https://doi.org/10.2352/EI.2024.36.14.CVAA-176

Light analysis of paintings at scale using spherical harmonics

Pepe Ballesteros Zapata[1], Darío Negueruela del Castillo[1], and Leonardo Impett[2]

[1]Institute for Digital Visual Studies, University of Zurich - Max Plank Society, Switzerland
[2]Cambridge Digital Humanities, University of Cambridge, United Kingdom

Abstract

Light is abstract. Light is difficult to describe with words. Unlike other pictorial motifs such as perspective, light depiction eludes fixed rules and defies straightforward measurement. This study presents a preliminary step towards measuring light information from paintings at scale. We use spherical harmonics to extract light environment features from depicted faces. Our initial evaluation assessed the model's ability to accurately represent environmental lighting, using natural images with ground truth lighting information. To evaluate performance on paintings on a large-scale setting, we compare our light direction estimates with human annotations over the course of five centuries. Finally, our methodology undergoes validation through an art historical case study. We track the use of diffuse light across seventeenth-century Netherlandish portrait painting to commensurate art historical literature with our quantitative measures.

Introduction

The use of light in painting is deeply embedded within its cultural and historical context, intricately linked to the painting techniques and materials of the time. Representational paintings depict scenes from our observable environment, where objects are arranged in a space, have colors, and illumination. It is in this context where light serves not only as an instrument to render the three-dimensional properties of the natural world, but also as an aesthetic tool for artistic expression. Among other pictorial motifs, light is used to suit the visual message by guiding the viewer's attention, evoke a particular mood, or arrange objects in space, while also carrying symbolic weight — as a reminder of the passage of time, portraying the liveness of emotions and intellectual faculty, or manifesting divine presence.

Unlike other pictorial features like perspective or human anatomy, light is challenging to measure. Moshe Barasch noted that Leonardo da Vinci expressed the impossibility to measure the correctness of lights and shadows by technical means, and continues: "the fact that shadows cannot be measured nor their 'correctness' ascertained by technical means makes their representation a definite manifestation of that artistic talent which, unlike 'rules', cannot be learned" [1, p. 48]. The detachment between light and fixed rules is tied to the complexities of describing light with language. As Hills observes [2, p. 4], light leaves room for interpretation, making accident and intention difficult to discern and describe. Barasch concludes: "We must consider light as a visual language, but not a textual language" [1].

The advent of computer vision methods to analyze paintings present a refreshing opportunity to study depicted light, enhancing existing techniques where human perception and language might encounter limitations. The quantification of light as a stylistic signature opens up avenues for detecting resonances between artists or schools who employ lighting in similar ways. This not only aids in appreciating aesthetic affinities but also entails the possibility of uncovering workshop practices, where consistent lighting might reveal an artist's practice of working directly from the subject, while contradictory lighting could suggest reliance on preparatory sketches, or the artist working directly from imagination [3]. Impett operationalized Lomazzo's theory of divine light using light source detection techniques, providing insights about the pictorial representation of divine sources in concrete paintings [4]. Researches in [5] quantified the extent to which artists employ lighting contrast in paintings using a Bayesian metric, addressing the problem of light quantification as a component of style. This method facilitates large-scale comparison of lighting contrast, yet it only takes into account this singular visual aspect of light. The adoption of spherical harmonics with the proposed method expands light analysis by capturing subtler nuances of the light environment, such as ambient and directional properties.

Figure 1. *On the left: A Young Roman Woman (detail). Sebastiano del Piombo (1512). On the right: 3D mesh of the face relighted with the extracted environment map.*

Stork proposed several methods for the analysis of light in paintings [3]. Spherical harmonics were used to analyze the lighting around the contour of an object and estimate the direction of illumination. This approach facilitates the inference of artist workshop practices by analyzing the illumination signature of different objects within the same painting [6]. In contrast, our methodology proposes a customized pipeline designed for light analysis at scale, enhanced by the integration of 3D information. This enables an increased number of spherical harmonic light coefficients, capturing more subtleties present in complex light settings. We use faces as light probes to extract the spherical harmonic coefficients of the light environment on a painting. Once the illumination is captured, it is possible to relight the 3D geometry of the face according that particular light environment, as showed in Figure 1.

The analysis of depicted light at scale presents a research opportunity yet to be defined by the digital art history community. We present here a preliminary work on the topic, proposing a computational pipeline for the systematic quantification of light features in a wide array of paintings. The paper begins by introducing the foundations of spherical harmonics, followed by the explanation of the proposed pipeline to extract three-dimensional light features from digital paintings. Subsequently, we introduce three distinct evaluation methods. First, we test the reliability of luminance values as a proxy for ground truth lighting by examining the capacity of the extracted coefficients to differentiate between light environments. Secondly, we conduct a large scale experiment to find the direction of illumination in paintings over five centuries, comparing our findings with light directions assessed by human evaluators. Finally, we propose a case study to test our method against art historical knowledge.

Methodology

Spherical Harmonics are a mathematical tool to represent any function on the surface of a sphere, such as the light distribution over a hemisphere. In computer graphics, they provide an efficient encoding of light used in real-time rendering applications [7]. Illumination conditions are described as a spherical function, which can be stored in a texture, or environment map. We use spherical harmonics to approximate the Bidirectional Reflectance Distribution Function (BRDF), which describes the reflection behaviour according to material properties of objects. The resulting BRDF describes a diffuse surface, which indicates that light is scattered uniformly in all directions. In the diffuse reflectance model, often referred as the Lambertian reflectance model, the calculation of incident light is only influenced by the angle of incident light relative to the surface normal. Visually, modeling the illumination of an object as a diffuse surface means that specular reflections —highlights that depend on the viewpoint— are not considered.

The intuition for the overall pipeline to extract lighting information from paintings comes from inverse rendering [8]. Contrary to a common computer graphics rendering pipeline, the task is to extract 3D information from the 2D images. While deep inverse rendering models rely on spherical harmonics as a latent representation of light features, we avoid this trend. Deep architectures require ground truth normal, depth, and shading or albedo maps. Such data is hardly accessible for natural images and impossible for paintings. Some self-supervised methods have been proposed for natural images [9], but the task remains an ill-posed problem and hence a more interpretable approach is favored. Our method sidesteps such issues by reducing the amount of required data to regress lighting information by assuming luminance values to be a good approximation for ground truth light intensities.

We chose to analyze lighting in faces for two reasons. First, it is the visual feature from which we most confidently can regress 3D data. This is an essential part of the pipeline as enables automatic feature extraction for distant viewing analysis. Secondly, human figures—and by extension faces—are the most widely represented visual motif in representational painting. We use an off-the-shelf 3D Morphable Model (3DMM) to (1) detect faces in a painting, (2) regress sparse 2D geometry features, specifically landmark points, and (3) regress dense 3D geometry features, including vertices and their corresponding normal vectors. We use the 3DMM called SynergyNet [10], a model tested on artistic data with fast inference processing time to extract 3D geometry, alignment, and orientation from a single monocular image. Overall, our lighting model assumes the following:

- Skin reflectance is Lambertian and thus acts as a low pass filter on the incident illumination.
- Distant illumination (e.g., incident light rays are parallel)
- Uniform albedo and convexity (e.g., no inter-reflections)
- Enough geometrical variance of the 3DMM to capture 3D face geometries on paintings
- Luminance values in the CIE-LAB color space are a reasonable good approximation for ground truth light intensities.

As we operate under the assumption of uniform albedo across the sampled light values, we propose a ground truth refinement pipeline. The primary goal is to produce a ground truth mask that approximates true lighting conditions in the skin of the face. We assume the ground truth illumination values can be well approximated from the CIE-LAB luminance pixel values from the face. Additional pre-processing techniques, such as applying inverse gamma correction to address the non-linearity of digital cameras, were explored but ultimately not included, as these steps did not yield any significant qualitative or quantitative improvement. The computation of the ground truth mask involves several steps, as showed in Figure 2. We create the first mask by projecting the 3D mesh onto the image plane. This projection is straight-forward and does not involve any transformation matrices since SynergyNet uses orthographic projection. This mask is the input to a skin segmentation model [11], which removes undesired pixels from the projection by segmenting the skin-related pixels. Finally, we use the sparse 2D landmark points regressed from SynergyNet to remove the area inside the eyes and lips.

Figure 2. *Steps for the calculation of the ground truth mask of Figure 1.*

Using a Lambertian reflectance model, the lighting environment $L(\vec{N})$ of the face can be described in terms of the first three order spherical harmonics [12]. This is described in equation 1, where l_n^m represents the first nine spherical harmonic coefficients, from l_0^0 to l_2^2. The superscript m indicates the degree, and the subscript n indicates the order of the spherical harmonic. Finally, the variable $Y_n^m(\vec{N})$ represents the evaluation of the spherical harmonic basis functions over the surface normal vectors $\vec{N}$.

$$L(\vec{N}) = \sum_{n=0}^{2} \sum_{m=-n}^{n} l_n^m Y_n^m(\vec{N}) \tag{1}$$

Hence, our hypothesis proposes the calculated illumination using spherical harmonics to be equal to the sampled luminance values of our ground truth mask. Such hypothesis can be postulated as a simple linear system of equations, as expressed in equation 2. In this equation, M is a $N_{vertices}$ x $N_{coefficients}$ matrix containing the sampled spherical harmonic basis functions. In our analysis, $N_{vertices}$ represent a subset of vertices from the regressed 3D mesh, selectively including only those mapped to positive values in the ground truth mask, thus varying with each facial evaluation. The vector $\vec{v}$ is the nine spherical harmonic coefficient vector to be found, and $\vec{b}$ is the ground truth vector representing the sampled luminance values.

$$M\vec{v} = \vec{b} \tag{2}$$

Two main sources of noise need to be considered. First, imperfections in the ground truth due to occlusions, such as beards, or cast shadows from a helmet or a hand. Secondly, a 3D face mesh that does not align well with the target face can cause large variations in the calculation of the coefficients. As suggested in [13], we add a regularization term C to dampen the effects of such noise sources. Equation 3 shows the error function to be minimized, where C is the diagonal matrix $(1, 2, 2, 2, 3, 3, 3, 3, 3)$. In practice, the analytical solution showed almost identical results to gradient optimization methods for small values of λ. In all the conducted experiments, we set the value of λ to 0.1.

$$E(\vec{v}) = \min_{\vec{v}} \|M\vec{v} - \vec{b}\|^2 + \lambda \|C\vec{v}\|^2 \tag{3}$$

Solving for $\vec{v}$ we aim to find the set of coefficients that minimizes the difference between the predicted lighting values and the luminance mask of the face. The optimized coefficients represent the contribution of each basis function to describe the overall illumination on the surface of the face. We can use such coefficients to generate a light environment map, a spherical visualization that represents the overall light intensity and how light varies in different directions. Each coefficient contributes to a certain frequency and pattern of light variation on the sphere, as shown in figure 3. In spherical harmonics, the first order ($n = 0$) is constant, and represents the ambient light. The second order ($n = 1$) are linear polynomials which encode the directional information of light. Finally, third order ($n = 2$) spherical harmonics are quadratic polynomials that capture more nuanced variations in lighting.

Ultimately, the light value at each vertex of the 3D mesh is the evaluation of the calculated coefficients on the basis functions, as expressed in equation 2. This enables us to render the final 3D mesh, illuminated under the derived lighting conditions, as showed in Figure 1.

Evaluation

In this section, we tackle the primary challenge of evaluating the estimated lighting conditions from 2D digital paintings where no ground truth is available. Our aim is to test the model's functionality for art historical research. Specifically, the ability of the coefficients to differentiate between light environments, and the extent to which the regressed 3D geometry and luminance ground truth serve as appropriate data to estimate light direction.

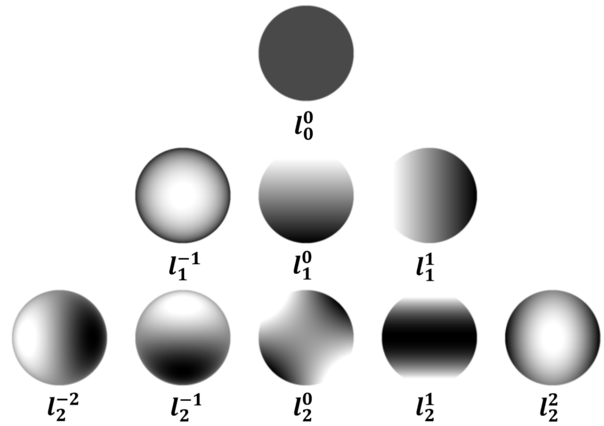

Figure 3. *Visualization of the overall contribution of each coefficient. First row (n=0), second row (n=1), thrid row (n=2). The environment map is the addition of all coefficient contributions, showed in Figure 1.*

Lighting environments

To test how well the proposed method distinguishes between lighting environments, we use a small dataset composed of three simple, yet different light scenarios. The Yale Face Database [1] contains images of 15 individuals under varying facial expression and lighting configuration. For each individual we select 3 images corresponding to each light environment: center light, left light and right light. We extract the spherical harmonic coefficients of each face to examine how sensitive each coefficient is to the direction of light.

Coefficients that vary significantly across light environments suggest a strong directional influence. From the first order spherical harmonic coefficients ($n = 1$), which encode directional information of light (see Figure 3), it is primarily the fourth coefficient l_1^1, aligned with the x-axis, that exhibits a substantial shift in value. This particular coefficient distinctly reflects the directional influence of the light source, exhibiting neutral values under centered lighting, negative values with left-ward lighting, and positive values when illuminated from the right, indicating its sensitivity to horizontal light positioning. Higher order coefficients related to the x axis, coefficients fifth l_2^{-2} and seventh l_2^0, follow the same behaviour, while the ninth coefficient l_2^2 is an exception. From the mathematical definition of the spherical harmonic basis functions, we know that positive values for l_2^2 means that there is always greater contribution of the x-axis over the y-axis, which is in correspondence to the expected behaviour. The presence of high variance within each coefficient reflects the intrinsic complexity of lighting environments, where subtle changes in the position of the light source, as well as the position and face geometry of the person being portrayed, can influence the assessment of the coefficients.

To further interpret the extent to which the extracted coefficients properly encode the lighting environments, we perform a classification evaluation using two linear algorithms: Logistic Regression and Support Vector Machine (SVM), and two non-linear algorithms: Gradient Boosting and Random Forest. This choice of models, while not exhaustive, provides a balanced overview,

[1] http://cvc.cs.yale.edu/cvc/projects/yalefaces/yalefaces.html

sufficient to assess the robustness of the coefficients. As the dataset is small, we use 10-fold cross-validation to prevent from overfitting. The aggregated results across models yield a mean accuracy of 90%, with precision, recall, and F1-scores averaging 91%, 90%, and 89%, respectively. The performance metrics span from a minimum accuracy of 89% to a maximum of 96%. The evaluated performance across all models suggest that the quantitative behaviour of the spherical harmonic coefficients matches the physical response of the actual light source position.

Light direction at scale

To further explore the consistency of the proposed method, we expand our performance evaluation to a distant viewing setting. In [14], authors proposed a study on light direction comprising around 10k artworks throughout western art history, where participants were asked to draw a 2D estimation of the light direction, along with their confidence of estimation. A comparison of the proposed method with human evaluations gives an idea of the extent to which the proposed 3D geometry and luminance ground truth serve as a reasonable approximation to encode light information for large scale analysis. This section explores the possibility of using human perception as another possible validation for our quantitative light measures.

We follow the strategy proposed in [15] to calculate the direction of illumination for 3D Lambertian surfaces from a distant light source. The least squares estimation expressed in Equation 4 allows us to calculate the 3D light direction given by coefficients L_x, L_y, L_z with an ambient light term A. This computation relies on the surface normal's components, described by matrix M, and their corresponding luminance values from the ground truth, captured in vector $\vec{b}$. Solving for $\vec{v}$, the calculation of the principal 2D light direction is straight forward using equation 5.

$$E(\vec{L},A) = \left\| M \begin{bmatrix} L_x \\ L_y \\ L_z \\ A \end{bmatrix} - \begin{bmatrix} I(x_1,y_1) \\ I(x_2,y_2) \\ \vdots \\ I(x_p,y_p) \end{bmatrix} \right\|^2 = \left\| M\vec{v} - \vec{b} \right\|^2 \qquad (4)$$

$$\theta = \tan^{-1}\left(\frac{L_x}{L_y}\right) \qquad (5)$$

We select paintings from 1400 to 1900 to focus the evaluation on representational paintings. To detect instances where there is a clear disagreement between human evaluations over the same painting, we adopted a criterion based on the Mean Absolute Deviation (MAD). We included paintings under two conditions: either they had only one evaluation (MAD $= 0$) and a confidence score above 4 (in a scale from 0 to 7), or they had multiple evaluations with a MAD less than or equal to two standard deviations above the MAD mean. For paintings with multiple evaluations, the mean value is used to determine the ground truth light direction. Recognizing the inherent complexity of light, we refine the analysis by selecting paintings of straightforward composition, specifically those featuring a single figure in a portrait-like setting, though not limited to traditional portraits. This approach makes it easier for human evaluators to discern light interactions, thereby facilitating a more precise examination of light.

Following this criteria, we use the zero-shot classification advantages of CLIP model [16] to filter the human evaluated dataset of the resulting 5887 paintings. Our approach involved assembling a benchmark dataset of 100 images containing a random sample of 50 portraits and 50 paintings. All images were gathered from the Web Gallery of Art [2]. While the Web Gallery of Art dataset may exhibit biases and does not offer a historically representative sample of painting schools, this bias does not significantly impact the assessment of light direction, making it sufficiently suitable for our analysis. For each image, a score value is calculated as the cosine distance between each visual CLIP embedding and a fixed prompt. Empirically, we found the following prompt to be effective in favour of classifying between the two set of images: "A photo of a figurative portrait painting." We calculate the optimal threshold using the benchmark dataset as the score value that maximizes the F1 score. Finally, we compute the score values for the human evaluated dataset and filter paintings below the optimal threshold. The resulting dataset after this filtering is a total of 1770 paintings.

We repeat the same procedure to remove paintings with artificial lighting, such as candles or divine light. Close light sources scatter light from multiple directions in the scene, so the direction of light is arbitrarily specified depending on a reference point that might differ among participants. In this case, we selected the prompt: "A photo of a painting depicting artificial light sources," and used 20 candlelight paintings from Gerrit van Honthorst (1592-1656) retrieved from the Web Gallery of Art. This process filtered 17 paintings from the dataset. A visual inspection over the resulting images after filtering reveals that the CLIP classification procedure is unstable, and undesired paintings are still found (e.g., landscapes, still-life, complex compositions). We tackle this issue by removing images where the model fails to detect any faces, or more than one face was detected.

Additionally, we want to detect instances where the 3DMM fails. When the 3D mesh is not regressed correctly, the misalignment between the depicted face and the 3D mesh compromises the light direction calculation. To automatically detect failure instances, we use the head pose information regressed from the 3D dense mesh. The head pose is defined by three rotational angles: yaw (rotation around the vertical axis), pitch (rotation around the horizontal axis), and roll (rotation around the longitudinal axis). In a portrait-like setting, art historical conventions generally exclude extreme values of these parameters, which allows us to detect failure instances of the regressed 3D mesh. Portrait-like paintings rarely show the back of the head, and the 3DMM is unlikely to detect faces in this orientation, so we limit the yaw axis to ±90 degrees. Pitch and roll axis are constrained within the range ±45 degrees. These thresholds are arbitrarily set within known anatomical limits, yet sufficiently broad to encompass the range of poses typically depicted in these artworks. Overall, a total number of 890 figurative portrait-like paintings were analyzed, from which 145 were detected as failure images and were hence discarded.

We propose the Mean Absolute Angular Error (MAAE) as the metric for performance evaluation. This metric provides a measure of the average magnitude of angular deviation from the ground truth light direction. It is calculated by averaging the ab-

[2] https://www.wga.hu/index_search.html

solute values of the differences between the ground truth and estimated angles, as expressed in equation 6.

$$\mathrm{MAAE} = \frac{1}{N}\sum_{i=1}^{N}\left|\theta_{\mathrm{gt},i} - \theta_{\mathrm{est},i}\right| \tag{6}$$

Early Renaissance paintings present a significant challenge, both in terms of successful processing by the off-the-shelf 3DMM model and in achieving precise light direction estimations with the proposed method, as illustrated in Figure 4. Stylistic conventions found in early Renaissance artworks minimize light contrast, and favored little use of self-shadowing compared to further art movements. This makes the task of identifying specific light directions more challenging for both, human evaluators and computational analysis.

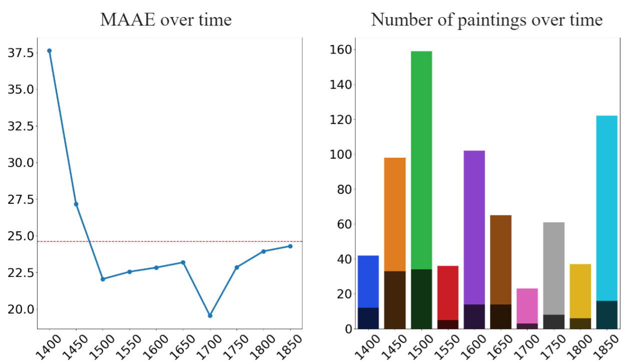

Figure 4. *On the left, the MAAE score per half-century. The red dashed line represents the overall average score. On the right, a count plot of paintings per half-century. The shaded areas represent the number of failed images.*

Moreover, our method calculates light direction in 3D. To compare our predictions to human evaluations, the 3D light direction estimations are mapped to the 2D plane of the image, which introduces distortions due to the loss of information along the z-axis. The type of lighting found in early Renaissance paintings, lacking clear clues of directional lights, tends to yield estimations where light appears to emanate directly towards the face. In this scenario, there is a significant contribution of the z-axis, which amplifies the distortion and the final error. As noted in [15], it is possible to calculate the 2D light direction using only the geometry information along the occluding boundary of the face, where the z-component of the surface normal is zero. This is only applicable for our method when the alignment between the 3D mesh and the face is perfect. That is the only scenario in which the 3D normal vectors, with a z-component of zero, actually correspond to the occluding boundary of the face. This strategy, hence, is not performative and thus not considered for our distant viewing analysis.

Numerous studies have underscored the challenges faced by human perception in precisely identifying and characterizing light environments [15] [17]. Naturally, the task becomes even harder for paintings [18] [19], where the room for interpretation increases. The standard deviation for human evaluations across the analyzed paintings shows that humans deviate 15.34 degrees on average, taking into account that 46% of the analyzed paintings had a single evaluation, and hence don't contribute to the deviation measure. The observed higher deviation in our method, averaging 24.4 degrees, shows that finding the light direction in paintings at scale remains a challenging task. While humans can use any other visual cues (e.g., cast shadows), our approach is limited to the information of the face. In particular scenarios where the face is not lit according to the overall scene, or objects cast shadows on the face (e.g., a hat), the light direction estimation can be compromised. The complexity of the task is further amplified by the inherent difficulty of working with paintings, where the diversity of stylistic variations, techniques, and visual features, introduce significant challenges to conduct a systematic analysis over centuries. We attempted to minimize this issue in our evaluation by not only selecting paintings that simplify the complexity of the lighting environments, but also paintings that bear a closer resemblance to the training dataset of the 3DMM model.

Case study: diffuse light in Netherlandish golden age painting

In *Light and shadow in Netherlandish art, 1600-1750: Theory and Practice*, Ulrike Kern studies light and shadow in the golden age by discussing the Netherlandish theories of art in relation to the pictorial practice of the time [20]. Over the seventeenth century, the handling of light and shadow changed considerably, as Kern describes it: "Shadows became more predominant in the paintings of Rubens and Rembrandt than in the art that was popular around 1600, while by contrast in the second half of the seventeenth century the demand for a more diffused light increased" [20, p. 20].

In this case study we propose to use our method to evaluate the diffuseness in lighting environments throughout seventeenth-century Netherlandish painting. As Kern's description refers to paintings in general, we assume that by extension, it also applies to portraits, a domain where our method exhibits lower susceptibility to inaccuracies. To this aim, we use data from the Web Gallery of Art, which offers an extensive collection of seventeenth-century paintings from the Dutch, Flemish, and Netherlandish schools. We consider this selection of data illustrative rather than evidential as the purpose of this study case is not to be historically representative but rather serve as an alternative evaluation method. Hence, our aim here is not to produce art historical knowledge, but to commensurate art historical writings with quantitative measures.

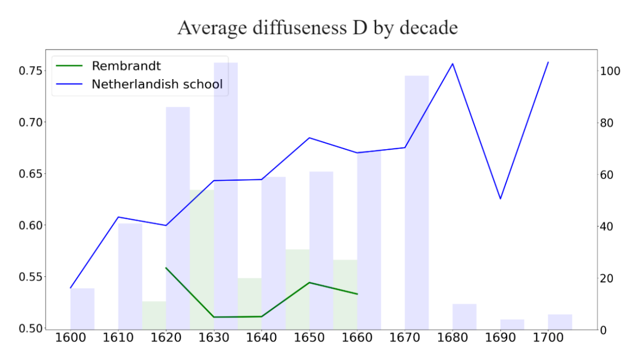

Figure 5. *Diffuseness metric D over time for Rembrandt and Netherlandish portraits. The left axis indicates the value of the D metric. The bar plots show the number of images per decade, indicated in the right axis.*

The extraction of spherical harmonics allows us to evaluate the magnitude of diffuseness in lighting environments. In a diffusely lit environment, faces are evenly illuminated, indicated by a high l_0 term for ambient light that lightens shadows, and low l_2 coefficients for subtle lighting variations. We propose to quantify diffuseness as the contribution of the low-order coefficients ($l_{0,1}$) with respect to the overall energy ($l_{0,1,2}$). The diffuseness metric, denoted as D, is confined to the range $D \in [0,1]$, where 0 represents no diffuseness and 1 signifies complete diffuseness. The calculation of D is straight forward, as expressed in equation 7, we calculate the ratio between Euclidean norms.

$$D = \frac{\|l_{0,1}\|^2}{\|l_{0,1,2}\|^2} \tag{7}$$

Following Kern's observation, we track the use of diffuse light in the Netherlandish school and compare it to the portraits of Rembrandt (1606-1669). Rubens (1577-1640) has been excluded from the analysis due to the limited availability of portraits with date information—only 33, while 173 Rembrandt portraits are available. Rembrandt's use of shadows are expected to produce a reduction in ambient light, manifested by pronounced self-shadows and darker tones on facial features. This reduction in ambient illumination inherently decreases diffuseness, so a lower value of D with respect to the Netherlandish school is expected.

We include the number of paintings analyzed by century in Figure 5 to acknowledge the interaction between the amount of data and the obtained metrics. It is noticeable that decades with a sparse amount of data correlate with significant fluctuations in the diffuseness metric. To derive towards preliminary conclusions we take into account only decades with enough portraits, ranging from 1610 to 1670. Figure 5 illustrates a trend that supports Kern's observations, revealing a gradual increase in the D metric over time for the Netherlandish school. The consistency of this upward trajectory reinforces the idea of an evolving artistic convention that increasingly favors diffuseness in light environments. Lower values of D in Rembrandt's portrait oeuvre reflects a stylistic preference for bold shadows and sharp contrasts, which in turn leads to decreased diffuseness.

Conclusions

Using faces as light probes, we extract spherical harmonic coefficients to analyze the light environment of paintings at scale. We conducted three evaluations to validate our approach for digital art history research. The evaluation on natural images for three different light environments showed that the extracted spherical harmonic coefficients match the physical response of the light environments. Large-scale analysis for light direction estimation remains a challenge, due to the inherent complexity of light, but primarily influenced by the stylistic variations and visual diversity present in paintings. These factors introduce significant challenges to conduct a systematic analysis over centuries. While the reported performance is close to human evaluation, the overall effectiveness of the model relies on the nature of the painting. We minimized this issue by analyzing light direction over portrait-like paintings, a type of image that bear a closer resemblance to the training set of the 3DMM model, and simplifies the complexity of the light environment. Our method faces challenges due to occlusion (e.g., cast-shadows, beards, or hats), that compromise the evaluation of lighting environments. Future work includes relaxing the constant albedo assumption by introducing a rigorous color segmentation method to remove noise from the ground truth mask. We finally observe that the proposed metric to evaluate the amount of diffuseness in paintings aligns with a trend supported by art historical context.

References

[1] M. Barasch, *Light and Color in the Italian Renaissance Theory of Art*. Print, New York: New York UP, 1978.

[2] P. Hills, *The Light of Early Italian Painting*. Yale University Press, 1987.

[3] D. G. Stork, "Computer analysis of lighting style in fine art: steps towards inter-artist studies," (San Francisco Airport, California, USA), p. 786903, Feb. 2011.

[4] L. L. Impett, "Painting by Numbers: Computational Methods and the History of Art," p. 326.

[5] Saboya Yang, G. Cheung, P. Le Callet, Jiaying Liu, and Zongming Guo, "Computational modeling of artistic intention: Quantify lighting surprise for painting analysis," in *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, (Lisbon, Portugal), pp. 1–6, IEEE, June 2016.

[6] D. G. Stork and M. K. Johnson, "Lighting analysis of diffusely illuminated tableaus in realist paintings: an application to detecting "compositing" in the portraits of Garth Herrick," (San Jose, CA), p. 72540L, Feb. 2009.

[7] P.-P. Sloan, J. Kautz, and J. Snyder, "Precomputed Radiance Transfer for Real-Time Rendering in Dynamic, Low-Frequency Lighting Environments,"

[8] R. Ramamoorthi and P. Hanrahan, "A signal-processing framework for inverse rendering," in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pp. 117–128, ACM, Aug. 2001.

[9] A. Dib, J. Ahn, C. Thebault, P.-H. Gosselin, and L. Chevallier, "S2F2: Self-Supervised High Fidelity Face Reconstruction from Monocular Image," Apr. 2022. arXiv:2203.07732 [cs].

[10] C.-Y. Wu, Q. Xu, and U. Neumann, "Synergy between 3DMM and 3D Landmarks for Accurate 3D Facial Geometry," Jan. 2024. arXiv:2110.09772 [cs].

[11] I. Perov, D. Gao, N. Chervoniy, K. Liu, S. Marangonda, C. Umé, M. Dpfks, C. S. Facenheim, L. RP, J. Jiang, S. Zhang, P. Wu, B. Zhou, and W. Zhang, "DeepFaceLab: Integrated, flexible and extensible face-swapping framework," June 2021. arXiv:2005.05535 [cs, eess].

[12] R. Ramamoorthi and P. Hanrahan, "An efficient representation for irradiance environment maps," in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pp. 497–500, ACM, Aug. 2001.

[13] M. K. Johnson and H. Farid, "Exposing Digital Forgeries in Complex Lighting Environments," *IEEE Transactions on Information Forensics and Security*, vol. 2, pp. 450–461, Sept. 2007.

[14] C.-C. Carbon and A. Pastukhov, "Reliable Top-Left Light Convention Starts With Early Renaissance: An Extensive Approach Comprising 10k Artworks," *Frontiers in Psychology*, vol. 9, p. 454, Apr. 2018.

[15] M. K. Johnson and H. Farid, "Exposing digital forgeries by detecting inconsistencies in lighting," in *Proceedings of the 7th workshop on Multimedia and security*, (New York NY USA), pp. 1–10, ACM, Aug. 2005.

[16] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning Transferable Visual Models From Natural Language Supervision," Feb. 2021. arXiv:2103.00020 [cs].

[17] J. Sun and P. Perona, "Where is the sun?," *Nature Neuroscience*, vol. 1, no. 3, pp. 183–184, 1998. Place: United Kingdom Publisher: Nature Publishing Group.

[18] T. Kartashova, H. d. Ridder, S. F. t. Pas, M. Schoemaker, and S. C. Pont, "The visual light field in paintings of Museum Prinsenhof: comparing settings in empty space and on objects," in *Human Vision and Electronic Imaging XX*, vol. 9394, pp. 554–563, SPIE, Mar. 2015.

[19] R. Casati and P. Cavanagh, *The Visual World of Shadows*. MIT Press, 2019.

[20] I. Kern, "Light and shadow in Netherlandish art 1600-1725 : Theory and practice," p. 245, 2010.

Author Biography

Pepe Ballesteros Zapata is a PhD fellow at the institute of Digital Visual Studies (University of Zurich - Max Plank Society) applying computational methods to the study of depicted light in art. He graduated from telecommunications engineering and received his M.Sc. degree in 'Signal Processing and Machine Learning for Big Data' from Universidad Politécnica de Madrid (UPM). His master thesis was selected winner of the national Spanish contest launched by RTVE called "Impulsa Visión Ayudas a la Investigación III", where he developed a text generation system to automatically write weather forecasts in Spanish. At the moment, his research interests focus on the intersection between art history, computer science and critical digital art history.