dulecha
tinsae.dulecha@univr.it

Filippo A. Fanni
filippoandrea.fanni@univr.it

Federico Ponchio
federico.ponchio@isti.cnr.it

Fabio Pellacini
pellacini@di.uniroma1.it

Andrea Giachetti
andrea.giachetti@univr.it

¹ Department of Computer Science, University of Verona, Verona, Italy

² ISTI-CNR, Pisa, Italy

³ Sapienza University of Rome, Rome, Italy

In this paper, we present a novel neural network-based encoding specifically designed to compress and relight a typical RTI dataset that can easily be integrated into the current RTI processing and visualization pipelines. Using a fully connected autoencoder architecture and storing per-pixel codes and the decoder coefficients as our processed RTI data, we obtain a relevant compression still enabling accurate relighting quality and limiting interpolation artifacts.

Another relevant issue in the RTI community is the lack of benchmarks to evaluate data processing tools on the kind of surfaces typically captured in real-world applications (e.g., quasi-planar surfaces made of heterogeneous materials with a wide range of metallic and specular behaviors). Relighting algorithms are typically tested on a few homemade acquisitions not publicly available. Some benchmarks do exist for Photometric Stereo algorithms (recovering surface normals), e.g., DiLiGent [20]. They could be used as well to test relighting quality. However, the images included are not representative of those typically captured in the real-world applications of RTI acquisition. Relightable images and Photometric Stereo algorithms are used mainly to analyze surfaces with limited depth, possibly made of challenging and heterogeneous materials. The performances of the different methods proposed are likely to vary with the shape complexity and the material properties. It would be, therefore, useful, to evaluate the algorithms on images of this kind, possibly with accurate lighting and controlled features.

For this reason, we propose, as a second, important contribution of the paper, novel datasets for RTI-related algorithms validation. We created a large synthetic dataset, SynthRTI, including MLIC generated with physical-based rendering techniques (PBR). This includes renderings with different sets of directional lights of three different shapes (with limited depth and variable geometric complexity) with different single and multiple materials assigned. The dataset will be useful not only to evaluate the compression and relighting tool but also to assess the quality of Photometric Stereo methods on realistic tasks. We release also another dataset, RealRTI, including real multi-light image collections with associated estimated light directions that can be used to evaluate the quality relighting methods on real surfaces acquired in practical applications of the technique. The datasets have been used to demonstrate the quality of NeuralRTI relighting.

The paper is organized as follows: Section 2 presents the context of our work, and Sect. 3 describes the proposed NeuralRTI framework and its implementation. Section 4 describes the novel datasets, and Sect. 5 presents the results of the evaluation of the novel compression and relighting framework, compared with the currently used ones, e.g., polynomial texture mapping (PTM), Hemispherical Harmonics fitting (HSH), and PCA-compressed Radial Basis Function interpolation.

2 Related work

According to a recent survey [14] on the state of the art of RTI processing and relighting of multi-light image collections, practical applications, especially in the cultural heritage domain, rely on the classic polynomial texture mapping (PTM) [9] or on Hemispherical Harmonics (HSH) [10] coefficients to store a compact representation of the original data and interactively relightable images.

An alternative to PTM and HSH has been proposed by Pitard et al. with their Discrete Modal Decomposition (DMD), [15,16]. As shown in [15], however, this method provides relighting quality quite similar to HSH if the same number of coefficients is chosen for the encoding. The method is not supported in public software packages. Radial Basis Function interpolation has been proposed to enhance the relighting quality as an alternative to simple parametric functions [5], but not suitable for interactive relighting. Ponchio et al. [17] combined principal component analysis (PCA)-based compression of the image stack and Radial Basis Function (RBF) interpolation in the light direction space to provide an efficient relightable image framework. Encoding software and web-based viewer are publicly distributed.

Despite the remarkable results obtained with neural network-based methods on other tasks related to the processing of multi-light image collections (e.g., Photometric Stereo [2,24]), no practical solutions based on them have been proposed to replace the solutions widely used in many applicative domains to create and visualize relightable images. Commonly used encoding software and viewers used in the practice only support the aforementioned methods: RTI Builder and RTI Viewer [3] support PTM and HSH encoding and relighting, Relight [17]), supports PTM, HSH and PCA/RBF.

A few neural network-based methods for image relighting have been proposed in the literature, but not in the classical RTI setting. Ren et al. [19] used neural network ensembles to interpolate light transport matrices, performing relighting based on a larger number of input images (e.g., 200–300), so that the method is not suitable for the relighting of common RTI datasets. An interesting method for image relighting was proposed by Xu et al. [23]: Their method learns from a training set of densely sampled MLIC a subset of light directions and, simultaneously, a method to relight the subset from arbitrary directions. The resulting method is global and allows the hallucination of non-local effects, but cannot be used to encode and relight classical RTI data due to the constraints on input light directions and the ability to encode different objects and materials depends on training data.

The method most closely related to ours is the one proposed by Rainer et al. [18], to recover bidirectional texture functions from multidimensional data. They adopt a

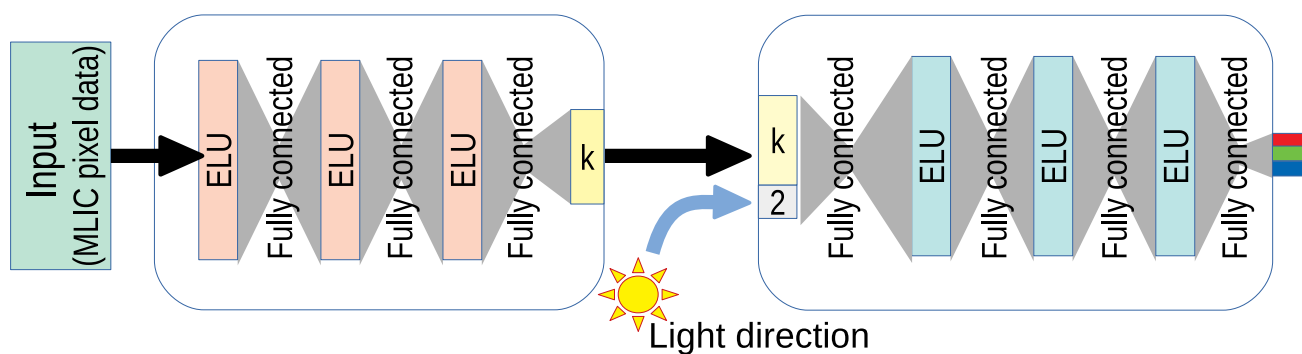


Fig. 1 Our NeuralRTI builds on an asymmetric encoder–decoder architecture. The encoder receives a per-pixel RTI data as input, i.e., a set of pixel values associated with a different lighting direction. These measurements are then passed through a sequence of fully connected layers

until a fully connected layer finally produces a k -dimensional code. The decoder concatenates the code vector with the light direction and passes them through a sequence of fully connected layers with componentwise nonlinearities. The last layer outputs a single RGB color

1D convolutional approach for encoding despite the multidimensional nature of the data and are focused on BTF compression, evaluating only the visual quality of the recovered material models. We follow a similar idea, but using a fully connected network architecture and focusing on practical applications requiring as accurate as possible recovery of details and material properties in relighted images.

3 Neural reflectance transformation imaging

The basic idea of NeuralRTI is to use the data captured in a multi-light image collection to train a fully connected asymmetric autoencoder mapping the original per-pixel information into a low-dimensional vector and a decoder able to reconstruct pixel values from the pixel encoding and a novel light direction.

Autoencoders [1] are unsupervised artificial neural networks that learn how to compress the input into a lower-dimensional code, called the latent-space representation, and then reconstruct an output from this representation that is as close as possible to the original input. In the recent past, autoencoders have been widely used for image compression [6,25,26] and denoising [8,22], as they can learn nonlinear transformations, unlike the classical PCA, thanks to nonlinear activation functions and multiple layers. We specialized this kind of architecture to create an RTI encoding that is compact like currently used RTI files and can be used as well for interactive relighting, provided better quality and fidelity to the original data.

Our encoder/decoder network, represented in Fig. 1, is trained end to end with the pixel data of each multi-light image collection to be encoded and the corresponding light directions as input. The training procedure minimizes the mean squared loss between predicted and ground truth pixel values on the set of given light directions.

The rationale of the network design is that we can then store the compact pixel code together with the coefficients of the decoder network (that are unique for all the pixel locations) as a relightable image. Given a light direction included in the original set, we can reconstruct the original images (compression), while using generic light directions we can generate relighted images. We assume that, if the sampling of the original data is reasonably good as in real-world RTI acquisitions, the network can provide better relighting results compared to traditional methods, as it can learn nonlinear light behaviors and not only the coefficients but also the relighting function is adapted to the input data (shape and materials).

The encoder network (Fig. 1, left) includes four layers, and each consists of one activation layer. The first three layers contain $3N$ units, where N is the number of input lights and the last layer is set to the desired size k of the compressed pixel encoding (9 in our tests). Each layer is equipped with the Exponential Linear Unit (ELU) activation function. Since the encoder network is used only during the training, there are no particular constraints on the architecture or its size. However, from empirical analysis, we found that nine coefficients seem sufficient to provide good relight quality and the size of intermediate layers of encoders and decoders equal to the input size provides good reconstruction results limiting overfitting.

The decoder network (Fig. 1, right) consists of three hidden layers consisting of $3N$ units each. The input is the concatenation of the pixel encoding and the 2D vector with the light direction. The output is the predicted single RGB pixel value, illuminated from the given light direction.

We use 90% of the total RTI data pixel as training data and a 10% sampled uniformly across pixel locations and light directions as the validation set. Training is performed with the Adam optimization algorithm [7] with a batch size of 64 examples, a learning rate of 0.01, a gradient decay fac-



Fig. 2 Sample images of the bas-relief object with the eight materials assigned

tor of 0.9, and a squared gradient decay factor of 0.99. The network is implemented using the Keras open-source deep learning library. After the training, the codes corresponding to pixel locations are converted to byte size with the offset+scale mapping and stored as image byte planes, and the coefficients of the decoder ($3N^2 + 5N$) are stored as a header information to be used as an input to the specific relighting tool and used at runtime to relight images to give novel arbitrary light directions.

In our tests, we trained the RTI encodings on a GeForce RTX 2080Ti machine with a single GPU, and training takes approximately one hour whereas the time required to create a 320×320 relighted image is approximately 0.007 s.

4 Evaluation datasets: SynthRTI and RealRTI

To evaluate the relighting quality, we created two datasets, SynthRTI and RealRTI.

4.1 SynthRTI

SynthRTI is a collection of 51 multi-light image collections simulated by using the Blender Cycles rendering engine. It is divided into two subsets: SingleMaterial, featuring 24 captures of three surfaces with eight different materials assigned (see Fig. 2), and MultiMaterial, with 27 captures of the same three surfaces with nine material combinations, as shown in Fig. 3. Each collection is subdivided into two sets of images, corresponding to two sets of light directions. The first is called Dome and corresponds to a classical multi-ring light dome solution with 49 directional lights placed in concentric rings in the l_x, l_y plane at five different elevation values (10, 30, 50, 70, 90 degrees). The second is called Test and includes 20 light directions at four intermediate elevation values (20,40,60,80). The idea is to use the classic dome acquisition to create the relightable images and evaluate the quality of the results using the test dataset. Figure 4 shows the Blender viewport with the simulated dome (left) and the

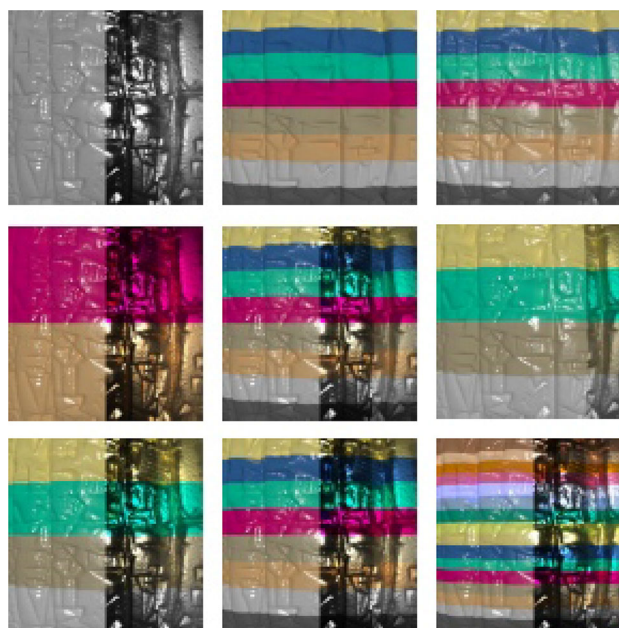


Fig. 3 Sample images of the cuneiform tablet with combinations of materials assigned

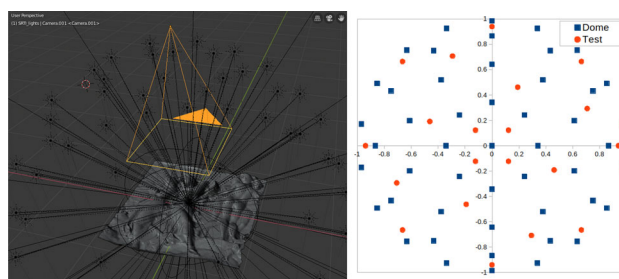


Fig. 4 Left, the virtual camera/lights configuration in the Blender viewport. We used directional (sun) lights. Right, the distribution of the light directions in the simulated dome subset (blue squares) and for the test evaluation (red dots)

light directions of the Dome and Test acquisition setups, represented in the l_x, l_y plane.

The three geometric models used to build the dataset are bas-reliefs with different depths. The first is a nearly flat surface, actually the 3D scan of an oil on canvas painting by W. Turner, performed by R.M. Navarro and found on SketchFab (<https://sketchfab.com>). The second is the scan of a cuneiform tablet from Colgate University. The third is the scan of relief in marble “The dance of the Muses on Helicon” by G. C. Freund, digitized by G. Marchal. All the models are distributed under the Creative Commons 4.0 license.

For the single-material collections, we used eight sets of cycles PBR parameters, to simulate matte, plastic and metallic behaviors, and a material with subsurface scattering. These sets are reported in Table 1. We rendered small images (320×320) with 8 bit depth.

Table 1 Parameters of the cycles principled BSDF model used to create the eight materials of the single-material dataset

#	Material	Base color	Metal	Spec	Rough	Subs. <i>R</i>
1	Matte white	0.8,0.8,0.8	0	0	0	
2	Plastic	0.8,0.8,0.8	0	1	0.4	
3	Plastic rough	0.8,0.8,0.8	0	1	0.5	
4	Plastic rough2	.25,.25,.25	0	1	0.6	
5	Plastic black	.25,.25,.25	0	1	0.4	
6	Metal smooth	0.8,0.8,0.8	1	0	0.5	
7	Metal rough	0.8,0.8,0.8	1	0	0.7	
8	Subsurface	0.5,0.5,0.5	0	1	0.5	1,0.2,0.1

Table 2 Combination of metallic, specularity, and roughness indexes used in the multimaterial dataset

	<i>r1</i>	<i>r2</i>	<i>r3</i>	<i>r4</i>	<i>r5</i>	<i>r6</i>	<i>r7</i>	<i>r8</i>
Metal	0	1	1	1	0	0	0	0
Roughness	1	0.4	0.5	0.6	0.5	0.6	0.7	0.8
Specularity	0	1	1	1	1	0	0	1

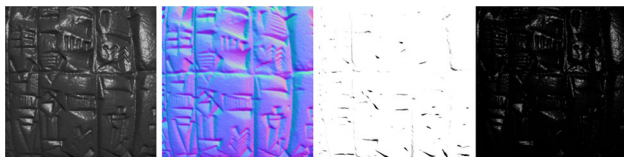


Fig. 5 Rendered multi-light image collections (a) come with associated normal maps (b) that can be used for Photometric Stereo algorithms evaluation, shadow maps (c), that can be used for shadow detection testing and specularity maps (d)

The MultiMaterial subset is created assigning different tints and material properties to different regions of the three geometric models. We created nine material combinations from eight base diffuse/specular/metallic behaviors (see Table 2) and 16 tints. The use of multiple materials may create problems in algorithms using global stats of the sampled reflectance to create mappings for dimensionality reduction.

SynthRTI can be exploited not only to test RTI relighting approaches, but also to evaluate Photometric Stereo methods, as the rendering engine can output normal maps, as well as shadow and specularity maps for each image (see Fig. 5). Ground truth normal maps and shadow maps are publicly released with the SynthRTI distribution that can be accessed at the following url: <https://github.com/Univr-RTI/SynthRTI>.

4.2 RealRTI

To test our approach and other methods on real images, we also created a dataset made of real RTI acquisitions, made with devices and protocols typically used in the cultural heritage domain. This dataset is composed of 12 multi-light

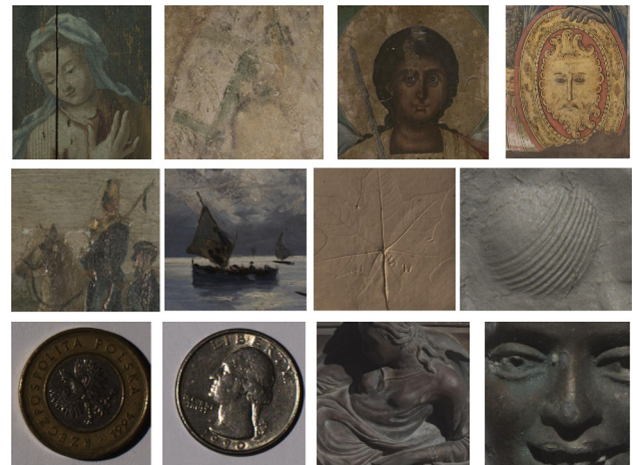


Fig. 6 Sample images of the RealRTI dataset representing the 12 different surfaces captured

image collections (cropped and resized to allow a fast processing/evaluation) acquired with light domes or handheld RTI protocols [14] on surfaces with different shape and material complexity. The items imaged are: (1) a wooden painted door (handheld acquisition, 60 light directions), (2) a fresco (dome acquisition, 47 lights), (3,4) two painted icons (handheld 63 and 72 lights), (5,6) two paintings on canvas (handheld, 49 lights and dome, 48 lights), impressions on plaster of a leaf (7) and a shell (8) (light dome, 48 lights), (9,10) two coins (both with light dome, 48 lights), and (11,12) two metallic statues (dome, 48 lights and handheld, 54 lights). This set allows testing relighting on different materials (matte, specular, metallic) and on shapes with different geometric complexities.

The RealRTI dataset can be accessed at the following url: <https://github.com/Univr-RTI/RealRTI>.

5 Results and evaluation

On the novel SynthRTI and RealRTI datasets, we performed several tests to evaluate the advantages of the proposed neural

encoding and relighting approach with respect to standard RTI encoding.

5.1 Evaluation methodology

We followed two different protocols for the evaluation of the relighting quality on SynthRTI and RealRTI. On the synthetic data, we created the relightable images with the different methods using the Dome subsets and tested the similarity of the images of the Test subset with the images relighted with the corresponding light directions. The similarity has been measured with the peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM).

On the RealRTI, as test data are missing, we followed a leave-one-out validation protocol: For each of the image collections, we selected five test images with different light elevations. Each of them has then, in turn, removed from the collection used to create relightable images and we evaluated the similarity of the removed images with the images relighted from the corresponding light direction. We finally averaged the similarity scores of the five tests.

5.2 Compared encodings

We compared the novel neural encoding with the RGB RTI encodings obtained with second-order polynomial texture maps (PTM), second- and third-order Hemispherical Harmonics RTI (HSH), and PCA-compressed RBF interpolation [17]. In PTM encoding, six fitting coefficients per color channel are obtained as floating-point numbers and compressed to byte size mapping the range between global max and min values into the 8 bits range. The total encoding size is therefore 18 bytes per pixel. In the second-order HSH, we similarly have an encoding with nine coefficients per color channel and a total of 27 bytes per pixel. In the third-order HSH, we have 16 coefficients per color channel and a total of 48 bytes per pixel. We implemented the encoding and relighting methods using MATLAB. Note that we could further compress the PTM/HSH encoding by using the LRGB version, using a single chromaticity for all the light direction. This choice, however, makes not possible to recover the correct tint of metallic reflections. In the PCA-RBF encoding, it is possible to choose the number of PCA components used for the input data projection. We decided to test the results obtained with 9 and 27 components. We used the original “Relight” code provided by the authors of [17] to create the encoding and relight images. In the neural encoding, as in PTM/HSH, we quantize coefficients to 8 bits integers with the min/max mapping so that they can be stored into image planes. This will allow easy integration of the novel format into existing viewers.

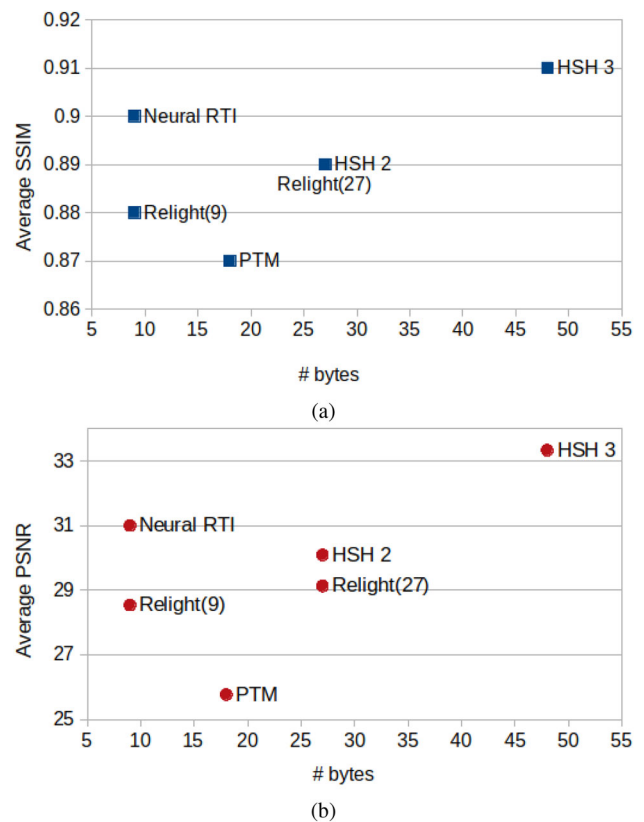


Fig. 7 Plots representing average SSIM (a) and PSNR (b) scores of the relighted images obtained with different methods from the test light directions of the SingleMaterial set. Scores are plotted versus the number of bytes used for the compressed encoding. Best methods should stay on top left

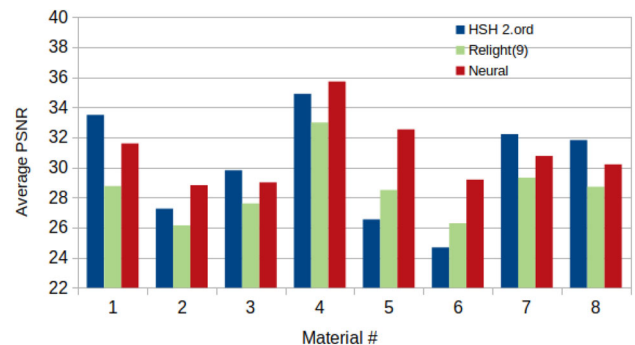


Fig. 8 Bar chart showing average SSIM of selected methods on the different materials. Neural relighting and second-order HSH perform similarly on rough materials, while neural relight is significantly better on glossy materials (5 and 6, see Table 1)

5.3 SynthRTI

If we look at the SSIM and PSNR values obtained on average on the three shapes with the single materials, we see that the neural relightable images are quite good, as the average scores estimated on all the dataset (eight materials times three shapes) are quite high despite the low number of parameters

Table 3 Average PSNR of the methods on the different shapes (all SingleMaterial sets). Parentheses include the number of coding bytes

	PTM(18)	HSH(27)	HSH(48)	[17](9)	[17](27)	Neur.(9)
Canvas	29.03	35.42	41.24	34.20	34.26	37.50
Tablet	23.79	27.63	29.92	25.87	26.80	28.06
Bas-relief	24.47	27.22	28.82	25.55	26.31	27.43
Avg	25.77	30.09	33.33	28.54	29.13	31.00

used by the coding (9). Figure 7 shows a plot representing the scores as a function of the number of encoding bytes and NeuralRTI stays on the top left of the plot. Scores of third-order HSH are higher, but the method requires 48 bytes for the encoding and, as we will see, it creates more artifacts on challenging materials. Other methods to compress the relightable image encoding are not as effective as NeuralRTI. This is demonstrated by the poorer results of the second-order HSH encoding that can be considered a compressed version of the third-order one and by the poorer results of the PCA-based compression of the full original information featuring 147 bytes per pixel (Relight).

Figure 8 shows the SSIM values for the different materials obtained by low-dimensional encoding methods as PCA/RBF (nine coefficients), neural (9), and HSH second order (27). Average SSIM values are close for HSH and Neural, but NeuralRTI performs better on challenging materials simulating dark plastic behavior and polished metals (5 and 6), despite the three times more compact code. Differences are statistically significant ($p < 0.01$ in a T -Test). HSH seems to handle significantly better the material with subsurface scattering effect on, probably thanks to color separation. Algorithms ranking does not depend on the geometric complexity of the model. Table 3 shows average PSNR values for the different models.

A visual comparison clarifies much better the advantages of NeuralRTI.

Figure 9 shows relighting performed with second-order HSH (a), third-order HSH (b), PCA/RBF with 9 or 27 coefficients (c,d), NeuralRTI (e) on the canvas object with assigned white “plastic” (material 2) behavior. (e) shows the ground truth image corresponding to the input direction. Only the neural relighting (e) is able to reproduce the specular highlight with a reasonable accuracy.

Figure 10 shows the relighting results obtained with different methods on the bas-relief shape with a metallic material (6) assigned, compared with the corresponding ground truth test image (f). It is possible to see that the highlights and shadows provided by the novel technique are the most similar to the real ones. HSH and PCA/RBF encodings with a limited number of parameters (a) and (c) appear matte. Adding more coefficients (b,d), the quality is better, but the contrast and the quality of the detail are not as good as in the result obtained with NeuralRTI (e), despite the heavier encoding. Like the

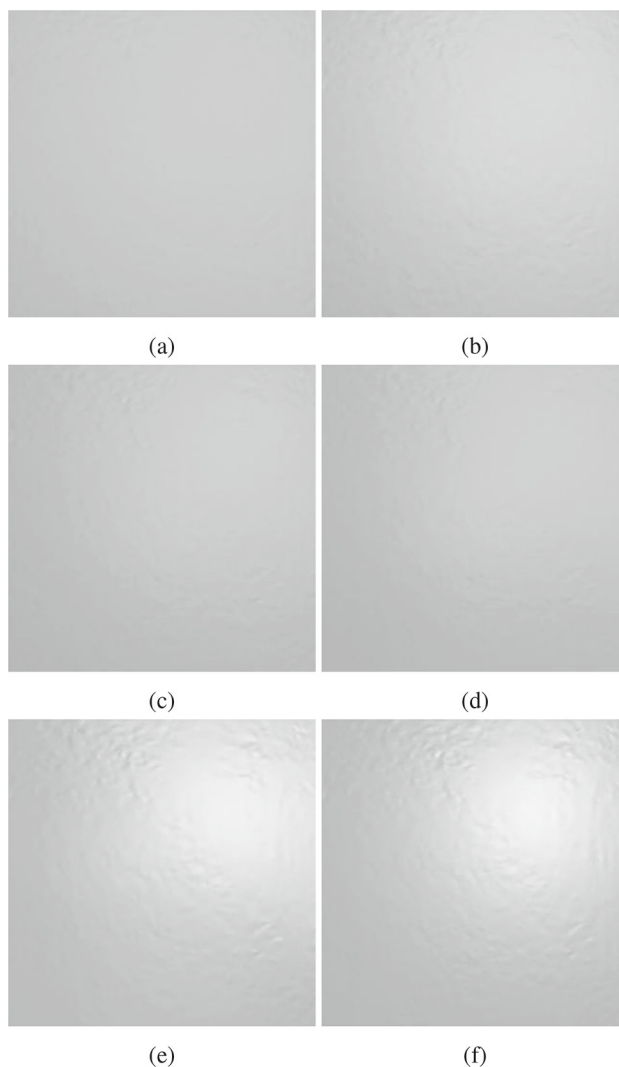


Fig. 9 Comparison of relighted images from simulated RTI data of a painted canvas, “plastic” material, light elevation 80 degrees. **a** second-order HSH. **b** Third-order HSH **c** PCA/RBF [17], nine coefficients. **d** PCA/RBF [17], 27 coefficients **e** NeuralRTI. **f** Ground truth image

other methods, NeuralRTI fails in reproducing correctly all the cast shadows (see, for example, the one on the right of the woman on the top right). This is expected, as the method is local, but the material appearance is definitely realistic on this challenging material.

Plotting the average relighting quality scores obtained on the MultiMaterial subset (Fig. 11), we see that Neural-

Table 4 Average PSNR of the methods on the different shapes (all MultiMaterial sets). Parentheses include the number of coding bytes

	PTM	HSH(27)	HSH(48)	[17](9)	[17](27)	Neur.(9)
Canvas	25.17	28.45	30.03	27.95	28.45	33.26
Tablet	20.56	22.76	24.24	20.89	22.28	23.74
Bas-relief	22.34	23.81	25.10	21.54	22.47	25.17
Average	22.69	25.01	26.46	23.46	24.40	27.39

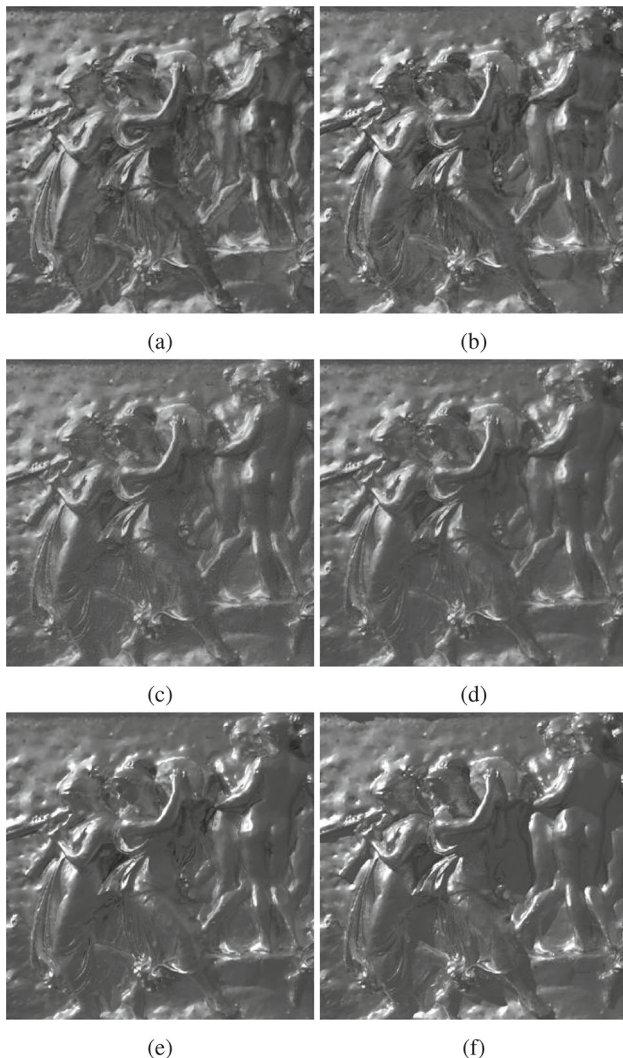


Fig. 10 Comparison of relighted images from simulated RTI data of a shiny (metal smooth) bas-relief. **a** Relighted with second-order HSH fitting (27 coefficients). **b** second-order HSH (46 coefficients). **c** Relighted with PCA/RBF (Relight, [17]), nine coefficients. **d** [17], 27 coefficients. **e** relighted with NeuralRTI (nine coefficients). **f** Ground truth image

RTI compares favorably with the other methods, as, despite the compact encoding, it demonstrates the highest relighting quality. This is because many of the multi-material combinations include metallic and glossy materials that are not rendered properly by the other methods. This fact appears clear looking at the bar chart in Fig. 12. The NeuralRTI

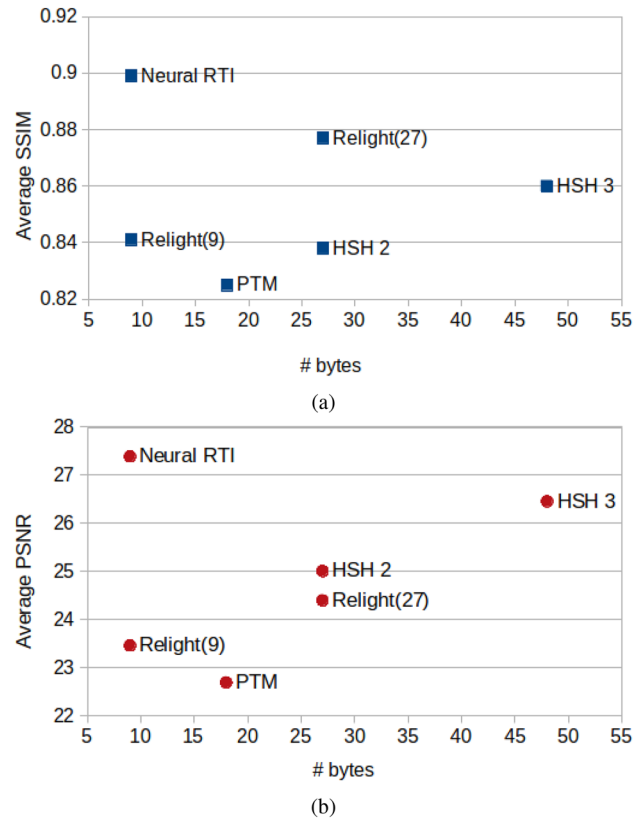


Fig. 11 Plots representing average SSIM (a) and PSNR (b) scores of the relighted images obtained with different methods from the test light directions of the MultiMaterial set. Scores are plotted versus the number of bytes used for the compressed encoding

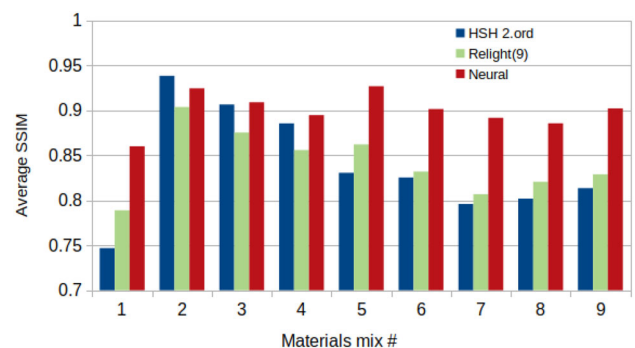


Fig. 12 Bar chart showing average SSIM of selected methods on the different materials. Neural relighting and second-order HSH perform quite similarly on rough materials, while neural relight is significantly better on glossy materials (5 and 6, see Table 1)

relighting has significantly higher SSIM values ($p < 0.01$ in T -tests) for all the material combinations except 2, 3, and 4. 2 is fully matte, 3 is all simulating rough plastic, and 4 is mostly composed of rough materials. Average PSNR values for the different geometric models are reported in Table 4.

A visual analysis of the results demonstrates here the advantages of NeuralRTI as well. Figure 13 shows relighted images obtained from the encodings made on the dome MLIC of material mix 5. Here, we have two tints and, on the left, plastic behavior with four levels of roughness and, on the right, metallic behavior with four levels of roughness (increasing from left to right). The second-order HSH fails in reproducing the correct highlights (a), compared with the ground truth ones (f). The third-order HSH better represents specularity (b), but still with artifacts, and requires 48 bytes per pixel. PCA/RBF-relighted image also appears matte with 9 bytes encoding (c) and is only slightly improved using 27 coefficients (d). NeuralRTI provides a good relighting of all the image regions, creating an image that is quite similar to the reference one in the test set. The directional light, in this example, comes here from an elevation of 60 degrees. In general, the ranking of the relighting quality is not changed by the input light elevation. Quality scores are all higher for higher elevation values.

5.4 RealRTI

Tests on real images confirm the evidence coming from those performed on the synthetic ones (and this also shows that the rendered materials have reasonably realistic behaviors). Looking at the average SSIM and PSNR plots obtained on the whole RealRTI dataset, it is possible to see that NeuralRTI provides the most accurate relighting even using only nine coefficients. The difference in the average score is mainly due to the different quality of the relighting of shiny metallic objects. The bar chart in Fig. 15 reveals that significant differences are found in items 9 and 10, which are metallic coins.

This fact can be seen by visually comparing relighted images with different methods. Figure 16 shows an image relighted with the leave-one-out procedure from a metallic coin acquisition. Relighting obtained with PCA/RBF with nine coefficients shows a wrong tint in the central part, limited highlights, and halos, the use of 27 coefficients removes the tint issue. Image relighted with HSH shows a correct tint, thanks to RGB decoupling, but relevant halos and missing highlights, especially in the second-order version. The NeuralRTI result is the only one with realistic highlights. Furthermore, it is possible to see that the proposed technique also avoids the typical artifact arising in the shadowed regions of RTI relighting, which appear as a blending of the shadows of the input images when obtained using PTM, HSH, and RBF.

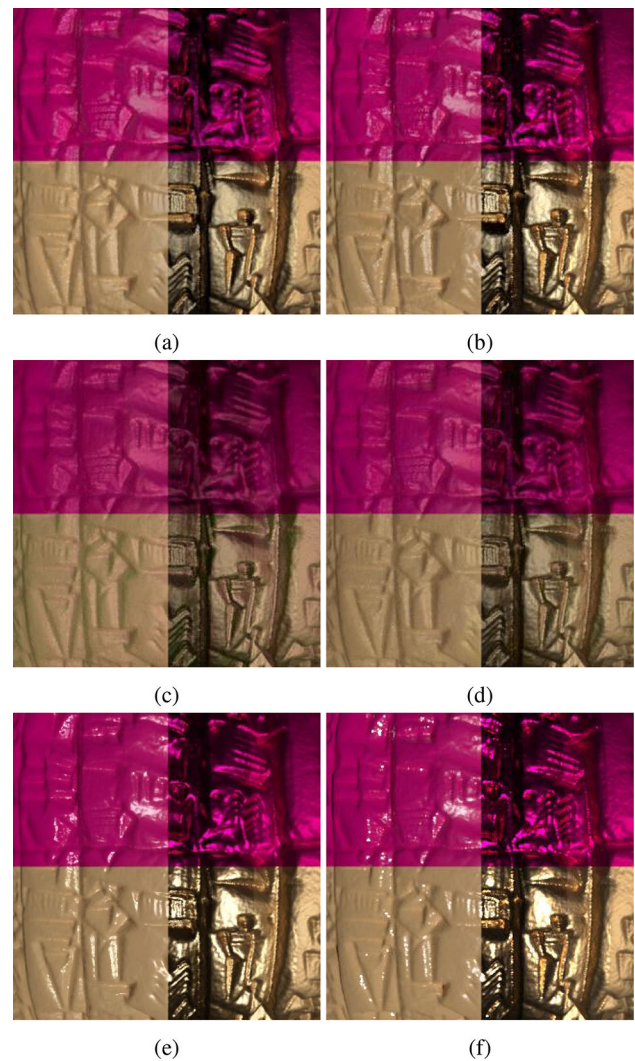


Fig. 13 Comparison of images relighted from the encodings created with the MultiMaterial 5 dome collection and the corresponding ground truth reference. **a** Second-order HSH. **b** Third-order HSH. **c** PCA/RBF [17], nine coefficients. **d** PCA/RBF [17], 48 coefficients. **e** NeuralRTI. **f** Ground truth image

5.5 Comparisons with other network architectures

As we pointed out in Sect. 2, there are no available frameworks for the neural network-based compressed encoding of relightable images, but there are actually methods for multi-light data relighting based on different network architectures, like those described [18,23]. We tested the use of these architectures, adapted for our scope and compared the relighting accuracy with the one provided by our method but did not include the results in the previous sections as it would not be fair to compare encodings not specifically designed and used for this kind of sampling. We show our results here just to confirm that our NeuralRTI encoding seems particularly suitable for the task also considering different network

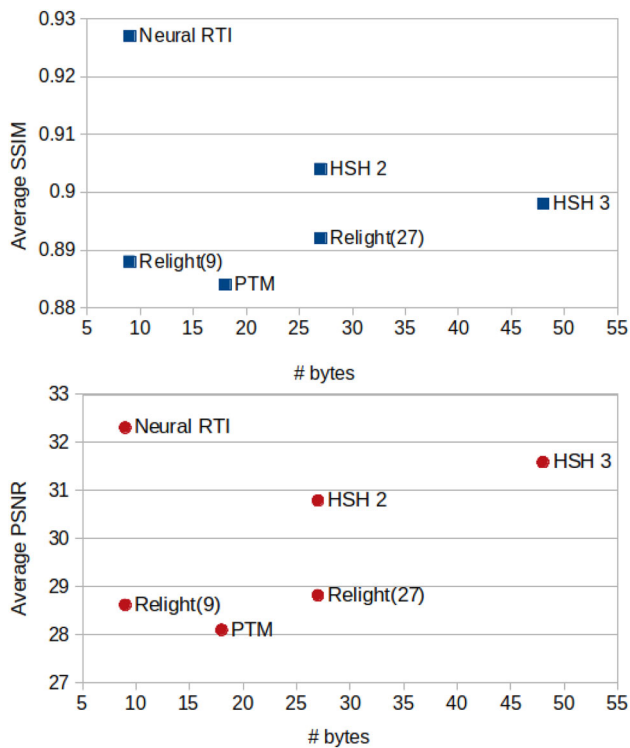


Fig. 14 Average quality scores of relighted images on the RealRTI dataset (all 12 objects), as a function of number of bytes used for the encoding

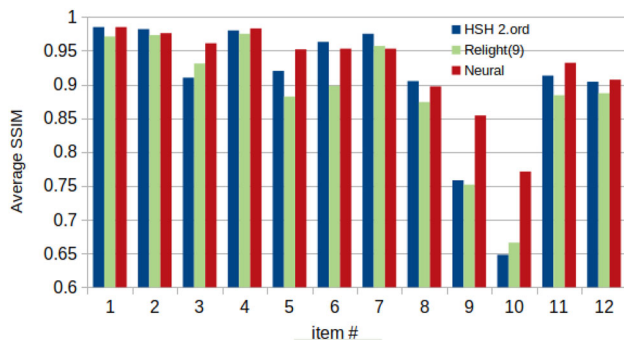


Fig. 15 Bar chart showing average SSIM of selected methods on the different objects. It is possible to see that the great improvement in the NeuralRTI relight is obtained for the shiny metallic objects (coins, items 9,10)

architectures. In [23], relighting is performed on five selected images, corresponding to those closer to an ideal sampling learned from a training set and to an encoding size of 5 bytes per pixel. This is clearly not the ideal way to compress a standard RTI dataset, but we still compared the relighting quality. Relighting is then global, based on the function learned on the training set. This can result in global illumination effects and realistic, even if hallucinated shadows.

The method in [18] is designed for a different application, e.g., compressing high-dimensional multi-light and multi-

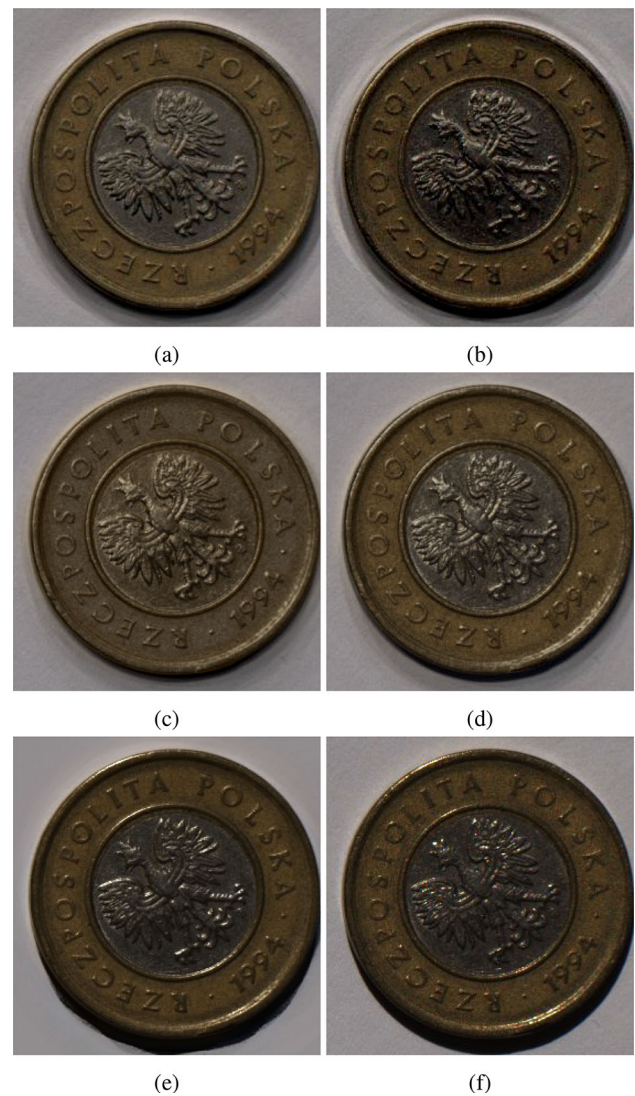


Fig. 16 Comparison of relighted images on a real challenging surface. **a** Second-order HSH. **b** Third-order HSH **c** PCA/RBF [17], nine coefficients. **d** PCA/RBF [17], 48 coefficients (e) NeuralRTI. **f** Ground truth image

view data of nearly planar patches. We instead focus on non-planar samples of potentially sharp BRDFs. The network of [18] uses 1D convolutions, while we avoid them and instead use more FC layers and different activation functions that work better with sharp-and-rotated BRDFs. We test an architecture similar to [18] in our case, with three 1D convolutional layers and a single fully connected layer, encoding with 9 bits and relighting similarly.

Looking at the results (see Table 5), it is possible to see that the adapted methods do not provide good results when compared with our method and other RTI encodings, showing that a specifically designed neural architecture is the right choice for the RTI relighting task.

Table 5 Adaptation of other network architectures not specifically designed for the relighting of typical RTI stacks resulted in poor results compared to our solution

		Average PSNR		
		Ours (9)	Adapted neural BTF (9)	Adapted deep relighting (15)
SynthRTI single		31.00	28.37	13.15
SynthRTI multi		27.39	23.49	12.46
Real		32.31	29.64	26.61
Average		30.23	27.16	17.41
		Average SSIM		
		Ours (9)	Adapted neural BTF (9)	Adapted deep relighting (15)
SynthRTI single		0.898	0.852	0.449
SynthRTI multi		0.899	0.808	0.436
Real		0.927	0.89	0.836
Average		0.908	0.850	0.574

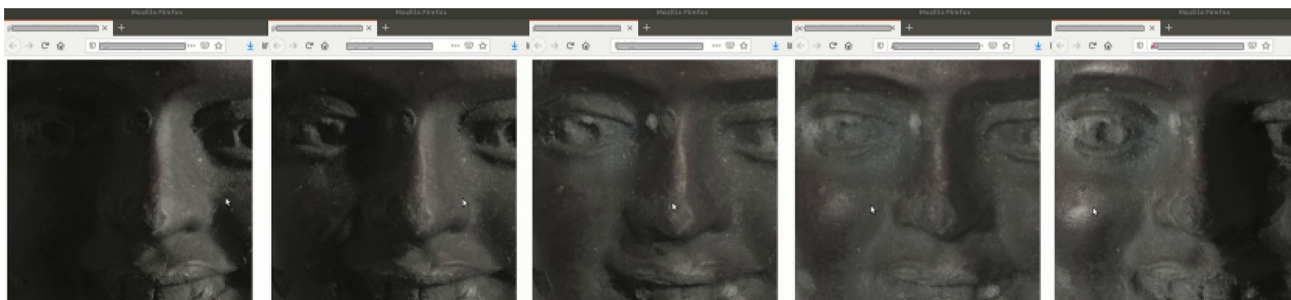


Fig. 17 Our web-based interactive relighting solution allows, like similar applications working with PTM/HSH/PCA-RBF encoding. Moving the cursor over the image, the user controls the light direction for the

novel illumination. In this case, moving from right to left, he can gradually move from raking light from right, to illumination from top, to raking light from left

5.6 Interactive relighting

We tested the NeuralRTI encoding in an existing web-based solution for the interactive relighting RTI framework. The current encoding results in a relighting that is far more complex than the simple weighted sum of the RTI/HSH coefficients and involves a number of multiplications that is proportional to the squared decoder layer size; however, we kept the number and the size of the decoder layers sufficiently small to allow interactive relighting.

To implement the web viewer, we used the tensorflow.js library [21]. The nine coefficients per pixel, quantized to 8 bits, are loaded in the browser as a binary type array and mapped back to the original floating-point coefficients range. The light direction parameter, varying with the mouse position, is concatenated with the coefficients and processed in the decoder network. Tensorflow.js library adopts a WebGL backend for the network processing and the data managed as textures in the GPU. The result is then rendered in a canvas element. Similarly, to PTM, HSH and PCA web viewers approach the coefficients can be combined and compressed in three JPEG or PNG images and split into a multiresolution pyramid of tiles.

Figure 17 shows some snapshots from an interactive surface inspection performed with the web viewer on our encoding.

Example videos captured in real time are included in the supplementary material, and a gallery of interactively relightable web-based visualizations of a selection of SynthRTI and RealRTI items can be seen at the project Webpage <https://univr-rti.github.io/NeuralRTI/>.

6 Discussion

In this paper, we provide two relevant contributions in the domain of multi-light image collection processing. The first is NeuralRTI, a neural network-based relighting tool that can provide better results than the current state-of-the-art methods, with reduced storage space. The quality of the images created with our novel technique is particularly good, especially on surfaces with metallic and specular behavior that are not well handled by the existing methods. Furthermore, the relighted images are less affected by the blended shadows artifacts typical of RTI.

NeuralRTI can be used immediately as an alternative to the classic PTM/HSH/PCA-RBF files and could be directly integrated into the existing visualization frameworks, as it is efficient enough to support real-time relighting. We demonstrate this with a specific web viewer based on tensorflow.js.

We believe that this tool will be particularly useful in the domain of cultural heritage and material science surface analysis, where RTI processing is widely employed. We will publicly release the encoding and relighting codes.

The main limitation of the proposed technique is that it is a “local” method, not learning global effects like cast shadows. This limitation, however, holds similarly for the currently used techniques, and it must be stressed that NeuralRTI provides better highlights and shadow simulations, avoiding blending artifacts. This is probably due to the ability to constrain nonlinearly the space of the reflectance patterns.

We plan, as future work, to investigate how our method developed to handle the typical sampling of RTI acquisition behaves with varying light directions sampling density, possibly specializing the codes for different acquisitions protocols. Our compressed encoding can be also used as a basis for further surface analysis, e.g., material and shape characterization. We plan to test methods to recover effectively normals and BRDF parameters from the compact encoding of the captured objects.

The second contribution is a synthetic dataset (SynthRTI) of images simulating light dome capture of surfaces with different geometric complexities and made of materials with different, realistic scattering properties. This dataset not only can be used to assess the quality of relighting methods on specific materials and shapes, but also to test Photometric Stereo algorithms (normal maps are obtained in the rendering process) and other tasks (e.g., shadow segmentation, highlights detection, etc), thanks to the information recorded in the multi-pass rendering. We plan to exploit it for a comprehensive evaluation of Photometric Stereo algorithms in future work. The dataset will be publicly released together with an archive of real calibrated MLICs (RealRTI) representing surfaces with different reflectance properties.

Acknowledgements Open access funding provided by Università degli Studi di Verona within the CRUI-CARE Agreement. This work was supported by the DSURF (PRIN 2015) project funded by the Italian Ministry of University and Research and by the MIUR Excellence Departments 2018-2022.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the

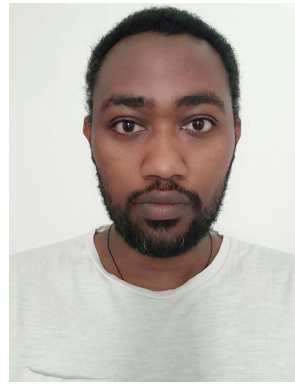
permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Bengio, Y., et al.: Learning deep architectures for ai. *Found. Trends® Mach. Learn.* **2**(1), 1–127 (2009)
2. Chen, G., Han, K., Shi, B., Matsushita, Y., and Wong, K.-Y.K.: Self-calibrating deep photometric stereo networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8739–8747, (2019)
3. CHI. Cultural heritage imaging website, 2019. [Online; accessed March-2019]
4. Earl, G., Basford, P., Bischoff, A., Bowman, A., Crowther, C., Dahl, J., Hodgson, M., Isaksen, L., Kotoula, E., Martinez, K., Pagi, H., and Piquette, K.E.: Reflectance transformation imaging systems for ancient documentary artefacts. In *Proc. International Conference on Electronic Visualisation and the Arts*, pp. 147–154, Swindon, UK, 2011. BCS Learning & Development Ltd
5. Giachetti, A., Ciortan, I.M., Daffara, C., Marchioro, G., Pintus, R., Gobbetti, E.: A novel framework for highlight reflectance transformation imaging. *Comput. Vis. Image Underst.* **168**, 118–131 (2018)
6. Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. *Science* **313**(5786), 504–507 (2006)
7. Kingma, D.P. and Ba, J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, (2014)
8. Lore, K.G., Akintayo, A., Sarkar, S.: Llnet: a deep autoencoder approach to natural low-light image enhancement. *Pattern Recognit.* **61**, 650–662 (2017)
9. Malzbender, T., Gelb, D., and Wolters, H.: Polynomial texture maps. In *Proceedings SIGGRAPH*, pp. 519–528, (2001)
10. Mudge, M., Malzbender, T., Chalmers, A., Scopigno, R., Davis, J., Wang, O., Gunawardane, P., Ashley, M., Doerr, M., Proenca, A., and Barbosa, J.: Image-based empirical information acquisition, scientific reliability, and long-term digital preservation for the natural sciences and cultural heritage. In *Eurographics (Tutorials)*, (2008)
11. Mudge, M., Malzbender, T., Schroer, C., and Lum, M.: New reflection transformation imaging methods for rock art and multiple-viewpoint display. In *Proceedings VAST*, vol. 6, pp. 195–202, (2006)
12. Palma, G., Corsini, M., Cignoni, P., Scopigno, R., Mudge, M.: Dynamic shading enhancement for reflectance transformation imaging. *J. Comput. Cult. Herit.* **3**(2), 1–20 (2010)
13. Pintus, R., Dulecha, T., Jaspe, A., Giachetti, A., Ciortan, I.M., and Gobbetti, E.: Objective and subjective evaluation of virtual relighting from reflectance transformation imaging data. In *GCH*, pp. 87–96, (2018)
14. Pintus, R., Dulecha, T.G., Ciortan, I., Gobbetti, E., and Giachetti, A.: State-of-the-art in multi-light image collections for surface visualization and analysis. In *Computer Graphics Forum*, vol. 38, pp. 909–934. Wiley Online Library, (2019)
15. Pitard, G., LeGoïc, G., Favrelière, H., Samper, S., Desage, S.-F., and Pillet, M.: Discrete modal decomposition for surface appearance modelling and rendering. In *Optical Measurement Systems for Industrial Inspection IX*, vol. 9525, pp. 952523:1–952523:10. International Society for Optics and Photonics, (2015)
16. Pitard, G., LeGoïc, G., Mansouri, A., Favrelière, H., Desage, S.-F., Samper, S., Pillet, M.: Discrete modal decomposition: a new approach for the reflectance modeling and rendering of real surfaces. *Mach. Vis. Appl.* **28**(5–6), 607–621 (2017)

17. Ponchio, F., Corsini, M., and Scopigno, R.: A compact representation of relightable images for the web. In *Proceedings ACM Web3D*, pp. 1:1–1:10, (2018)
18. Rainer, G., Jakob, W., Ghosh, A., and Weyrich, T.: Neural btf compression and interpolation. In *Computer Graphics Forum*, vol. 38, pp. 235–244. Wiley Online Library, (2019)
19. Ren, P., Dong, Y., Lin, S., Tong, X., and Guo, B.: Image based relighting using neural networks. *ACM TOG*, 34(4):111:1–111:12, (2015)
20. Shi, B., Wu, Z., Mo, Z., Duan, D., Yeung, S.-K., and Tan, P.: A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. In *Proceedings CVPR*, pp. 3707–3716, (2016)
21. Smilkov, D., Thorat, N., Assogba, Y., Yuan, A., Kreeger, N., Yu, P., Zhang, K., Cai, S., Nielsen, E., Soergel, D., Bileschi, S., Terry, M., Nicholson, C., Gupta, S.N., Sirajuddin, S., Sculley, D., Monga, R., Corrado, G., Viégas, F.B., and Wattenberg, M.: Tensorflow.js: Machine learning for the web and beyond. *CoRR*, abs/1901.05350, (2019)
22. Xie, J., Xu, L., and Chen, E.: Image denoising and inpainting with deep neural networks. In F.Pereira, C.J.C. Burges, L.Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pp. 341–349. Curran Associates, Inc., (2012)
23. Xu, Z., Sunkavalli, K., Hadap, S., and Ramamoorthi, R.: Deep image-based relighting from optimal sparse samples. *ACM TOG*, 37(4):126:1–126:13, (2018)
24. Zhang, W., Hansen, M.F., Smith, M., Smith, L., Grieve, B.: Photometric stereo for three-dimensional leaf venation extraction. *Comput. Ind.* **98**, 56–67 (2018)
25. Zheng, B., Sun, R., Tian, X., Chen, Y.: S-net: a scalable convolutional neural network for jpeg compression artifact reduction. *J. Electr. Imaging* **27**(4), 043037 (2018)
26. Zini, S., Bianco, S., and Schettini, R.: Deep residual autoencoder for quality independent jpeg restoration. *arXiv preprint arXiv:1903.06117* (2019)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Tinsae G. Dulecha received a BSc degree in computer science from Hawassa University, Ethiopia, in 2007, he then worked at Ethio telecom for 5 years till he joined Ca'Foscari University (October 2012) where he received his MSc in Computer Science in October 2014. December 2014, he won a 1-year research fellow to work on a project entitled “Augmented reality framework for museums” at IUAV University of Venice, Department of culture. In March 2015 he joined Hawassa University, Institute of Technology, School of informatics, and served as a lecturer. October 2016, he won a 1-year research fellow to work on a project entitled “Automated inventory of a cooler” at Ca'Foscari University, European Center for Living Technology(ECLT). Since October 2017 he is a Ph.D. student of Verona University working on material appearance analysis within the DSurf project under the supervision of Prof. Andrea Giachetti. His research interests are in the areas of Measurement, analysis, and modeling of material appearance, Machine learning, and computervision.



Filippo A. Fanni received both the BSc. and MSc. degree in Computer Science from the University of Cagliari in 2017 and 2019, respectively. Currently, he is a Ph.D. student at the University of Verona under the supervision of prof. Andrea Giachetti. His research interests include computer graphics, geometry processing, computational design, and digital fabrication.



Federico Ponchio is a Researcher at the ISTI Institute of the National Research Council of Italy. He received his Ph.D. in computer graphics at the Technical University of Clausthal DE. His research interest include multiresolution, geometry compression, huge models visualization, image processing, and their web applications.



Fabio Pellacini is a Full Professor of Computer Science at Sapienza University of Rome where he leads the Computational Design Lab. His research focuses on algorithms for intuitively designing complex 3D environments with an emphasis on creative and entertainment industries and personal fabrication. Pellacini received his Laurea degree in Physics from the University of Parma (Italy), and his MSc. and Ph.D. in Computer Science from Cornell University. Pellacini received an NSF CAREER

award and an Alfred P. Sloan fellowship for his research contributions. Before Sapienza, he was a faculty member at Dartmouth College and worked for Pixar Animation Studios on lighting algorithms, where his work received credits on three featurefilms.



Andrea Giachetti is an associate professor of Computer Science at the University of Verona. His research focuses on various aspects of Visual Computing such as image processing, 3D modeling and shape analysis, interaction in virtual reality. He received a Ph.D. degree in Physics from the University of Genova, Italy, in 1993. Before Verona, he worked with CRS4 (Center for Advanced Studies, Research, and Development, Sardinia), where he has been head of the Medical Image Processing of CRS4, and key staff of various EU funded projects and with the University of Cagliari.