Interactive Relighting Tools for Virtual Avatars, Digital Doubles, and Telepresence

Fajar Pratama[1] and Ayu Rahmani[2]
[1]Garuda Institute of Technology, Department of Computer Science and Intelligent Systems, Jalan Malioboro 78, Yogyakarta 55212, Indonesia
[2]Citra Nusantara University, Faculty of Computing and Digital Media, Jalan Asia Afrika 59, Bandung 40111, Indonesia

ABSTRACT Virtual avatars, digital doubles, and telepresence systems increasingly mediate professional collaboration, social interaction, and entertainment, yet the visual plausibility of these digital humans remains strongly constrained by lighting. Illumination modulates appearance, reveals or suppresses geometric detail, conveys mood, and communicates subtle social cues. Traditional offline rendering pipelines can synthesize photorealistic relighting of captured or modeled humans, but they are not suitable for interactive scenarios where lighting must respond to user input, dynamic environments, or real-time communication. This paper examines interactive relighting tools for controllable virtual identities, with a focus on their application to real-time avatars, high-fidelity digital doubles, and telepresence experiences. The discussion spans physical and mathematical models of light transport, data-driven relighting with neural representations, system-level design of low-latency pipelines, and perceptual aspects of lighting control interfaces. A unified perspective is developed in which avatar appearance is expressed as a low-dimensional, controllable mapping from scene illumination and viewing configuration to rendered imagery, and in which interaction is understood as exploration of this appearance manifold under tight computational constraints. The paper analyzes trade-offs between physical accuracy, temporal stability, and latency, and it considers the role of capture, learning, and prior knowledge in building robust relighting models. Special attention is given to the design of user-facing tools that support intuitive yet precise lighting control for communication, performance, and content creation. The paper concludes with open challenges related to generalization, cross-device consistency, and the integration of relighting with broader telepresence and mixed reality ecosystems.

KEYWORDS

1. Introduction

Relighting describes the process of changing the illumination of a scene or subject while preserving underlying geometry, reflectance, and identity [1]. For virtual avatars, digital doubles, and telepresence, relighting is not only an offline post-production tool but an interactive capability that must operate at frame rates compatible with real-time communication. Interactive relighting tools allow users, systems, or autonomous agents to adjust light direction, color, intensity, and spatial distribution in response to context, intent, or scene understanding. In practice this capability has to coexist with diverse constraints, including limited compute budgets on mobile devices, heterogeneous capture hardware, compressed network streams, and strongly varying environmental conditions.

The rapid adoption of video conferencing, virtual production stages, and mixed reality has created an ecosystem in which digital representations of people are routinely embedded into synthetic or hybrid environments. Virtual avatars range from stylized characters in social platforms to photorealistic digital doubles used in film and live events. Telepresence systems seek to reproduce remote participants with sufficient fidelity and responsiveness that participants can maintain mutual gaze, perceive subtle nonverbal cues, and feel co-present. Lighting is central to these goals because it shapes perceived material, surface detail, and depth, and because it influences the reading of facial expressions and eye contact [2]. At the same time, illumination provides a powerful design handle: by manipulating lighting, a system can harmonize an avatar with a virtual background, adapt to cultural or organizational lighting practices, or emphasize important communicative signals [3].

Interactive relighting tools must reconcile several tensions [4]. Physically based models of light transport suggest high-dimensional integrals over geometry, materials, and illumination, while real-time constraints demand compact representations and efficient numerical approximations. High-fidelity telepresence requires faithful reproduction of an individual's appearance under diverse lighting, yet interactive tools must expose a manageable number of intuitive controls. Production pipelines for digital doubles often rely on complex capture stages that are poorly matched to consumer hardware, whereas telepresence applications must function with commodity sensors and noisy data. Neural rendering approaches offer powerful new abstractions for relighting, but they introduce their own trade-offs in terms of training data requirements, generalization, and predictability of failure modes.

In this paper, interactive relighting is treated as a controlled mapping from illumination parameters and scene or avatar descriptors to rendered output, subject to latency and quality constraints imposed by interactive use. The analysis focuses on three interrelated use cases. Virtual avatars, which may be stylized or realistic, must be lit in a manner that is consistent with their environment and performance context while often operating under hardware acceleration in game engines. Digital doubles, constructed with high-fidelity geometry and reflectance data, are typically associated with more controlled production environments and stricter visual expectations. Telepresence systems prioritize responsiveness and robustness, often using partial geometry or learned representations derived from live camera feeds [5]. A unified discussion across these regimes clarifies shared mathematical structures and system challenges.

The remainder of the paper examines the foundations and design of interactive relighting tools, from physical image formation models and linear approximations to neural representations and telepresence-oriented system architectures. Emphasis is placed on the interplay between mathematical structure, numerical methods, and human interaction design, with the aim of understanding how to build tools that are both technically sound and practically usable in real-time settings.

2. Background on Relighting and Telepresence

Interactive relighting builds on a long history of research in computer graphics, vision, and image-based modeling. At its core lies the image formation process, in which scene radiance is mapped to sensor irradiance and ultimately to pixel intensities. For avatars and digital doubles, the scene typically consists of a human head and body, clothing, and nearby props, represented with explicit geometry, volumetric structures, or learned implicit fields. Telepresence scenarios add the complexity of live acquisition, in which this representation must be inferred or updated continuously from camera streams, depth sensors, or other modalities.

Classical rendering pipelines treat relighting as a forward problem: given geometry, materials, and light sources, they simulate the transport of light through the scene. For offline digital doubles, physically based path tracing remains common, supporting complex reflectance, global illumination, and volumetric effects [6]. However, direct path tracing is computationally expensive for interactive applications, motivating approximations that reduce the dimensionality of the problem. Precomputed radiance transfer, spherical harmonic lighting, and basis decompositions of environment maps all illustrate the view that lighting can be represented compactly and that surface response can be expressed as a linear operator acting on this representation. These constructions are particularly appealing for avatars because they allow parts of the computation to be moved offline or amortized across frames.

In parallel, image-based relighting has emerged as a powerful alternative in which explicit light transport simulation is replaced, wholly or in part, by interpolation or reconstruction from captured examples. For digital doubles, specialized capture systems such as light stages acquire many images of a subject under varied illumination and sometimes varied viewpoint. The resulting reflectance field encapsulates how each point on the subject responds to changes in lighting. While traditional light stages assume controlled multi-source illumination, telepresence systems often operate with single-view or few-view setups and uncontrolled environment lighting. This mismatch motivates the development of compact reflectance and geometry models that can be inferred from limited data but still support realistic relighting.

Telepresence introduces additional constraints that shape the design of relighting tools. Remote participants often appear on heterogeneous devices, ranging from phones to large displays and head-mounted displays [7]. Network conditions impose variable bandwidth and latency, constraining the amount of lighting information that can be transmitted and the complexity of remote rendering. In some systems, relighting occurs entirely on the sender side, with the transmitted avatar already lit for a target environment. In others, a more neutral or canonical representation is sent, and relighting is performed on the receiver side to match local context. Both strategies require mechanisms to estimate, represent, and exchange illumination parameters, as well as mechanisms to keep remote participants visually coherent even when environments differ.

Another important aspect is the distinction between photorealism and stylization. Interactive social platforms may employ non-photorealistic avatars whose materials and shading are deliberately abstracted. Even in such cases, lighting remains crucial: stylized shading models must convey shape and expression reliably, and lighting changes can be used as expressive tools. For digital doubles used in film or live performance, expectations are different, and photorealistic relighting under varied conditions is often a primary goal. Telepresence systems may sit between these extremes, seeking a balance between fidelity, robustness, and computational cost [8]. Across these regimes, interactive relighting tools must provide consistent abstractions that allow artists, end-users, or automated systems to reason about and manipulate lighting.

Finally, there is a growing interest in cross-modal and cross-

Target	Typical representation	Primary objectives	Dominant constraints
Stylized virtual avatars	Low-poly mesh, toon shading, simplified materials	Expressive communication, recognizability, platform branding	Mobile-grade GPUs, stylistic coherence across worlds
Photorealistic digital doubles	High-resolution mesh, displacement maps, multi-layer skin BRDFs	Visual realism, continuity with live-action footage	Offline/near-real-time budgets, high capture cost, strict quality expectations
Telepresence participants	Parametric head/body models, point clouds, neural avatars	Low latency, robustness to noisy capture, gaze and expression fidelity	Commodity sensors, bandwidth limits, heterogeneous displays
Mixed reality human holograms	Volumetric grids or implicit fields embedded in real spaces	Consistent lighting with physical scene, spatial presence	On-device inference, spatial tracking errors, cross-device consistency

Table 1 Representative targets for interactive relighting, spanning stylized avatars, high-end digital doubles, and telepresence-oriented representations with differing objectives and system constraints.

Illumination representation	Mathematical form	Typical dimension	Usage in avatars / telepresence
Directional light set	Finite set of unit directions with RGB intensities	$3p$ for p lights	Art-directable rigs, key/fill/rim control, real-time engines
Low-order spherical harmonics (SH)	Coefficient vector $c \in \mathbb{R}^k$	$k \approx 9$–27	Diffuse environment lighting, smooth global shading
Environment map (cube / lat-long)	Discrete function $L\omega_i$ on sphere	$O10^3$–10^4 samples	Image-based lighting, background-aware telepresence
Mixture of spherical Gaussians	Weighted lobes on sphere with mean and covariance	$O6m$ for m lobes	Compact specular-capable environment approximation
Learned latent lighting code	Vector $\ell \in \mathbb{R}^d$ without explicit semantics	8–64	Neural relighting, joint optimization with appearance

Table 2 Common illumination parameterizations, trading off compactness, physical interpretability, and suitability for real-time avatar relighting.

device consistency of lighting. A participant might appear as a talking-head video stream on one device, as a 3D avatar in a virtual world on another, and as a holographic-like projection in an extended reality environment elsewhere. Ideally, the perceived lighting characteristics of that participant would stay coherent across these manifestations, even when the underlying rendering pipelines differ. Achieving this coherence requires shared representations of illumination and appearance, as well as robust estimation and control algorithms that can operate across diverse acquisition and rendering stacks. Interactive relighting tools form a key component in this broader effort to synchronize visual appearance in distributed, heterogeneous telepresence ecosystems.

3. Physical and Mathematical Models of Relighting

Relighting rests on the physics of light transport and its mathematical abstractions. A standard starting point is the rendering equation, which relates outgoing radiance at a point to incident radiance, surface reflectance, and emitted light. For a point on a surface with position vector [9]

$$x \in \mathbb{R}^3$$

and unit normal

$$n \in \mathbb{R}^3,$$

the outgoing radiance in direction

$$\omega_o \in \mathbb{S}^2$$

can be written conceptually as an integral over incident directions

$$\omega_i \in \mathbb{S}^2.$$

Model class	Core mapping	Strengths	Typical limitations
Linear diffuse model	$I\ell = a\ B\ell$	Fast evaluation, analytic inversion, SH-friendly	Limited to near-Lambertian behavior, weak speculars
Precomputed radiance transfer (PRT)	$s = Ac$ with transfer matrix A	Handles soft shadows, reusable across frames	Precomputation cost, less flexible under topology changes
Analytic microfacet shading	Discrete sum over light samples	Physically meaningful parameters, engine integration	Complex to calibrate from sparse telepresence data
Fully neural renderer	$f_\theta : \ell, z, v \mapsto I$	Expressive, captures complex view dependence	Data-hungry, opaque failure modes, control entanglement
Hybrid analytic + neural residuals	Analytic base plus learned correction term	Better interpretability, improved realism per FLOP	Requires careful training to avoid artifacts in real time

Table 3 Comparison of representative relighting models, from linear diffuse approximations to fully neural and hybrid approaches.

Capture regime	Sensors	Recovered quantities	Relighting affordances
Light stage session	Dozens of cameras, programmable LED sphere	Dense reflectance field, high-res geometry, albedo, normals	High-fidelity digital doubles, basis images for relighting
Studio multi-view rig	Several RGB(D) cameras with calibrated lights	Pose-tracked meshes, medium-quality materials	Live performance capture with moderate relighting
Commodity desktop / laptop	Single RGB webcam, optional depth sensor	Sparse normals, coarse geometry, estimated environment map	Everyday telepresence with robust but approximate relighting
Mobile mixed reality device	RGB camera, depth, IMU, spatial mapping	Local light probes, environment structure	Scene-aware avatar insertion with shared lighting

Table 4 Representative capture configurations and the relighting capabilities they enable for avatars and telepresence participants.

The integral couples the bidirectional reflectance distribution function and the incident illumination on the hemisphere oriented around the surface normal. In full generality this integral is high-dimensional and expensive to evaluate directly for each point and direction.

For many avatar and telepresence applications, simplifications are introduced to obtain tractable models. A common approximation is to treat skin and many clothing materials as predominantly diffuse at the scale of interest, at least for the purposes of coarse relighting. In the diffuse or Lambertian case, outgoing radiance depends on the cosine of the angle between the normal and the incident direction and on a scalar albedo coefficient. If the incident illumination is represented as a function [10]

$$L\omega_i,$$

the diffuse shading term at a surface point may be expressed as a spherical convolution. When illumination is represented in a low-dimensional basis, such as low-order spherical harmonics, this convolution reduces to a pointwise product in coefficient space. One may write, in abstract form,

$$s = A\,c,$$

where

$$c \in \mathbb{R}^k$$

denotes illumination coefficients in the chosen basis,

$$A \in \mathbb{R}^{m\times k}$$

denotes a precomputed transfer matrix encoding per-vertex or per-texel response, and

$$s \in \mathbb{R}^m$$

denotes the resulting shading values over the avatar surface. This linear structure underlies many interactive relighting tech-

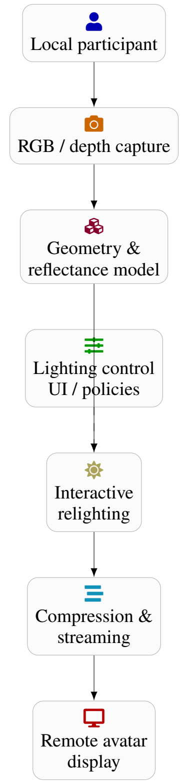

Figure 1 High-level interactive relighting pipeline for telepresence: live capture feeds an avatar model, which is driven by lighting controls and rendered in real time before being encoded and streamed to remote displays.

niques, since it allows shading to be recomputed rapidly as the coefficients change.

In more general cases, material behavior is anisotropic, specular, or spatially varying. Microfacet models approximate specular reflection by averaging contributions from many microscopic surface facets with a distribution of orientations [11]. The reflectance function then depends on both the half-vector between view and light direction and the roughness and index of refraction parameters. While the underlying integrals remain complex, practical implementations often discretize illumination into a finite set of sampled directions or prefiltered environment representations. One can write discrete shading at a point as a finite sum

$$I = \sum_{j=1}^{p} w_j \, r_j,$$

where the weights

$$w_j$$

capture illumination samples and visibility factors, and the terms

$$r_j$$

encode per-sample reflectance response. In matrix form, this becomes

$$I = u^{\top} v,$$

with

$$u, v \in \mathbb{R}^p.$$

Such low-dimensional representations naturally lend themselves to linear algebraic optimization and precomputation.

Interactive relighting tools often rely on the observation that the mapping from illumination to pixel intensities can be approximated as a linear or mildly nonlinear operator [12]. For a set of image pixels indexed by

$$i = 1, \ldots, N,$$

and an illumination parameter vector

$$\ell \in \mathbb{R}^d,$$

a linear model takes the form

$$I_i \ell = a_i \; b_i^{\top} \ell,$$

where

$$a_i \in \mathbb{R}$$

and

$$b_i \in \mathbb{R}^d$$

are coefficients derived from geometry and materials. In matrix notation, stacking the pixel intensities into a vector

$$I \ell \in \mathbb{R}^N$$

yields

$$I \ell = a \; B \ell,$$

with

$$B \in \mathbb{R}^{N \times d}.$$

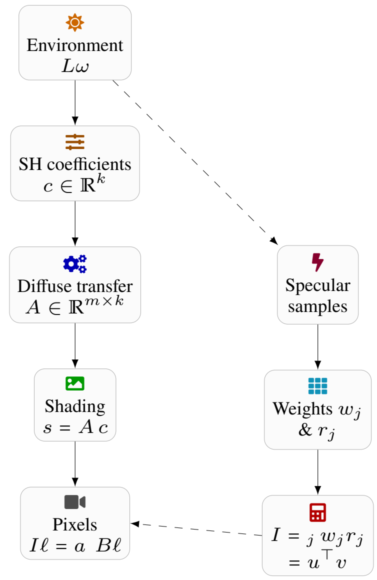

Figure 2 Physical and linear formulations of avatar relighting: diffuse shading is expressed as a matrix–vector product between precomputed transfer and illumination coefficients, while specular components are approximated by finite sums that can be written as low-dimensional inner products.

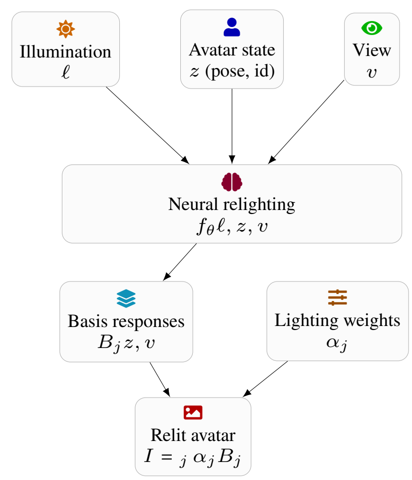

Figure 3 Data-driven relighting model: a neural function takes illumination, avatar state, and view as input, predicts basis responses, and combines them with lighting coefficients to produce relit images while preserving controllable lighting parameters.

This formulation is convenient for numerical analysis, because the map from [13]

$$\ell$$

to

$$I\ell$$

is affine and supports efficient evaluation on parallel hardware. It also clarifies how relighting can be posed as an inverse problem: given a set of observed images under different illuminations, one may solve for the underlying transfer matrix and albedo parameters using least-squares methods and regularization.

From a multivariate calculus perspective, interactive relighting benefits from differentiable formulations of the rendering process. Suppose the appearance model is parameterized by a vector

$$\theta \in \mathbb{R}^q,$$

which may encode geometry, material parameters, or learned latent codes. The rendered image

$$I\ell, \theta$$

depends on both illumination and these parameters [14]. Differentiable rendering techniques compute gradients

$$\nabla_\theta I\ell, \theta$$

and sometimes

$$\nabla_\ell I\ell, \theta,$$

enabling gradient-based optimization or learning. For example, during calibration of a digital double, one may minimize an energy functional

$$E\theta = {}_\Omega \left(I\ell, \theta - I_{\text{obs}}\right)^2 \, d\Omega,$$

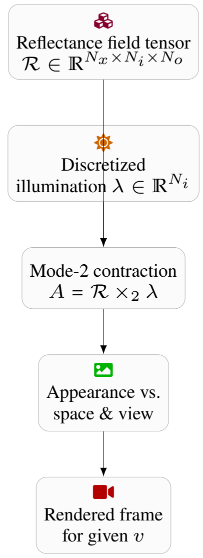

Figure 4 Tensor view of relighting: the reflectance field is treated as a high-order tensor that is contracted with a discretized illumination vector to yield spatial and view-dependent appearance, which is then sampled to form rendered frames.

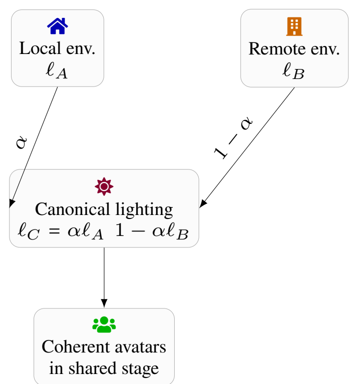

Figure 5 Coordination of lighting descriptors in telepresence: local and remote environment estimates are blended into a canonical illumination used to relight all participants, promoting visually coherent avatars despite differing physical spaces.

over an image domain

$$\Omega, 15$$

using gradient descent. Numerical optimization schemes such as Gauss–Newton or quasi-Newton methods exploit the local linear structure of the problem to accelerate convergence, although they must be adapted to respect real-time constraints when used in an interactive loop.

Tensor formulations provide additional structure for avatars and telepresence settings. The reflectance field of a subject can be viewed as a tensor

$$\mathcal{R}$$

whose indices correspond to spatial location, incident direction, and viewing direction. For a discrete representation with

$$N_x$$

spatial samples,

$$N_i$$

incident directions, and

$$N_o$$

outgoing directions, one may think of [16]

$$\mathcal{R} \in \mathbb{R}^{N_x \times N_i \times N_o}.$$

Relighting corresponds to contracting this tensor along the incident-direction mode with an illumination vector. If

$$\lambda \in \mathbb{R}^{N_i}$$

denotes a discretized illumination distribution, the resulting appearance tensor along spatial and viewing dimensions can be written abstractly as

$$A = \mathcal{R} \times_2 \lambda,$$

where

$$\times_2$$

denotes mode-2 tensor–vector multiplication. Low-rank tensor approximations such as canonical polyadic or Tucker decompositions reduce the storage and evaluation cost by expressing

$$\mathcal{R}$$

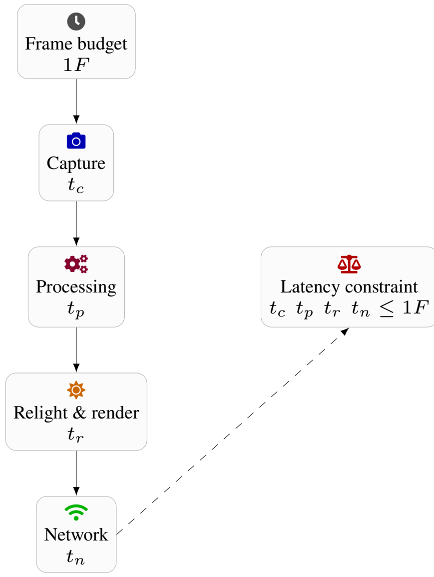

Figure 6 Latency model for interactive relighting: capture, processing, rendering, and networking must collectively fit within the frame budget to support real-time telepresence with responsive lighting control.

as a sum of separable factors. In practice, such decompositions correspond to learning a small set of basis images or basis lighting patterns and exploiting linear combinations to reconstruct new configurations.

Probability and statistics enter in several ways. Illumination in real environments is uncertain and time-varying, and measurements are noisy. One may treat illumination parameters [17]

$$\ell$$

as random variables with a prior distribution

$$p\ell,$$

and avatar appearance as a conditional distribution

$$pI \mid \ell, \theta.$$

Bayesian formulations seek posterior distributions

$$p\theta \mid I_{\text{obs}}$$

over appearance parameters given observed images. In simplified settings, Gaussian assumptions produce linear Gaussian models in which posterior means and covariances can be computed in closed form using standard multivariate formulas. Even when full Bayesian inference is not employed, statistical reasoning is useful for regularization, for estimating confidence in relighting operations, and for designing robust estimators that mitigate the impact of outliers and sensor artifacts.

These mathematical abstractions provide the foundation on which more complex neural and data-driven relighting techniques are constructed [18]. They guide the choice of parameterizations, suggest numerically stable algorithms, and inform the design of user interfaces that expose a limited number of meaningful controls within an underlying high-dimensional appearance space.

4. Neural and Data-Driven Relighting for Avatars

Neural rendering has introduced new ways of representing and manipulating appearance under changing illumination. Rather than explicitly modeling geometry and reflectance with analytic functions, neural approaches learn a mapping from inputs such as pose, viewing direction, and lighting parameters to output images or feature maps. For avatars and digital doubles, such models can encode complex phenomena including view-dependent specular highlights, subsurface scattering, and soft self-shadowing, while still supporting interactive relighting under certain conditions.

A generic neural relighting model can be abstractly described as a function

$$f_\theta : \ell, z, v \mapsto I,$$

where

$$\ell$$

denotes illumination parameters,

$$z$$

denotes a representation of the avatar state, such as pose, expression, or latent identity codes,

$$v$$

denotes viewpoint parameters, and [19]

$$I$$

denotes the rendered image or image features. The function

$$f_\theta$$

is represented by a neural network with parameters

$$\theta.$$

Pipeline stage	Time budget symbol	Example value at F = 60 fps	Typical operations
Capture	t_c	4–6 ms	Sensor readout, demosaicing, basic denoising
Processing / reconstruction	t_p	2–5 ms	Segmentation, normals/albedo estimation, pose tracking
Relighting / rendering	t_r	3–6 ms	Shading, neural inference, compositing, post-processing
Network transmission	t_n	2–8 ms (variable)	Encoding, transport, decoding across endpoints
Total per frame	$t_c\ t_p\ t_r\ t_n$	$\leq$ 16.7 ms	Must satisfy latency constraint $\leq 1F$

Table 5 Illustrative latency budget for an interactive relighting pipeline targeting 60 fps, showing how capture, processing, rendering, and networking share the per-frame time constraint.

Source of uncertainty	Mathematical view	Impact on appearance	Mitigation strategy
Noisy illumination estimates	Random illumination vector $\ell \sim p\ell$	Flickering brightness, unstable shading	Temporal filtering of ℓ_t, Bayesian smoothing
Imperfect geometry / normals	Uncertain shape parameters θ	Distorted highlights, incorrect self-shadowing	Shape priors, multi-view fusion, learned corrections
Sensor noise / compression artifacts	Observation noise in I_{obs}	Speckling, blockiness, unstable fine detail	Robust losses, denoising, confidence-weighted updates
Domain shift across devices	Mismatch between training and deployment distributions	Bias in relit skin tone, material mismatch	Device-specific calibration, adaptation of $p\theta$

Table 6 Examples of uncertainty sources in avatar relighting, their effect on perceived appearance, and common strategies for handling them within statistical and data-driven frameworks.

During training, the parameters are optimized to minimize a loss functional

$$\mathcal{L}\theta = \mathbb{E}_{\ell,z,v}\left[\Phi\big(f_\theta\ell, z, v, I_{\text{target}}\big)\right],$$

where

$$\Phi$$

measures discrepancy between predicted and target images or features [20]. Common choices for

$$\Phi$$

include per-pixel squared error, robust norms, and perceptually motivated feature-based distances. In practice, the expectation is approximated with empirical averages over dataset samples.

For digital doubles acquired in controlled capture setups, the training data can span a wide range of lighting and viewing configurations. This enables the network to learn detailed relighting behavior that approximates the underlying reflectance field.

Neural architectures vary widely and may include convolutional encoders and decoders, attention mechanisms, and volumetric or implicit representations. In implicit models, the network is queried at continuous spatial coordinates and directions, akin to neural radiance fields, allowing rendering by volumetric integration along rays. In more mesh-based schemes, neural networks augment traditional shading by predicting correction terms or residuals that refine analytic models.

Interactive relighting imposes constraints on these architectures [21]. Evaluation of

$$f_\theta$$

must be efficient enough to maintain target frame rates, and the mapping from intuitive lighting controls to the internal parameterization must be stable. One strategy is to use the neural model to precompute basis responses. For example, the network can be trained to predict appearance under a fixed set of canonical basis lights. At runtime, the user's chosen lighting configuration is decomposed into the same basis, and the final

Control dimension	User-facing metaphor	Internal parameterization	Perceptual priorities
Primary light direction	Orbital gizmo around face, "sun position" slider	Rotation of SH/environment map, key light azimuth/elevation	Stable shadow direction, eye highlight placement
Global brightness / contrast	Single "brightness" or "punch" control	Scale and offset on illumination coefficients ℓ	Avoid clipping, maintain facial detail
Color temperature	Warm–cool slider	Low-dimensional color transform on light RGBs	Natural skin tones, white balance consistency
Relighting style preset	Named modes (e.g., "office", "studio", "dramatic")	Preset bundles of ℓ and shading parameters	Predictable mood changes, temporal coherence
Autoadapt to background	Toggle for automatic matching	Optimization over ℓ from background cues	Reduce mismatch between avatar and scene lighting

Table 7 Illustrative mapping between intuitive lighting controls presented to users and the underlying illumination parameterizations manipulated by interactive relighting systems.

image is obtained by linearly combining the basis responses. If the network's output for each basis light is denoted

$$B_j z, v,$$

and the coefficients of the current lighting in that basis are

$$\alpha_j,$$

then the rendered image becomes

$$I = \sum_j \alpha_j \, B_j z, v.$$

This decomposition recovers the linear structure exploited in classical precomputed radiance transfer, but with the basis response learned directly from data rather than computed from explicit transport simulation [22].

Another line of work uses neural networks to infer physically meaningful parameters such as environment maps, surface albedo, and normal fields from monocular observations. The estimated parameters can then feed into a conventional real-time renderer that handles relighting analytically. Formally, one can consider an inverse module

$$g_\phi : I_{\text{obs}} \mapsto \hat{\ell}, \hat{\theta},$$

and a forward renderer

$$R : \ell, \theta \mapsto I.$$

The composite map

$$R \circ g_\phi$$

serves as a relighting system: it infers latent illumination and appearance parameters from input images and synthesizes output images under modified lighting. Training can proceed either in a supervised manner, if ground truth parameters are available, or in a self-supervised manner by enforcing photometric consistency between observed and rendered images under varying known or estimated lighting.

From a probabilistic standpoint, neural relighting models can be interpreted as learning conditional distributions [23]

$$pI \mid \ell, z, v,$$

potentially with explicit uncertainty estimates. Generative models such as variational autoencoders or diffusion-based approaches can sample multiple plausible relit images, accounting for ambiguities in the underlying geometry or reflectance. While such generative diversity is less common in telepresence applications that demand stability and consistency, uncertainty estimates can nonetheless inform adaptive algorithms that adjust reliance on learned components based on confidence.

For telepresence, the interaction between neural relighting and streaming constraints is critical. When a remote participant is represented as a neural avatar, the receiving device may hold a copy of

$$f_\theta$$

and locally control

$$\ell$$

based on the user's environment or preferences. Alternatively, relighting may be performed server-side, with only compressed images transmitted [24]. Both options raise questions about how to parameterize and synchronize illumination. For example, if the sender transmits coefficients

$$\ell$$

in a spherical harmonic basis, the receiver can apply these coefficients to local objects or other avatars, achieving coherent cross-object lighting. Conversely, if the receiver estimates its

Challenge area	Underlying cause	Affected scenarios	Possible research directions
Cross-device appearance consistency	Different cameras, displays, and rendering stacks	Participants joining from phones, desktops, XR headsets	Shared illumination spaces, cross-modal calibration, perceptual metrics
Generalization across users and environments	Limited training coverage of identities and lighting	Scalable telepresence with global user base	Meta-learning, domain adaptation, richer priors over θ and ℓ
Temporal stability vs. responsiveness	Noisy sensing vs. need for quick feedback	Rapidly changing environments, mobile use	Adaptive filters on ℓ_t, predictive models, user-tunable smoothing
Ethics, privacy, and fairness in relighting	Appearance normalization and editing	Virtual meetings, content creation platforms	Transparent controls, bias audits, user agency over relighting policies

Table 8 Selected open challenges for interactive relighting of avatars, digital doubles, and telepresence participants, emphasizing generalization, consistency, dynamics, and societal considerations.

own illumination from camera or sensor data, it must map this estimate into the parameterization expected by the avatar's relighting model.

Neural and data-driven approaches also influence how interactive tools for artists and users are designed. Because the mapping from parameters to images is learned, it may exhibit nonlinearities or entanglements that complicate control. To address this, some systems introduce intermediate interpretable controls that are linearly related to user input but nonlinearly related to network internal states. Others constrain training to align certain latent directions with semantically meaningful changes in lighting, such as azimuthal rotation of the primary key light or adjustment of overall contrast. Mathematically, this can be understood as shaping the Jacobian [25]

$$\frac{\partial I}{\partial \ell}$$

so that changes in

$$\ell$$

correspond to predictable and smooth variations in appearance.

Overall, neural and data-driven relighting techniques extend the expressive range of interactive systems, particularly for complex materials and capture conditions where analytic models struggle. At the same time, they inherit the underlying physical and linear structures that have long guided relighting research, and their successful deployment in avatars, digital doubles, and telepresence depends on careful integration with those structures.

5. System Design, Interaction, and Telepresence Integration

Interactive relighting tools must be realized as complete systems that integrate capture, computation, rendering, and user interaction. For avatars and digital doubles, such systems typically consist of a capture path that acquires geometry and reflectance information, a processing path that builds or updates an appearance model, and a rendering path that responds to interactive lighting controls in real time. Telepresence scenarios add live networking and synchronization, turning the system into a distributed pipeline.

At capture time, the system may receive monocular or multi-view RGB streams, depth maps, infrared data, or other sensor modalities. For high-end digital doubles, dedicated capture sessions can obtain dense multi-view imagery under controlled lighting, enabling construction of detailed meshes and reflectance maps [26]. For consumer telepresence, capture is often limited to a single commodity camera and possibly a depth sensor. In both cases, the captured data must be processed to extract a representation suitable for subsequent relighting. This representation might be a polygonal mesh with texture maps, a point cloud, a volumetric grid, an implicit neural field, or a hybrid of these structures. The processing stage may involve geometric reconstruction, segmentation of the subject from the background, estimation of per-pixel normals and albedo, and fitting of anatomical priors such as parametric head and body models.

The rendering path implements the actual relighting operation. A common architecture separates global illumination descriptors from local shading evaluation. Global illumination descriptors encode the lighting environment in a compact form, such as a small set of directional lights, a low-resolution environment map, a spherical harmonic coefficient vector, or a mixture of spherical Gaussians. Local shading evaluation then uses precomputed transfer data or a neural model to map these descriptors to per-vertex or per-pixel color. The use of compact descriptors is essential for real-time performance, since it allows the same descriptors to be reused across many surface points and potentially across multiple avatars in a scene.

The interactive control interface sits atop this rendering path [27]. For expert users, such as lighting artists in virtual production, the interface may expose parameters closely aligned with physical lighting concepts, including key and fill lights, color temperature, and physical units of intensity. For general users in telepresence applications, controls are usually more abstract or goal-oriented, such as presets that adjust overall brightness and contrast, or context-aware modes that attempt to match background lighting automatically. In both cases, the system must translate user input into illumination descriptors and update the rendering in a temporally coherent manner. Sudden changes in lighting parameters can produce visual discontinuities or flickering, particularly when combined with noise from live capture.

From a numerical standpoint, the system faces latency constraints that can be described in terms of frame time budgets. If the target frame rate is

$$F$$

frames per second, the total time available to acquire, process, relight, and display each frame is approximately

$$\frac{1}{F}$$

seconds. Let

$$t_c, t_p, t_r, t_n$$

denote the times for capture, processing, rendering, and network transmission respectively [28]. A basic requirement is

$$t_c \ t_p \ t_r \ t_n \leq \frac{1}{F}.$$

Meeting this inequality often requires exploiting parallelism, pipelining, and approximate algorithms. For example, processing and rendering for frame

$$k$$

can overlap with capture for frame

$$k \ 1,$$

and neural inference can be accelerated with specialized hardware and model compression.

Telepresence integration poses additional challenges in synchronizing lighting across participants and scenes. Suppose a remote participant is represented as an avatar whose appearance depends on a local lighting descriptor

$$\ell_A,$$

while the local user's environment is described by

$$\ell_B.$$

A telepresence system may wish to harmonize these descriptors so that both participants appear under comparable lighting, even though they occupy different physical spaces [29]. One strategy is to map both

$$\ell_A$$

and

$$\ell_B$$

into a shared canonical illumination space, perhaps by normalizing for overall intensity and color balance, and then choose a compromise descriptor

$$\ell_C$$

that balances local context and mutual visibility. The mapping from

$$\ell_A, \ell_B$$

to

$$\ell_C$$

can be formulated as an optimization problem that trades off local fidelity and cross-participant consistency according to a simple energy function, for instance

$$E\ell_C = \alpha \left\|\ell_C - \ell_A\right\|^2 \ 1 - \alpha \left\|\ell_C - \ell_B\right\|^2,$$

with

$$\alpha \in 0, 1.$$

The minimizer is a convex combination [30]

$$\ell_C = \alpha \, \ell_A \ 1 - \alpha \, \ell_B,$$

illustrating how linear illumination parameterizations facilitate such coordination.

User interaction with relighting tools is shaped by human perception of lighting. Observers are sensitive to inconsistencies in shadow direction, shading continuity, and eye highlights, particularly on faces. At the same time, many aspects of absolute intensity and color temperature can be adjusted without breaking plausibility, as long as shading patterns remain coherent. This asymmetry suggests interface designs that emphasize control over lighting direction and relative contrast while automating global exposure and white balance. For example, a system may estimate a local environment map from the user's camera feed and then allow the user to rotate this environment around the avatar, thereby changing apparent light direction while preserving relative incident distributions. Mathematically, rotation of an environment map expressed in spherical harmonics corresponds to multiplying the coefficient vector by a rotation matrix with block structure. Precomputing these block matrices makes it possible to update illumination descriptors interactively as the user rotates a virtual light rig [31].

In collaborative telepresence, relighting tools also affect social signaling. Adjusting lighting to emphasize a speaker or to align all participants with a shared virtual stage can aid turn-taking and attention management. Conversely, overly aggressive or inconsistent relighting can undermine trust by making participants' appearance unstable. System designers must therefore balance autonomy and control: automated relighting based on scene understanding can reduce cognitive load, but users often require manual overrides to preserve their preferred appearance. Logging and predictability of lighting adjustments are important for maintaining a sense of agency.

Integration with broader mixed reality ecosystems introduces further considerations. When avatars are embedded into spatially tracked environments, relighting tools may draw on spatial light probes, depth maps, and surface reconstructions to infer local illumination. The same descriptors can be shared across avatars and virtual objects, enabling consistent lighting across the entire scene. Temporal filtering and prediction algorithms can smooth lighting estimates over time, using simple dynamical models such as autoregressive filters [32]. For instance, a first-order filter updates the illumination descriptor via

$$\ell_t = \beta\,\ell_{t-1}\ 1 - \beta\,\tilde{\ell}_t,$$

where

$$\tilde{\ell}_t$$

is the raw estimate at time

$$t,$$

and

$$\beta \in 0, 1$$

controls the trade-off between responsiveness and stability. Such filtering reduces flicker in relit avatars when sensor estimates of lighting are noisy.

Overall, system design for interactive relighting requires coordination between numerical models, hardware capabilities, network constraints, and human factors. The tools must expose clear and interpretable controls while internally operating on mathematical representations of illumination and appearance that support efficient computation and robust telepresence integration.

6. Conclusion

Interactive relighting tools for virtual avatars, digital doubles, and telepresence sit at the intersection of physical modeling, numerical approximation, data-driven learning, and human–computer interaction [33]. Starting from the physics of light transport, the discussion has outlined how simplified models, such as linear mappings between illumination coefficients and shading, create a foundation for real-time computation. These models can be extended to tensor formulations of reflectance fields and framed within probabilistic perspectives that account for measurement uncertainty and variability in real-world illumination.

Neural and data-driven methods expand the expressive range of relighting, enabling high-fidelity reproduction of complex materials and subtle view-dependent effects even from limited capture setups. By treating relighting as a learned function of illumination, pose, and viewpoint, these methods can match or approximate the behavior of meticulously constructed digital doubles, while remaining compatible with consumer-level telepresence scenarios. Their integration with analytic representations allows hybrid systems that benefit from both physical interpretability and learned flexibility.

At the system level, interactive relighting tools must meet strict latency budgets and operate robustly in heterogeneous environments. They must map intuitive user controls and environmental measurements into internal illumination descriptors, coordinate these descriptors across distributed participants, and maintain temporal coherence in the presence of noise and network variability. When avatars are embedded into richer mixed reality scenes, relighting tools also become a mechanism for harmonizing real and virtual content, sharing lighting information across objects, and managing transitions between different devices and display modalities.

From a user perspective, lighting is both a technical constraint and an expressive resource [34]. Relighting tools influence how remote participants are perceived, how attention is guided, and how identity and mood are communicated. Effective designs therefore require not only accurate rendering but also careful consideration of perceptual sensitivities and social expectations. Interfaces that foreground control over direction, contrast, and shadow structure, while automating global adjustments, align well with these considerations. several challenges remain. Generalization of relighting models across individuals, environments, and devices continues to be an open problem, especially under the practical limitations of telepresence systems. Consistency of lighting across multiple representations of the same person, such as video, avatars, and holographic displays, requires further work on shared illumination parameterizations and cross-modal calibration. Privacy and fairness concerns arise when relighting modifies or normalizes appearance in ways that may inadvertently obscure identity cues or introduce bias. Addressing these issues requires combining advances in mathematical modeling and machine learning with careful evaluation in real-world usage contexts. Despite these challenges, the core mathematical and computational structures described here provide a coherent framework for reasoning about interactive relighting. By viewing avatars, digital doubles, and telepresence participants as entities embedded in a controllable appearance space shaped by illumination and viewing conditions, it becomes possible to design tools that support flexible, responsive, and perceptually grounded control of lighting in a wide variety of interactive settings [35].

Acknowledgments

We would like to thank the reviewers of this research for their helpful comments and suggestions.

References

[1] S. Angioni, N. Lincoln-DeCusatis, A. Ibba, and D. R. Recupero, "A transformers-based approach for fine and coarse-grained classification and generation of midi songs and soundtracks.," *PeerJ. Computer science*, vol. 9, pp. e1410–e1410, 6 2023.

[2] R. J. Haslam, ""this wonderful machine": How should we teach humanities texts like gulliver's travels in the time of chatgpt?," *Critical Humanities*, vol. 2, 1 2024.

[3] J. Lee, P. N. Duong, and H. Lee, "Configurable encryption and decryption architectures for ckks-based homomorphic encryption.," *Sensors (Basel, Switzerland)*, vol. 23, pp. 7389–7389, 8 2023.

[4] T. Mahalingam and M. Subramoniam, "Optimal object detection and tracking in occluded video using dnn and gravitational search algorithm," *Soft Computing*, vol. 24, pp. 18301–18320, 11 2020.

[5] A. Muhammad, Z. Salman, K. Lee, and D. Han, “Harnessing the power of diffusion models for plant disease image augmentation.,” *Frontiers in plant science*, vol. 14, pp. 1280496–, 11 2023.

[6] A. Revanur, N. I. Kolkin, D. Agarwal, S. Agrawal, H. Zhang, M. Harikumar, and E. Shechtman, “Image relighting using machine learning,” Oct. 2 2025. US Patent App. 18/949,023.

[7] Z. Zhou, X. Qian, J. Hu, G. Chen, C. Zhang, J. Zhu, and Y. Dai, “An artificial intelligence-assisted diagnosis modeling software (aims) platform based on medical images and machine learning: a development and validation study.,” *Quantitative imaging in medicine and surgery*, vol. 13, pp. 7504–7522, 7 2023.

[8] A. Generosi, T. Agostinelli, S. Ceccacci, and M. Mengoni, “A novel platform to enable the future human-centered factory,” *The International Journal of Advanced Manufacturing Technology*, vol. 122, pp. 4221–4233, 8 2022.

[9] A. Kohli and A. Gupta, “Detecting deepfake, faceswap and face2face facial forgeries using frequency cnn,” *Multimedia Tools and Applications*, vol. 80, pp. 18461–18478, 2 2021.

[10] W. Li, F. Viola, J. Starck, G. J. Brostow, and N. D. F. Campbell, “Roto++: accelerating professional rotoscoping using shape manifolds,” *ACM Transactions on Graphics*, vol. 35, pp. 62–15, 7 2016.

[11] R. Fiebrink, P. R. Cook, and D. Trueman, “Chi - human model evaluation in interactive supervised learning,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 147–156, ACM, 5 2011.

[12] null Tahmid Noor Rahman, null Nusaiba Khan, and null Zarif Ishmam Zaman, “Redefining computing: Rise of arm from consumer to cloud for energy efficiency,” *World Journal of Advanced Research and Reviews*, vol. 21, pp. 817–835, 1 2024.

[13] T. Morita, S. Akashiba, C. Nishimoto, N. Takahashi, R. Kukihara, M. Kuwayama, and T. Yamaguchi, “A practical teacher–robot collaboration lesson application based on printeps,” *The Review of Socionetwork Strategies*, vol. 12, pp. 97–126, 5 2018.

[14] A. T. Sonwane, “Face mask detection system,” *International Journal for Research in Applied Science and Engineering Technology*, vol. 9, pp. 639–642, 11 2021.

[15] H. Lee and Y.-J. Shin, “Video learning based image classification method for object recognition,” *Journal of Central South University*, vol. 20, pp. 2399–2406, 9 2013.

[16] H. T. Macedo, T. Almeida, L. N. Matos, and B. O. Prado, “Frequency maps as expert instructions to lessen data dependency on real-time traffic light recognition,” 8 2020.

[17] S. Waldner, E. Wendelspiess, P. Detampel, C. M. Schlepütz, J. Huwyler, and M. Puchkov, “Advanced analysis of disintegrating pharmaceutical compacts using deep learning-based segmentation of time-resolved micro-tomography images.,” *Heliyon*, vol. 10, pp. e26025–e26025, 2 2024.

[18] M. Handrik, J. Handriková, and L. Jakubovičová, “Use of the boinc system for distributed data collection in the "internet of things",” *MATEC Web of Conferences*, vol. 157, pp. 05009–, 3 2018.

[19] A. Nadeem, A. Jalal, and K. Kim, “Accurate physical activity recognition using multidimensional features and markov model for smart health fitness,” *Symmetry*, vol. 12, pp. 1766–, 10 2020.

[20] F. Binzagr, “Explainable ai-driven model for gastrointestinal cancer classification.,” *Frontiers in medicine*, vol. 11, pp. 1349373–, 4 2024.

[21] J. A. Omiye, H. Gui, R. Daneshjou, Z. R. Cai, and V. Muralidharan, “Principles, applications, and future of artificial intelligence in dermatology.,” *Frontiers in medicine*, vol. 10, pp. 1278232–, 10 2023.

[22] Y. Chen, J. Chi, T. Shen, B. You, Y. Wang, and C. Zhang, “A multi-granularity facial extreme makeup transfer and removal model with local-global collaboration,” *Applied Intelligence*, vol. 54, pp. 9741–9759, 7 2024.

[23] C. Herrera, R. Baker, J. Sheaves, and M. Sheaves, “A computer vision approach for studying fossorial and cryptic crabs,” 5 2020.

[24] K. A. Ritter and T. L. Chambers, “Educational gaming and use for explaining alternative energy technologies,” *International Journal for Innovation Education and Research*, vol. 2, pp. 30–42, 3 2014.

[25] Y. Yao, A. Mohan, E. Bliss-Moreau, K. Coleman, S. M. Freeman, C. J. Machado, J. Raper, J. Zimmermann, B. Y. Hayden, and H. S. Park, “Openmonkeychallenge: Dataset and benchmark challenges for pose tracking of non-human primates,” 9 2021.

[26] P. K. D. Pramanik, S. Pal, and P. Choudhury, “Mobile crowd computing: potential, architecture, requirements, challenges, and applications,” *The Journal of Supercomputing*, vol. 80, pp. 2223–2318, 7 2023.

[27] R. Shao and E. J. Delp, “Forensic scanner identification using machine learning,” 2 2020.

[28] S. Sudiyono, R. Rusiyanto, A. D. N. I. Musyono, A. Septiyanto, A. B. Anggoro, and A. Setiawan, *Development of E-Modules Based on Project-Based Learning on Tapper Turning Materials*, pp. 235–242. Atlantis Press SARL, 12 2022.

[29] J. Tang, R. Zhong, R. Zhang, and Y. Zhang, “Imuc: Edge–end–cloud integrated multi-unmanned system payload management and computing platform,” *Drones*, vol. 8, pp. 19–19, 1 2024.

[30] C. K. Mohan and B. Yegnanarayana, “Classification of sport videos using edge-based features and autoassociative neural network models,” *Signal, Image and Video Processing*, vol. 4, pp. 61–73, 12 2008.

[31] C. Hua, X. Cao, B. Liao, and S. Li, “Advances on intelligent algorithms for scientific computing: an overview.,” *Frontiers in neurorobotics*, vol. 17, pp. 1190977–, 4 2023.

[32] T. Pamuła and W. Pamula, “Detection of safe passage for trains at rail level crossings using deep learning.,” *Sensors (Basel, Switzerland)*, vol. 21, pp. 6281–, 9 2021.

[33] H. Cecotti, L. Huisinga, and L. G. Peláez, “Fully immersive learning with virtual reality for assessing students in art history,” *Virtual Reality*, vol. 28, 1 2024.

[34] D. Zhou, B. Pei, X. Li, D. Jiang, and L. Wen, “Innovative bim technology application in the construction management of highway.,” *Scientific reports*, vol. 14, pp. 15298–, 7 2024.

[35] G. Huang, “Retracted article: Intelligent image denoising based on optical imaging equipment in physical education teaching simulation,” *Optical and Quantum Electronics*, vol. 56, 1 2024.