https://doi.org/10.2352/lim.2024.5.1.14

A Pipeline for Characterising Virtual Reality Head Mounted Displays

Ujjayanta Bhaumik, Laurens Van de Perre, Frédéric B. Leloup; Light&Lighting Laboratory, KU Leuven (ESAT), BE-9000 Ghent, Belgium

Abstract

With the increasing popularity across various scientific research domains, virtual reality serves as a powerful tool for conducting colour science experiments due to its capability to present naturalistic scenes under controlled conditions. In this paper, a systematic approach for characterising the colorimetric profile of a head mounted display is proposed. First, a commercially available head mounted display, namely the Meta Quest 2, was characterised by aid of a colorimetric luminance camera. Afterwards, the suitability of four different models (Look-up Table, Polynomial Regression, Artificial Neural Network and Gain Gamma Offset) to predict the colorimetric features of the head mounted display was investigated.

Introduction

Virtual Reality (VR) has emerged as an innovative technology, finding its use in diverse fields such as healthcare [1, 2, 3], education [4, 5] and psychology [6, 7, 8]. Its capacity to create immersive, controlled environments has made it invaluable for conducting scientific experiments [6, 9], particularly in colour science. In this context, the fidelity of colour reproduction in VR environments, facilitated by Head Mounted Displays (HMDs), is critical for ensuring the realism and accuracy of visual stimuli. VR HMDs are used to present stimuli to participants, necessitating precise colour reproduction to maintain experimental accuracy. Such experiments underscore the importance of establishing precise relationships between device settings and device-independent photometric and colorimetric attributes [6, 10, 11].

While past research about display characterisation has focused on the colour characterisation of conventional displays such as Liquid Crystal Displays [12, 13], research regarding VR HMDs is still ongoing. Early HMDs were less colorimetrically accurate and had problems, for instance, producing an output that has blue distortion. Current HMDs perform much better, as evidenced by experiments. [14, 15, 16].

Modern displays can be prone to additivity issues though, which arise when the sum of different channel outputs does not match that of the combined channel output. Therefore, basic characterisation models may not be suitable for HMD characterisation [17, 18]. Prior to the presented work, a study was conducted that systematically compared the colorimetric performance of various colour characterisation models through simulation [19]. The current work transcends the domain of simulations and introduces field experiments with a commercially available HMD, namely the Meta Quest 2. A pipeline is designed to perform a colour characterisation of the HMD. First, the stability of the HMD is measured, followed by testing the linearity and additivity. After these basic tests, four carefully chosen characterisation models are trained on selected training points spread across the chromaticity gamut of the HMD. Recognising the inherent challenges in HMDs, such as colour channel interactions and additivity issues, empirical data from the headset is investigated and rigorously evaluated. By doing so, the aim is to validate the effectiveness of the characterisation models in practical, real-world scenarios and to present a proper method to characterise an HMD. The presented method can easily be generalised for profiling and characterising HMDs, and also makes it easy to compare different devices. The results and discussions provided in this paper present the colorimetric characteristics of Meta Quest 2 HMD and shed light on the performances of each model, allowing one to make informed decisions when choosing the most suitable colour characterisation method for their specific applications in VR and other high-fidelity settings. The culmination of these efforts aims to establish a robust framework for precise colorimetric characterisation, address the limitations of existing models, choose a proper colour characterisation model, and facilitate more accurate colour rendering in scientific experiments.

Setup

An Alienware m15 laptop (16 GB RAM, 64-bit Windows 10 Enterprise operating system, Intel® Core™ i7-8750H CPU operating at 2.2 GHz, NVIDIA GeForce GTX 1060 graphics card with 6 GB Video RAM) was used in the study. The Alienware laptop and the Meta Quest 2 HMD were connected by the Oculus Link cable in conjunction with the Oculus desktop app, which is the primary software interface for managing the headset.

Methodology

HMD characterisation pipeline

Stability

The stability assessment of the HMD followed a structured method. Initially, a fully black image was displayed for 30 minutes, after which four different images (Red, Green, Blue, and White) were successively displayed, for 120 minutes each. Between the four presented images, the black screen was shown for 30 minutes each time.

A colorimetric camera (TechnoTeam LMK-6-12 Colour), incorporating a high-resolution image sensor of 4078 by 2998 pixels and measuring in CIE XYZ colour space (1931 2° observer), was used for HMD characterisation.

For each of the four images, colorimetric luminance images were continuously captured and analysed. Stability was assessed by calculating the relative difference between the maximum and minimum luminance channel outputs recorded within a fifteen minutes timeframe. The HMD was assessed as stable when the calculated relative difference is less than 0.5% of the minimum observed reading in the time window [20].

Linearity and Additivity

The linear relationship between driver values and displayed luminance was checked for each channel (R,G,B) of the Meta Quest 2 HMD. To this end, the luminance of a set of input images was measured. The input images included the R, G and B channels set to values encompassing the entire scale from 0 to 255, with an increment of 6. Additivity assessment was conducted through the comparison of the sum of tristimulus values obtained from separately displayed RGB colour channels with the tristimulus values derived from the corresponding achromatic stimulus. Initially, images showcasing maximum intensity in the red, green and blue channels were captured. The mean XYZ values were computed from the region of interest (ROI), comprising the top 5 percentile pixels in each luminance image. Subsequently, an image featuring the maximum white stimulus ($R = G = B = 255$) was presented. To assess additivity, the cumulative XYZ values of the individual colour channels were compared with the XYZ values of the white stimulus. In an ideal scenario, the sum of the luminance values for the red, green, and blue channels should equal the luminance value obtained for the corresponding white image. The same procedure was repeated for all other equal $R/G/B$ combinations, set to the same value ranging between 0 and 255, with an increment of 6.

Assessment of prediction models

In a prior study by the authors [19], the suitability of matrix-based models (e.g., Gain Offset Gamma Offset (GOGO), Gain Gamma Offset (GGO), Gain Offset Gamma (GOG)), and other models like Polynomial Regression (POR), Look-up Table (LUT) and Artificial Neural Network (ANN)-based models were assessed using simulation data. Findings revealed that POR, LUT and ANN-based models outperformed the other alternatives, achieving a CIELAB colour difference (ΔE^*_{ab}) of less than 1 between input and predicted colours. Consequently, these three models were selected for testing the Meta Quest 2 HMD. For comparison, a simple GGO model was also included.

Colorimetric accuracy of each of the four different models was tested with the following procedure. First, all models were trained on a dataset of 630 images. This dataset includes a series of images that spans the three individual channels, as well as images formed from combinations of two and all three channels, resulting in cyan, magenta, yellow and achromatic hues. A systematic increment of 6 was applied to each individual and distinct combination of colour channels extending up to a driver value of 252, resulting in a total of 295 images (including black). As such, the dataset counts 42 pure red, green and blue images. The same number of cyan, magenta, yellow and grey images is produced, while there is one pure black image. Apart from these combinations, 7 additional driver RGB values were chosen from the set $I = \{0, 42, 85, 128, 170, 212, 255\}$, with which 335 additional unique combinations of values were created. Following on the presentation of each image in the HMD in a random order, a measurement was performed with the colorimetric luminance camera. From the defined ROI the average tristimulus XYZ values were calculated, resulting in 630 (XYZ, RGB) training pairs.

To test the accuracy of each trained colour characterisation model, 19 test points were carefully selected, encompassing the chromaticity gamut of the HMD device. The chromaticity gamut of the device was determined from the maximum red, green and blue output, converting the respective XYZ values into CIE 1976 Y, u', v' coordinates using LuxPy [21]. The 19 different test points were selected to uniformly cover the chromaticity gamut of the HMD. First, 9 points were selected on the boundary of a triangular 80% area of the HMD gamut. Let the vectors v_r, v_g, v_b represent the vertices of the triangle in $u'v'$ space, corresponding to the vertices of that selected area. The test points on the boundary are given by Eqs. 1-3. To ensure that all test points are inside the gamut of the HMD, the luminance value of each point was calculated by selecting the ten nearest points from the training set, and their average luminance was selected as the luminance of that test point. Next, 10 additional test points were chosen from the interior of the selected area of the chromaticity gamut. The colour coordinates are calculated from Eqs. 4-8 where $P_{b,x} (1 \le x \le 9)$ and $P_{i,y} (1 \le y \le 10)$ refer to the boundary and interior points, respectively.

$$P_{b,1} = \frac{v_r + v_g}{2}, \qquad P_{b,2} = \frac{v_g + v_b}{2}, \qquad P_{b,3} = \frac{v_r + v_b}{2} \tag{1}$$

$$P_{b,4} = \frac{3v_r + v_g}{4}, \qquad P_{b,5} = \frac{3v_r + v_b}{4}, \qquad P_{b,6} = \frac{3v_g + v_b}{4} \tag{2}$$

$$P_{b,7} = \frac{v_r + 3v_g}{4}, \qquad P_{b,8} = \frac{v_g + 3v_b}{4}, \qquad P_{b,9} = \frac{v_r + 3v_b}{4} \tag{3}$$

$$P_{i,1} = 0.25v_b + 0.75\frac{v_r + v_g}{2}, \quad P_{i,2} = 0.50v_b + 0.50\frac{v_r + v_g}{2} \tag{4}$$

$$P_{i,3} = 0.75v_b + 0.25\frac{v_r + v_g}{2}, \quad P_{i,4} = 0.25v_r + 0.75\frac{v_b + v_g}{2} \tag{5}$$

$$P_{i,5} = 0.50v_r + 0.50\frac{v_b + v_g}{2}, \quad P_{i,6} = 0.75v_r + 0.25\frac{v_b + v_g}{2} \tag{6}$$

$$P_{i,7} = 0.25v_g + 0.75\frac{v_r + v_b}{2}, \quad P_{i,8} = 0.50v_g + 0.50\frac{v_r + v_b}{2} \tag{7}$$

$$P_{i,9} = 0.75v_g + 0.25\frac{v_r + v_b}{2}, \quad P_{i,10} = \frac{v_r + v_g + v_b}{3} \tag{8}$$

Finally, to test the trained characterisation models, a testing procedure was used in which the trained models are employed to predict RGB values ($RGB_{predict}$) corresponding to target XYZ values (XYZ_{target}) of the 19 test points. These RGB values are displayed in the HMD and from the acquired luminance image, XYZ values are calculated ($XYZ_{measured}$). To determine the colorimetric accuracy of the trained models, XYZ_{target} and $XYZ_{measured}$ are compared.

Results and Discussion

Stability

Stability data were obtained according to the method described earlier. The blue, green and (combined) white channels reached stability within the minimum stabilisation time period of 30 minutes. The red channel took around 36 minutes to stabilise.

To assess the repeatability of the HMD, all measurements were done twice (on two different days). Thus, two training datasets were obtained, each containing 630 images. The average ΔE^*_{ab} between the two days for all the 630 points was found to be 0.32, which clearly indicates a good repeatability of the device.

Linearity and Additivity

The variation in luminance (Y values) with driver values shows a non-linear relationship. The sum of tristimulus values of the individual primaries $[(R = 255, G = B = 0); (G = 255, R = B = 0); (B = 255, R = G = 0)]$ exceeds the tristimulus value of the white stimulus ($R = G = B = 255$). This observation aligns with prior research findings which also pointed toward additivity issues with modern HMDs [17, 18]. These findings are exemplified numerically in Table 1, in which the XYZ values of the red, green, and blue primary at maximum driver output are presented together with the XYZ values of the white stimulus (first day measurements). The summed X, Y and Z values of the RGB primaries are 3.49%, 2.27% and 3.63% higher than the X, Y and Z values of the white stimulus, respectively. This is consistent with the results described in the literature [17, 18].

Table 1: Colour space specifications for Meta Quest 2

Primary	***X***	***Y***	***Z***	***u'***	***v'***
Red	27.98	14.83	1.47	0.439	0.524
Green	26.68	52.08	9.56	0.128	0.560
Blue	14.44	6.53	72.90	0.174	0.178
White	66.77	71.81	80.98	0.193	0.466
Combined *RGB*	69.10	73.45	83.93	0.194	0.465

Characterisation

To characterize the Meta Quest 2 HMD, the ANN, LUT, POR and GGO models were used. For the ANN model, a network with three deep hidden layers comprising 200, 200, and 100 neurons, respectively, was chosen based on results obtained from the colorimetric performance of various colour characterisation models through simulation. The used LUT was three dimensional and the chosen interpolation method was linear. Several tests were done to discern the best degree of a polynomial, resulting in the adoption of a degree 4 polynomial for the POR model.

In Fig. 1, the chromaticity gamut of the Meta Quest 2 HMD is presented as a black triangle in CIE 1976 $u'v'$ colour space. The 19 test points to assess the four different colour characterisation models are annotated as ‘x’ markers and ‘+’ markers, indicating the boundary and interior points, respectively.

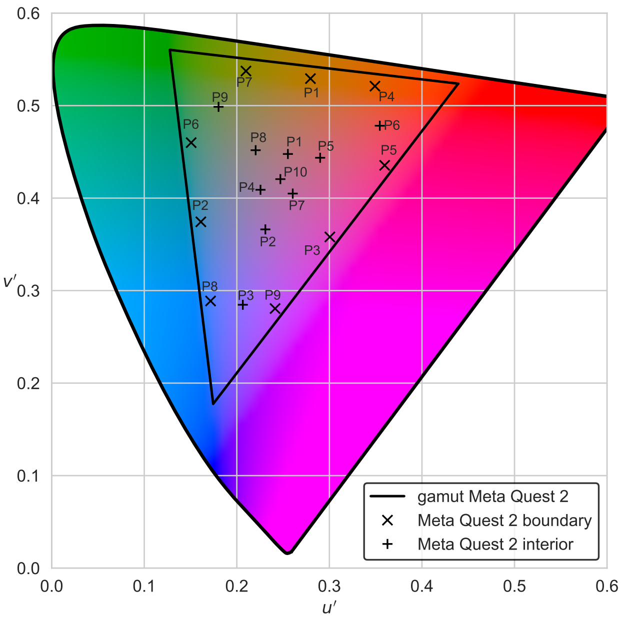

Figure 1: The chromaticity gamut of the Meta Quest 2 HMD is shown as a black triangle. The boundary points are annotated as ‘x’ markers and the interior points as ‘+’ markers. Note that the colours on the diagram are solely for reference and may not accurately reflect specific colours.

In Fig. 2, boxplots are presented that show the colour difference between the 19 defined test colours and their matching set colours as derived from the four characterisation models, expressed as ΔE^*_{ab}. The boxplots are grouped based on the capture days (Day 1 and Day 2), followed by the type of test point (left vs. right part of individual graphs: 10 interior points vs 9 boundary points). The results for each applied calibration model are presented in a different colour.

The choice of POR or LUT as a colour characterisation model seems to be the best option. Especially for the interior test points, both models outperform the ANN and GGO models by a large margin. Taking into account only the interior points, the POR model achieves a mean ΔE^*_{ab} of 1.81 over both capture days, similarly the LUT achieves a mean ΔE^*_{ab} of 1.84. In general, the ANN model does not perform well. This could potentially be attributed to the fact that building a good neural network requires a considerable amount of data and parameter tuning. Since the LUT and POR models already perform better, they are a better choice considering the time and effort of data collection. Besides the inferior general performance, among all models, the ANN also possesses the largest interquartile range, followed by the GGO model. This shows that there is a higher degree of variability in observed colour differences depending on the test points for these two types of colour characterisation models. This finding is even more valid for the boundary test points, for which for all tested characterisation models, large interquartile ranges are observed. Thus, all models performed considerably worse for the boundary test points than for the interior test points.

Conclusion

In this paper, four different colour characterisation models (LUT, POR, ANN and GGO) to characterise an HMD were tested in practice, on a commercially available HMD: the Meta Quest 2. From the results, it can be concluded that the POR and LUT colour characterisation models perform significantly better than the ANN or GGO models. The performance of colour characterisation models is affected by the location of a point on the colour gamut of a device, for instance, whether it is located closer to the gamut boundary or situated considerably inside the gamut. All four models perform considerably worse for boundary points, possibly due to higher energy output and power consumption. In summary, for conducting colour science experiments in which achieving high colour accuracy is important, the use of a suitable colour characterisation model is advised. In this practical study, the POR and LUT models were the preferable colour characterisation models, since they performed significantly better than the other tested models. Future work will consist of testing the presented colour characterisation pipeline on a more recent HMD from the same manufacturer.

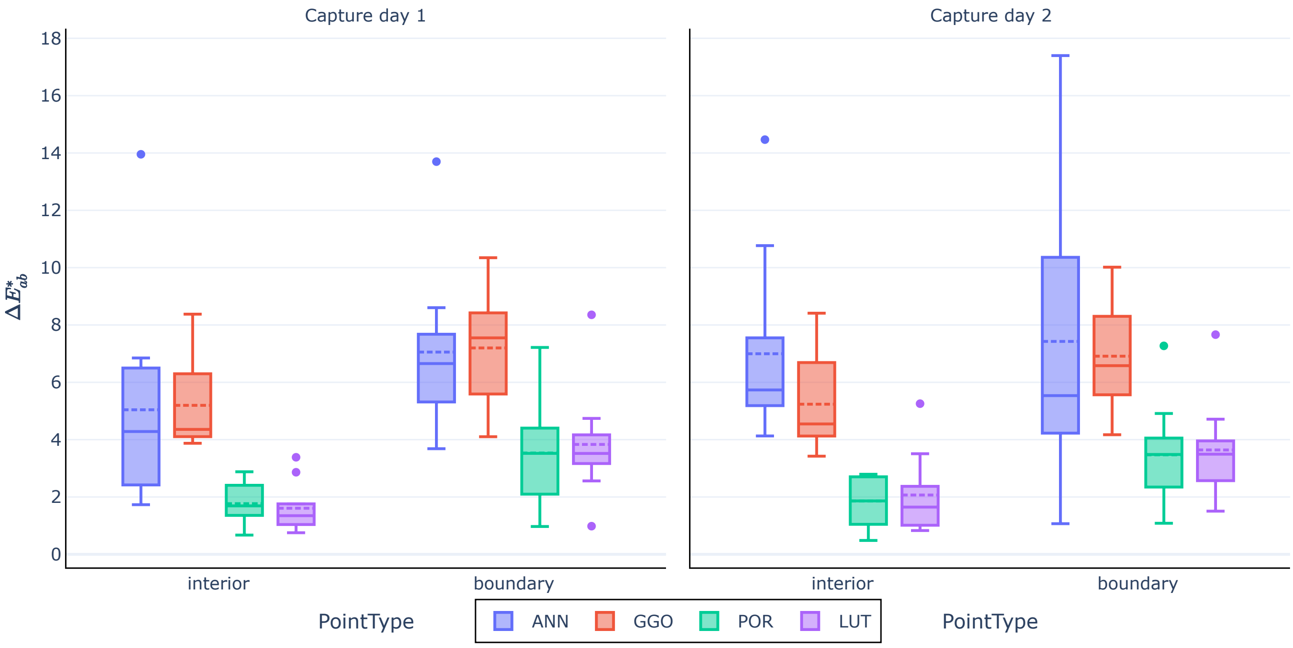

Figure 2: Performance of various colour characterisation models across two different days for Meta Quest 2 HMD. The colour difference between the 19 defined test colours and their matching set colours as derived from the four characterisation models are expressed as ΔE^*_{ab}. The dashed lines in each boxplot represent the mean ΔE^*_{ab}, while the solid lines represent the median ΔE^*_{ab} values. The bar represents the interquartile range of the data, and the small circular dots represent the outliers.

Acknowledgments

Author Ujjayanta Bhaumik and Laurens Van de Perre are supported through a junior research project G057021N of the Research Foundation – Flanders (FWO).

References

[1] H. Makinen, E. Haavisto, S. Havola, J.-M. Koivisto, User experiences of virtual reality technologies for healthcare in learning: an integrative review, Behaviour & Information Technology. 41 (2022) 1–17. https://doi.org/10.1080/0144929X.2020.1788162

[2] O.L. Chavez, L.-F. Rodriguez, J.O. Gutierrez-Garcia, A comparative case study of 2D, 3D and immersive-virtual-reality applications for healthcare education, International Journal of Medical Informatics. 141 (2020) 104226. https://doi.org/10.1016/j.ijmedinf.2020.104226

[3] G. Riva, Virtual reality for health care: the status of research, Cyberpsychology & Behavior. 5 (2002) 219–225. https://doi.org/10.1089/109493102760147213

[4] M.A. Rojas-Sanchez, P.R. Palos-Sanchez, J.A. Folgado-Fernandez, Systematic literature review and bibliometric analysis on virtual reality and education, Education and Information Technologies. 28 (2023) 155–192. https://doi.org/10.1007/s10639-022-11167-5

[5] S. Kavanagh, A. Luxton-Reilly, B. Wuensche, B. Plimmer, A systematic review of virtual reality in education, Themes in Science and Technology Education. 10 (2017) 85–119. https://www.learntechlib.org/p/182115

[6] R. Gil Rodrıguez, F. Bayer, M. Toscani, D. Guarnera, G.C. Guarnera, K.R. Gegenfurtner, Color calibration of a head mounted display for Color vision research using virtual reality, SN Computer Science. 3 (2022) 1–10. https://doi.org/10.1007/s42979-021-00855-7

[7] R.A. Elphinston, A. Vaezipour, J.A. Fowler, T.G. Russell, M. Sterling, Psychological therapy using virtual reality for treatment of driving phobia: a systematic review, Disability and Rehabilitation. 45 (2023) 1582–1594. https://doi.org/10.1080/09638288.2022.2069293

[8] C.J. Wilson, A. Soranzo, others, The use of virtual reality in psychology: A case study in visual perception, Computational and Mathematical Methods in Medicine. 2015 (2015). https://doi.org/10.1155/2015/151702

[9] M.A. Cohen, T.L. Botch, C.E. Robertson, The limits of Color awareness during active, real-world vision, Proceedings of the National Academy of Sciences. 117 (2020) 13821–13827. https://doi.org/10.1073/pnas.1922294117

[10] R.G. Rodriguez, M. Toscani, D. Guarnera, G.C. Guarnera, F. Bayer, K. Gegenfurtner, Color constancy in a virtual reality environment, Journal of Vision. 20 (2020) 1226–1226. https://doi.org/10.1167/jov.20.11.1226

[11] H. Cwierz, F. Diaz-Barrancas, J.G. Llinas, P.J. Pardo, On the validity of virtual reality applications for professional use: A case study on color vision research and diagnosis, IEEE Access. 9 (2021) 138215–138224. https://doi.org/10.1109/ACCESS.2021.3118438

[12] J.E. Gibson, M.D. Fairchild, Colorimetric characterization of three computer displays (LCD and CRT), Munsell Color Science Laboratory Technical Report. 40 (2000). http://www.sgidepot.co.uk/vw/PDFs/Colorimetric_Characterization.pdf

[13] E.A. Day, L. Taplin, R.S. Berns, Colorimetric characterization of a computer-controlled liquid crystal display, Color Research & Application. 29 (2004) 365–373. https://doi.org/10.1002/col.20046

[14] A. Mehrfard, J. Fotouhi, G. Taylor, T. Forster, N. Navab, B. Fuerst, A comparative analysis of virtual reality head-mounted display systems, ArXiv Preprint ArXiv:1912.02913. (2019). https://doi.org/10.48550/arXiv.1912.02913

[15] A. Mehrfard, J. Fotouhi, G. Taylor, T. Forster, M. Armand, N. Navab, B. Fuerst, Virtual reality technologies for clinical education: evaluation metrics and comparative analysis, Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization. 9 (2021) 233–242. https://doi.org/10.1080/21681163.2020.1835559

[16] Francisco Díaz Barrancas; Raquel Gil Rodríguez; Avi Aizenman;

Florian S. Bayer; Karl R. Gegenfurtner, Color Calibration in Virtual Reality Using Different Head Mounted Displays, Vision Sciences Society Annual Meeting Abstract, Journal of Vision. (2023). https://doi.org/10.1167/jov.23.9.5257

[17] H. Ha, Y. Kwak, H. Kim, Y. Seo, Discomfort luminance level of head-mounted displays depending on the adapting luminance, Color Research & Application. 45 (2020) 622–631. https://doi.org/10.1002/col.22509

[18] M. Toscani, R. Gil, D. Guarnera, G. Guarnera, A. Kalouaz, K.R. Gegenfurtner, Assessment of OLED head mounted display for vision research with virtual reality, in: 2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), IEEE, 2019: pp. 738–745. https://doi.org/10.1109/SITIS.2019.00120

[19] U. Bhaumik, F.B. Leloup, K. Smet, Systematic comparison of head mounted display Colorimetric performance using various Color characterization models, Optics Continuum. 2 (2023) 1490–1504. https://doi.org/10.1364/OPTCON.493238

[20] CIE, Test methods for LED lamps, LED luminaires and LED modules, CIE S 025/E: 2015, in: CIE Vienna, Austria, 2015.

[21] K.A. Smet, Tutorial: the LuxPy Python toolbox for lighting and color science, Leukos. (2019). https://doi.org/10.1080/15502724.2018.1518717

Author Biography

Ujjayanta Bhaumik holds an MSc in Computer Graphics, Vision, and Imaging from University College London and a B.Tech. in Computer Science and Engineering from the National Institute of Technology Jamshedpur, India. Since 2021, he has been pursuing a PhD at the Light&Lighting Laboratory, KU Leuven, Belgium on the perception of light and colour in real and virtual environments.

Laurens Van de Perre earned a Professional Bachelor's degree in Electronics – ICT and a Master of Engineering Technology in Electronics-ICT from KU Leuven. He began his career in 2014 as a scientific researcher at the Light&Lighting Laboratory, where he focused on the correlation between objective and subjective descriptors of lit indoor environments, earning his doctorate in 2023. Currently, Laurens is a postdoctoral researcher investigating light perception in virtual reality environments at KU Leuven. He also holds a postgraduate degree in industrial business policy, with a specialization in project management and business administration.

Frédéric Leloup received his MSc in Electromechanical Engineering in 2001 and PhD in Engineering in 2012, both from KU Leuven (Belgium). He worked as a research fellow and postdoc researcher at the Light&Lighting Laboratory research group, before being appointed associate professor in the Department of Electrical Engineering at KU Leuven in 2022. His research focuses on hard and soft lighting metrology. Frédéric is president of the Belgian Institute on Illumination and National Representative within CIE Division 1 (Vision and Colour).