The Cube++ Illumination Estimation Dataset

EGOR ERSHOV[1], ALEXEY SAVCHIK[1], ILLYA SEMENKOV[1], NIKOLA BANIĆ[2], (Member, IEEE),
ALEXANDER BELOKOPYTOV[1], DARIA SENSHINA[1],
KARLO KOŠČEVIĆ[3], (Student Member, IEEE),
MARKO SUBAŠIĆ[3], (Member, IEEE), AND SVEN LONČARIĆ[3]

[1]Institute for Information Transmission Problems, Russian Academy of Sciences, 119991 Moscow, Russia
[2]Gideon Brothers, 10000 Zagreb, Croatia
[3]Faculty of Electrical Engineering and Computing, University of Zagreb, 10000 Zagreb, Croatia

Corresponding author: Egor Ershov (ershov@iitp.ru)

ABSTRACT Computational color constancy has the important task of reducing the influence of the scene illumination on the object colors. As such, it is an essential part of the image processing pipelines of most digital cameras. One of the important parts of the computational color constancy is illumination estimation, i.e. estimating the illumination color. When an illumination estimation method is proposed, its accuracy is usually reported by providing the values of error metrics obtained on the images of publicly available datasets. However, over time it has been shown that many of these datasets have problems such as too few images, inappropriate image quality, lack of scene diversity, absence of version tracking, violation of various assumptions, GDPR regulation violation, lack of additional shooting procedure info, etc. In this paper a new illumination estimation dataset is proposed that aims to alleviate many of the mentioned problems and to help the illumination estimation research. It consists of 4890 images with known illumination colors as well as with additional semantic data that can further make the learning process more accurate. Due to the usage of the SpyderCube color target, for every image there are two ground-truth illumination records covering different directions. Because of that, the dataset can be used for training and testing of methods that perform single or two-illuminant estimation. This makes it superior to many similar existing datasets. The datasets, it's smaller version SimpleCube++, and the accompanying code are available at https://github.com/Visillect/CubePlusPlus/.

INDEX TERMS Color constancy, dataset, illumination estimation, white balancing, multiple illumination, mixed illumination.

I. INTRODUCTION

The human visual system is able, in some conditions, to recognize colors despite the influence of the illumination on their appearance through the ability known as color constancy [1]. It is not yet fully understood how this ability functions and therefore it is not possible to directly model it. Nevertheless, various computational color constancy methods are used in the pipelines of digital cameras. They are usually designed to first identify the chromaticity of the light source and then to remove its influence on the scene. The last one is described in details here [2]–[5]. For both of these tasks, the commonly used image formation model that also includes the

Lambertian assumption is usually given as

$$f_c(\mathbf{x}) = \int_\omega I(\lambda, \mathbf{x})R(\lambda, \mathbf{x})\rho_c(\lambda)d\lambda \qquad (1)$$

where $\mathbf{x}$ is a pixel in the image $\mathbf{f}$, $c \in \{R, G, B\}$ is the color channel, λ is a wavelength in the visible light spectrum ω, $I(\lambda, \mathbf{x})$ is the spectral distribution of the light source, $R(\lambda, \mathbf{x})$ is the surface reflectance, and $\rho_c(\lambda)$ is the camera sensitivity for the color channel c. It is often assumed that the scene illumination is uniform. This means that the spatial information is not required in the illumination estimation equations and so the color of the observed light source $\mathbf{e}$ is

$$\mathbf{e} = \begin{pmatrix} e_R \\ e_G \\ e_B \end{pmatrix} = \int_\omega I(\lambda)\boldsymbol{\rho}(\lambda)d\lambda. \qquad (2)$$

For a somewhat satisfying color correction, it is already enough to know the direction of **e** [6], which means that **e** can be described by chromaticities instead of colors. For example, *r*, *g*, and *b* chromaticity components are calculated as *R*, *G*, and *B* color components divided by their sum so that $r + g + b = 1$. Thus, knowing only two of them is enough.

Since there are more unknowns than equations, illumination estimation is an ill-posed problem and additional assumptions have to be made in order to tackle it. Because of that, numerous illumination estimation methods with various assumptions have been proposed and they are often divided into two groups: the low-level statistics-based methods and the learning-based methods.

The low level statistics-based methods include White-Patch [7], [8] and its improvements [9]–[11], Gray-World [12], Shades-of-Gray [13], 1st and 2nd order Gray-Edge [14], Weighted Gray-Edge [15], using bright pixels [16], gray pixels [17] or bright and dark colors [18], exploiting illumination perception [19] and expectation [20], etc. Interesting to note that ≪gray balancing≫ occurs also in scope of printer calibration [21].

Learning-based methods include neural networks [22], high-level visual information [23], natural image statistics [24], Bayesian learning [25], [26], spatio-spectral learning [27], methods restricting the illumination solution space [28]–[30], color moments [31], regression trees with simple features from color distribution statistics [32], spatial localizations [33], [34], convolutional neural networks [35]–[38] and genetic algorithms [39], modelling color constancy by using the overlapping asymmetric Gaussian kernels with surround pixel contrast based sizes [40], finding paths for the longest dichromatic line produced by specular pixels [41], detecting gray pixels with specific illuminant-invariant measures in logarithmic space [42], channel-wise pooling the responses of double-opponency cells in LMS color space [43], sensor-independent learning [44], [45], and numerous others. Learning-based methods have much higher accuracy than statistics-based ones, but they are usually slower [46].

While the number of the proposed illumination estimation methods is ever-growing, there are not too many illumination estimation datasets and even the existing ones have various problems. These include too few images, inappropriate image quality, lack of scene diversity, multiple poorly synchronized versions of the same dataset, violation of various assumptions, etc. A high-quality illumination estimation dataset should be:

- *Diverse*. The more content and illumination cases are covered, the higher is the testing quality.
- *Large*. It is important that the datasets are not only diverse but that they also contain many images for each particular case. This makes it possible to notice quality improvement even for rare cases [47].
- *Informative*. Dataset should contain as much information about each captured image as possible. Precisely the information available during shooting procedure,
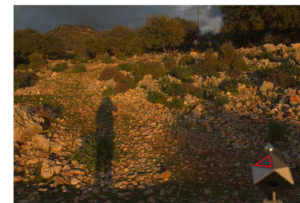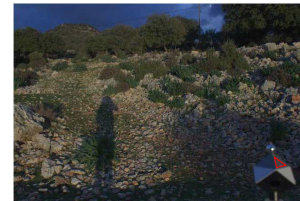
CUBE++
4890 images

META INFORMATION:

DAYTIME:	day
ILLUMINATION:	natural
LIGHT_OBJECTS:	sky
PLACE:	outdoor
SHADOWS:	true
RICHNESS:	rich
SHARPNESS:	sharp
KNOWN_OBJECTS:	false

FIGURE 1. Examples of chromatic adaptation based on two captured ground-truth illumination colors for an image from Cube++, a new large dataset where each image is accompanied by ground-truth illumination from several directions and semantic information about the scene content. This enables the research of single and multiple illumination scenarios as well as selection of images by various criteria.

meta-information about scene properties, information about light sources from different angles, etc.
- *Updatable*. Every illumination estimation dataset usually contains ground-truth illumination errors. Because of that, the dataset infrastructure should provide simple and reliable way for dataset debugging and tracking of its versions.
- *Verifiable*. From the previous point, it follows that the dataset should be available for verification, namely all provided markup and ground-truth can be collected and, if necessary, recreated by anyone who just downloads the source images.
- *Accessible*. The value of a dataset is decreasing when the downloading process is too complicated or time-consuming.
- *GDPR compliant*. Even a very good dataset can be of limited use for European researches if it is not compliant with GDPR, because it may prevent the researchers from publishing some of their results without breaking the regulations.

In this paper a new illumination estimation dataset named Cube++ with all of these properties is described. It contains 4890 images (see Fig. 2) carefully calibrated so as to get highly accurate ground-truth illumination. The images were collected in numerous countries, places, and illumination conditions. The countries in question include Austria, Croatia, Czechia, Georgia, Germany, Romania, Russia, Slovenia, Turkey, and Ukraine. In order to enable easy selection of images with specific properties, each image is accompanied by additional semantic information such as whether there are shadows in the image, whether it is an indoor or an outdoor image, whether the scene contains objects with known coloration, etc. An example of an image from Cube++ is shown in Fig. 1. The dataset is appropriate for different light source
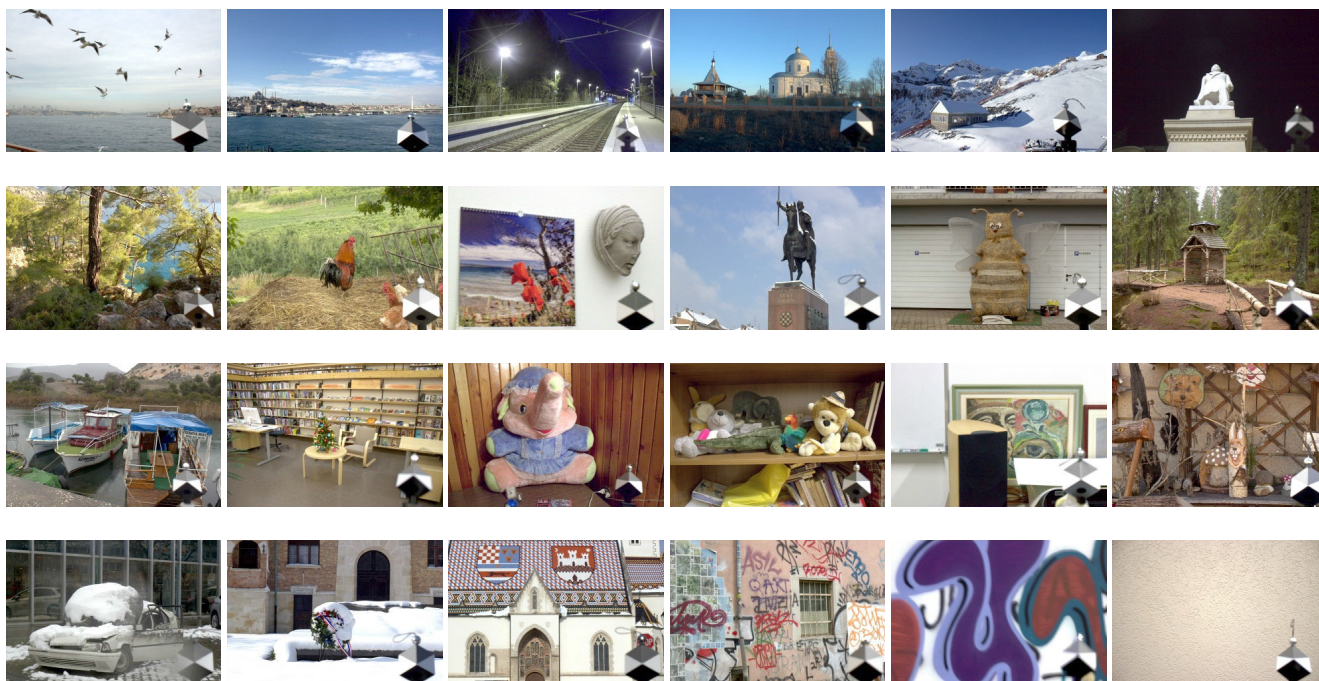
FIGURE 2. Example images from the newly created Cube++ dataset.

estimation use cases such as: single light source estimation, two light sources estimation, or estimation of at least one significant light source in the scene. Finally, some of the collected images were not included in the dataset and they are kept aside to be released later as part of a future illumination estimation benchmark somewhat similar to [48].

The paper is structured as follows: Section II describes the most important existing illumination estimation datasets and the problems associated with them, Section III gives the motivation for creating a new dataset, Section IV describes the methodology used to collect the dataset, Section V describes the newly proposed Cube++ dataset, Section VI presents a discussion about the scientific usefulness of contemporary datasets' form and about a potential improvement, and finally, Section VII concludes the paper.

II. INFLUENTIAL EXISTING DATASETS

One of the first illumination estimation datasets with a large number of images was the GreyBall dataset [49]. A gray ball was placed in the scene of each of the 11346 images to extract the ground-truth illumination. The main problem with this dataset is that the images are non-linearly processed and as such, they do not comply with the image formation model given in Eq. (1). Furthermore, the images in the GreyBall dataset are relatively small with a size of 240×360. Finally, the images were extracted from a video that was captured at several locations, which means that many of them have highly correlated illuminations and content. To cope with this problem of high redundancy, it has been proposed to use only a subset of 1135 images from GreyBall [50], [51].

In 2008, the ColorChecker dataset [25] with its 568 images was published and the ground-truth illumination was extracted by means of putting a color checker instead of a gray ball in the image scenes. This dataset was created by two different cameras and its images, which are individually bigger than the ones in the GreyBall dataset, also underwent non-linear processing, which means that similarly as with the GreyBall dataset they are given as 8-bit per-channel JPEG images.

In 2011, the reprocessed version of the ColorChecker dataset that contains only linearly processed images was published [52]. However, as observed already in 2013 [53], it was not mentioned clearly enough that the black level was supposed to be subtracted before using the images. Despite this observation, a lot of papers continued publishing results of methods obtained on the technically unprepared images with the black level included. This effectively led to the circulation of at least three versions of the ColorChecker datasets and the problem was formally addressed in [54] by also bringing into question the alleged advances in the illumination estimation research. In 2018, there was an attempt to rehabilitate the ColorChecker dataset by publishing the recalculated accuracy of various methods by using the allegedly correct ground-truth [55]. However, this attempt was marred by serious technical faults and wrong calculations that included comparing the estimations obtained on older versions to the new ground-truth, which only introduced further confusion [56]. This effectively opens the possibility of more future versions of the results on the ColorChecker dataset. In short, using the ColorChecker dataset can be very confusing and

problematic due to many circulating versions of the alleged results and consequent inappropriate comparisons, and therefore, to avoid problems, it should probably be omitted as the primary dataset choice.

In 2014, nine new NUS datasets with each of them taken by one of nine different cameras were published [18]. The images were only linearly processed, the black level subtraction was performed from the start in the initial paper, and the number of images was sufficiently high. The calibration object used to extract the ground-truth illumination was again a color checker. However, the problems with the NUS datasets include violations of uniform illumination assumption when having only a single ground-truth illumination, a relatively small number of images per camera with 268 being the maximum, having the same scenes repeated in images, and not being GDPR compliant as well as neither of the previous datasets.

In 2017, the Cube dataset [44] was published with 1365 images taken with a single camera and with a SpyderCube [1] calibration tool used for calibration. Due to its geometry that is superior to the one of a color checker, SpyderCube allows for easier detection of the presence of two illuminations and their extraction. This was extensively used to carefully calibrate each of the images of the Cube dataset and to obtain an accurate ground-truth. Special care was also taken to avoid the violation of the uniform illumination assumption as much as possible. The main drawback of the Cube dataset is that it contains only outdoor images, which also negatively affects the ground-truth illumination distribution. This drawback was alleviated in the Cube dataset's extension named the Cube+ dataset [44]. It contains 342 additional indoor images for a total of 1707 images and a wider span of ground-truth illumination distribution similar to the one in other datasets.

A relatively recent dataset is the INTEL-TAU dataset [57], a successor to the INTEL-TUT dataset [58], with 7022 images taken by three different cameras. While the number of images is sufficiently high, its main drawback is the fact that most of its images do not contain a calibration object in their scenes. Namely, it was removed after the initial calibration. Although this removes the requirement for masking it out, it also makes it impossible to reliably check and verify the ground-truth calibration and it is known that such errors occur [59]. Additionally, since the original raw image files are not provided, the EXIF data with the meta-information that may be important to some methods is also not available. The INTEL-TAU dataset is also completely GDPR compliant. Instead of avoiding problematic scenes, GDPR compliance was achieved by having "privacy masking applied on all sensitive information" such as "recognizable faces, license plates, and other privacy-sensitive information". The masking was performed so that "color component values inside the privacy masking area were averaged".

[1] https://www.datacolor.com/photography-design/product-overview/spydercube/

However, this effectively changes the original content and it may be undesirable in some cases.

A relatively recent dataset is the one for temporal color constancy [60], which contains 600 sequences of varying length between 3 and 17 frames. The dataset has not yet been made publicly available at the moment of writing this paper.

It is also important to mention that in contrast to all the described datasets that contain real-world images taken in mostly uncontrolled conditions, there are a lot of datasets made in fully controlled or even laboratory conditions, such as [6], [61]–[66].

The main advantage of the laboratory dataset is that it allows to research particular problem in fully-controlled conditions, but the variability of such datasets is often too low.

While other illumination estimation benchmark datasets also exist, it can be argued that the ones mentioned here are the most influential ones. They also share many problems with other existing datasets and thus their descriptions also cover most of the problems of other datasets. Some characteristics of the datasets mentioned here are summarized in Table 1.

III. MOTIVATION

After laying out the brief descriptions of some of the best-known illumination estimation benchmark datasets, it is possible to identify some of their main problems already recognized by the wider interested research community. Therefore, the motivation for creating a new illumination estimation is to try to reduce or entirely eliminate some of the mentioned problems of the existing datasets.

A. SIMPLE TECHNICAL FAULTS

Probably the most serious and most detrimental problem is the one connected to the technical shortfalls that can happen when creating and publishing a dataset. Some of the main such shortfalls are using non-linearly processed images and providing confusing information on black level subtraction.

As for the non-linearly processed images, the solution is to simply avoid performing non-linear processing and this can be simply carried out.

In the case of the black level subtraction, with the earlier datasets, this problem occurred due to a lack of explicit mentioning of the black level value in the papers that originally described these datasets. Additionally, in some cases, even a script that demonstrates the proper handling of the black level was either missing or put to a somewhat obscure location. In the case of the ColorChecker, such problems have led to multiple circulating versions of the ground-truth data and experimental results. Therefore, in the case of publishing a new dataset, such and similar problems motivate to clearly provide all necessary details on the required data for the black level subtraction and also to provide an example of how to do it.

TABLE 1. Characteristics of different illumination estimation datasets; the Cube++ dataset is described later in Section V.

Characteristic	GreyBall [49]	ColorChecker [52]	NUS [18]	Cube+ [44]	INTEL-TAU [57]	TCC [60]	Cube++
Number of images	11346 [2]	568	1736	1707	7022	approx. 10000 [2]	4890
Number of sensor types	1	2	8	1	3	1	1
GDPR compliance	-	-	-	-	✓	✓	✓
Undistored content	✓	✓	✓	✓	-	✓	✓
Color target in the scene	✓	✓	✓	✓	-	✓	✓
Color target type	gray ball	color checker (CC)	CC	SpyderCube	CC	SpyderCube	SpyderCube
Meta-info about the scene	-	-	-	-	-	-	✓
Night images	-	-	-	-	-	-	✓
Winter season	-	-	-	-	-	✓	✓
Year	2003	2020 [3]	2014	2019	2020	2020 [4]	2020
License	-	-	-	-	CC BY-SA 4.0	MIT	CC BY 4.0

B. RELIABLE GROUND-TRUTH

One of the probably least detectable technical faults with serious consequences is erroneous calibration and ground-truth illumination extraction. Based on the experience with existing datasets, it usually happens that there are multiple illuminations in the scene and the calibration object is under the influence of only one of them, which may not even be the dominant one. In that case, even if a method estimates the dominant illumination, it will be penalized because the ground-truth is based on another illumination. As mentioned earlier, this was already reported for the ColorChecker dataset.

To make the ground-truth reliable, one should use such a calibration objects that can detect the presence of multiple illuminations. Examples of such calibration objects include a gray ball such as in the GreyBall dataset or a SpyderCube instance such as in the Cube+ dataset, because they make it possible to simultaneously observe illuminations coming from different angles, and these can then be checked for any significant difference. An example of capturing two significantly different illuminations with a SpyderCube instance and showing the difference in how they affect color correction is given in [44]. If a significant difference is present, additional steps can be taken to either correctly determine which of the illuminations is the dominant one or to discard the image to prevent any future problems, which finally results in a correctly extracted and reliable ground-truth illumination.

C. VERIFIABLE GROUND-TRUTH

While the ground-truth should primarily be reliable, it should also be verifiable in order to add an additional layer of reliability. The simplest way of making the ground-truth of a dataset verifiable is to have all the dataset images contain a calibration objects in their scenes. In that way the ground-truth can easily be extracted by other researchers and then compared to the originally provided one to look for

potential errors. Additionally, the very visual information can help identify cases such as e.g. having the calibration object in a shadow while the majority of the scene is outside of that shadow.

D. CONTENT VARIETY

A new illumination estimation dataset should have a high content variety. While this seems rather obvious, it is not always put into practice to the full extent. For example, while the GrayBall dataset contains over 11k images, they are highly correlated and thus effectively not as rich in content as it may seem at first. In the case of datasets such as the ColorChecker dataset or the NUS datasets, all images were taken at the same geographical location and during the same season. None of the images there were taken e.g. during winter or at night. Such content choice restriction results in failure to cover many interesting and challenging environments that illumination estimation methods encounter in real-world applications and that should also be included in the research.

E. ILLUMINATION VARIETY

Having an appropriate ground-truth illumination variety in an illumination estimation dataset is important for several reasons. The most important one is to closely cover as much as possible of the illuminations that are encountered in the real-world applications because in that way the illumination estimation methods can be properly trained and tested.

An additional reason to have a sufficient ground-truth illumination variety is to prevent abuses of some often used error statistics that are possible if the ground-truth illumination are too clustered [67]. Such abuses can lead to false conclusions about the performance of the tested methods and consequently be detrimental for the research community and practitioners.

F. CHECKING FOR MULTIPLE ILLUMINATIONS

The majority of the illumination estimation datasets provide only a single ground-truth illumination per image. This effectively means that in terms of evaluation these datasets implicitly assume an uniform illumination. However, it is know that in illumination estimation datasets this is generally not the case [56]. As a matter of fact, any image with shadows has already effectively at least two illuminations that may differ significantly and this can also have a significant outcome on the later color correction step [44]. Additionally, even if there are no shadows, it is still possible for an image to be under the influence of multiple illuminations. In that case having a calibration object that is designed to successfully capture the illumination from only one direction at a time will fail to capture all the illuminations in the scene, let alone to detect their presence. Capturing only a single illumination when there are more present also leads to a problem during the evaluation. Namely, if a method correctly estimates one of the illuminations, but the other one is marked as the ground-truth, it may be argued that in this case the method is being unfairly judged. Because of that, an illumination estimation dataset should preferably use calibration objects that can simultaneously capture the illumination color from multiple directions. This would solve at least two problems. First, it would detect whether there are multiple illuminations in the first place, and second, if there really are multiple illuminations in the scene, then such a calibration object will capture more information on them. An example of such a calibration object is the SpyderCube object that has been described earlier.

G. NUMBER OF IMAGES

While some of the previous datasets with non-linearly processed images are obviously disadvantageous, some of them like the GreyBall dataset have the advantage of having thousands of images, which still makes them attractive to many researchers. Therefore, besides having a technically correct dataset, it is also important to make it have a sufficiently large number of images. This can result in both making the dataset desirable by offering a lot of useful data as well as simultaneously discouraging researches from using the inferior older datasets just because of their size. As for how large exactly a dataset should be, it should contain several thousands of images to outsize the existing datasets of lower quality and also to enable new breakthroughs. Finally, having a dataset with a large number of images is a prerequisite for achieving the previously mentioned content and illumination variety.

H. SEMANTIC DATA

In numerous cases, additional semantic information can be useful for research of specific illumination

---

estimation methods. For example, some of the methods may be interested in being explicitly trained only on indoor or outdoor images. Others may be interested in training images that contain no shadows whatsoever since they introduce additional illuminations. More generally, it may be useful to know whether there is a violation of the uniform illumination assumption on a given image. In such cases, it can be highly practical to be able to efficiently filter out images from a dataset based on some given criteria.

Because of that, a useful addition to a new illumination estimation benchmark dataset would be semantic information for each image. In that way, the research could be speeded up by not requiring researchers to label the images from scratch. Additionally, if such semantic information were given in advance, a lot of potential label mismatches between the labels created by different researchers could also be prevented.

I. PRIVACY CONCERNS

With the recent arrival of regulations such as the General Data Protection Regulation (GDPR), it becomes ever more important to respect privacy in publicly available images. This also means that using images from previous datasets with e.g. recognizable people or registration plates may nowadays be potentially seen as problematic. With respect to this, for the sake of respecting privacy, any new illumination estimation dataset should also take care of avoiding images that would contain any content that could compromise someones privacy.

On the other hand, if a public dataset is also supposed to be useful for development of methods that rely on e.g. faces [68], [69] or sclera [70], then it should obviously also contain images with faces. However, in that case it would be appropriate to obtain the consent for public use from the persons present in the image scenes. That would enable the researchers to use and show these images publicly in papers.

J. MULTIPLE INSTANCES OF THE SAME SENSOR CLASS

There can be significant differences between spectral characteristics of different sensors used by various cameras. This effectively means that a learning-based method that has been successfully trained on the images created by a camera of one model will not necessarily perform well on images created by a camera of another model without some adjustments. As a result, the problem of inter-camera color constancy has recently started to gain ever more attention [44], [45], [71]. Since almost every dataset was taken with another camera sensor, there is no shortage of training and testing images.

On the other hand, it is known and it can be shown that even for the instances of the same sensor class there are measurable differences in the spectral characteristics [72]. Hence, to check the significance of the impact of these differences on the accuracy of illumination estimation methods, an interesting feature of an illumination estimation dataset would be to have images created by several instances of the same sensor class. In addition to ground-truth illumination,

---

[2]Images are highly correlated. The GreyBall images are taken from 2 hours of video. The TCC images are taken from 600 video sequences.

[3]The cited version of ground-truth values were published in 2020, the original dataset was published in 2008.

[4]Expected in 2020, the data is not published at the time of writing.

such a dataset would also have to provide the sensor instance labels for each image.

IV. ACQUISITION METHODOLOGY

By identifying the problems with the existing datasets and describing some desired properties of the future datasets, the guidelines for creating a new illumination estimation dataset have been laid out. One of the main goals of this paper is not just to provide a theoretical framework, but also to create and propose a dataset by following these guidelines. The first step in doing so is to describe the used acquisition methodology.

A. TECHNICAL SETUP

1) COLOR TARGET (SpyderCube) CHARACTERISATION

As the calibration tool in the newly proposed dataset, the SpyderCube instances were used. SpyderCube is a color target for photographers whose main purpose is to help them to adjust the white balance manually. The general look of the SpyderCube is given in Fig. 3. A chrome ball is used to analyze specular highlights, the white on two faces is used to estimate true highlight value, the gray on two faces represents the midtone of the image and its color temperature, and the bottom black face is used to evaluate shadow values in the scene in relation to the black trap i.e. the hole, which represents absolute black.
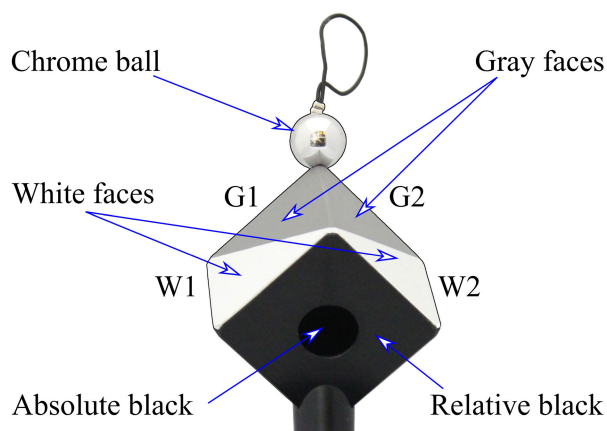

FIGURE 3. SpyderCube general look.

According to the manufacturer company Datacolor, the gray cube faces are neutral gray with a reflection coefficient of 18%.

Since SpyderCube is a relatively low-cost tool, some doubts about its declared optical properties could arise. To validate its properties, two SpyderCube instances, labeled SC1 and SC2, were compared. Individual faces of these SpyderCube instances were named G1, G2, and W1, W2, as shown in Fig. 3.

Reflection spectra of SpyderCube parts were measured using a Eye-One Pro spectrophotometer by X-Rite in the high-resolution mode of 3.3 nm with the help of the spotread
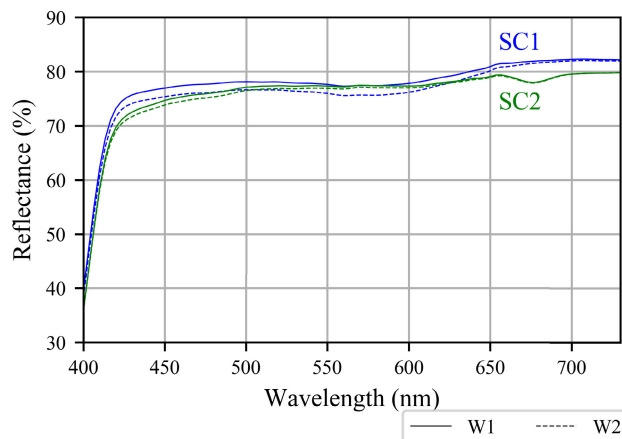

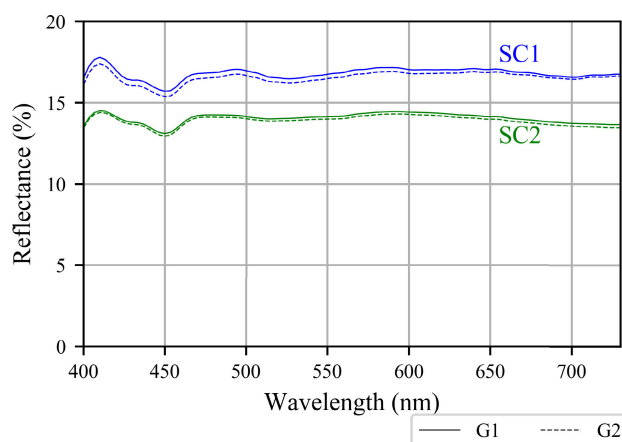**FIGURE 4. SC1 and SC2 white parts reflectance spectra.**

FIGURE 5. SC1 and SC2 gray parts reflectance spectra.

utility from Argyll CMS.[5] For each part, three measurements were made and the results were averaged. Figures 4 and 5 show the spectral reflection coefficients of the white and gray parts of the SC1 and SC2, respectively.

These measurements lead to the following observations:

- Gray parts of both SpyderCube instances are not "ideal" gray, i.e. the reflection spectra slightly depend on the wavelength. The sensitivity of the blue sensor in many cameras has a maximum at around 450 nm wavelength, and the reflection coefficients of gray parts G1 and G2 have a noticeable drop in the blue band.
- Each SpyderCube instance has small differences between reflection coefficients of its own gray parts G1 and G2.
- There are rather big differences between the gray parts reflection coefficients of the two measured SpyderCube instances.

---

[5]http://www.argyllcms.com/

- White parts of both SpyderCube instances are also not "ideal" white, i.e. the reflection spectra are not horizontal lines.
- Differences of the white parts W1 and W2 reflection coefficients of the both SpyderCube instances are small.

The idea behind SpyderCube as a calibration tool is that it does not distort the color of the illumination source, i.e. it is assumed to be "color neutral". From this point of view, what is important is the similarity between the shapes of the curves of reflection coefficients for the two SpyderCube instances and not the differences between the curves' values. From the measurements, it can be concluded that the curve shapes are indeed very similar. Therefore, the Spyder-Cube "color neutrality" assumption generally holds with one exception being the blue region of the spectrum for the white faces.

The degree of SpyderCube's color neutrality is one of the most important factors for accurate ground-truth extraction. The height of the grey reflection coefficient curve does not significantly influence the ground-truth extraction. Compared to other uncertainties, the measured deviations from color neutrality have only rather a small impact on the ground-truth.

Nevertheless, to measure the amount of this impact in terms of practical use, several images of two SpyderCube instances that were simultaneously in the same scene were captured with a Canon 600D camera under a D50-like illumination. The average difference between the ground-truth illuminations extracted from the faces of each SpyderCube instance and measured in terms of the angular reproduction error [73] was about 0.15°, which is in terms of color reproduction insignificant and invisible [74].

Performed measurements effectively demonstrate that using grey faces of different SpyderCube instances in different images has no significant effect on the overall ground-truth extraction quality. Still, SpyderCube quality should be studied in details also for all types of complicated artificial light sources (such as gas discharge lamp, etc.).

2) HANDHELD SETUP
To collect the dataset in natural conditions, the following equipment shown in Fig. 6 was used:

- Canon 550D camera or Canon 600D,
- SpyderCube calibration tool, and
- special attachment of the cube to the camera.

Special cube fasteners were built that allows the cube to be positioned so that it appears near the lower right corner of the image. The fasteners can also be rotated both in horizontal and vertical planes. The distance of the cube from the camera can be adjusted using a telescopic monopod and during the dataset, images capturing it was set to 50 cm. The experience gained while collecting the dataset images has led to the conclusion that the custom-built handheld setup is convenient to use.
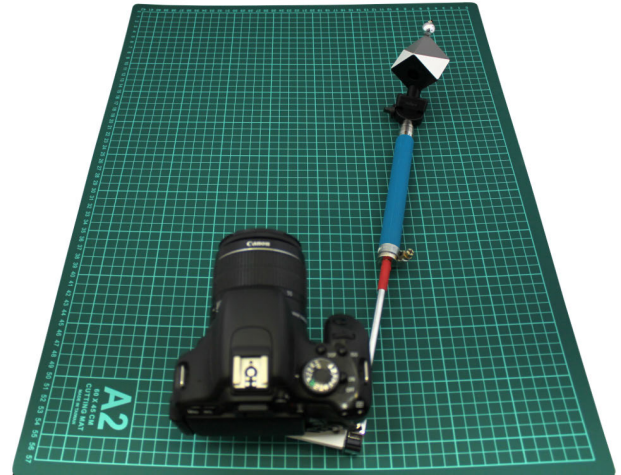

FIGURE 6. The general look of the handheld setup with Canon 600D camera.

(a)   (b)

(c)   (d)

FIGURE 7. Examples of images that should be excluded from the dataset: a) the color target is illuminated by the local lantern from the near shop, the color is different from the lighting of the most of scene; b) the color target is illuminated by sources that have almost no effect on the lighting of the observed scene; c) overexposed color target; d) the overly dark image.

B. DATA COLLECTION AND FILTRATION
The main thing to pay attention to during the image capturing was to assure that the used target cube and the majority of the observed scene are under the same illumination or illuminations. Examples of images with scenes where this requirement was not met are shown in Fig. 7.

Another significant factor that prevents accurate ground-truth extraction is the occurrence of glare on the color target. Images with this issue are usually characterized by clipping of the values in one of the color channels on the gray or white faces of the color target. The overexposure can be avoided in at least two ways: either by using manual camera settings or by specifying relative exposure compensation.

Dimming by one step usually turned out to be enough during the image collecting. Manual camera settings and one step lighting can also help to properly deal with the overly dark images. Examples of an overexposed and a too dark image are shown in Fig. 7c and Fig. 7d, respectively.

It should again be mentioned that there may often be several different illumination sources in one scene, commonly two, e.g., sun and sky or sky and streetlight. In this case, especially if the areas of the scene parts illuminated by different light sources are comparable, it is preferable to place the cube so that both illumination colors are captured by different cube faces. By doing so, it is later possible to simultaneously extract the illumination color of both influential scene light sources.

One of the main problems during image acquisition was to find the right position for the photographer to avoid differences between illuminations influencing the target cube and most of the observed scene. A lot of interesting scenes are available only in urban areas where there are a lot of different artificial illuminations. However, scenes in urban areas are usually full of different personal data like faces or plate numbers, which means that there are some difficulties related to GDPR. To make Cube++ GDPR compliant, the images with humans in the scene were filtered out and removed. This was done both automatically by using YOLOv3 [75] and manually by additionally checking each image.
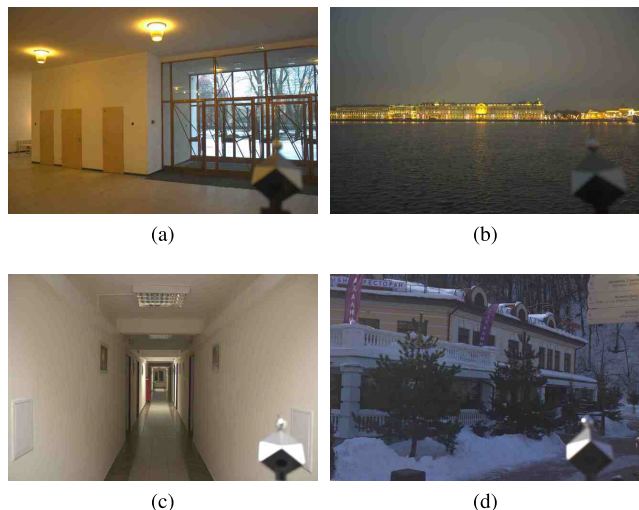

FIGURE 8. Examples of images with partial illumination estimation: a) and b) some scene parts are illuminated by the light source not captured by the cube; c) illumination significantly varies in the scene due to the interreflections d) all of the scene is in the shadow, while one of the cube faces is illuminated by the sun.

During the final quality filtration, all images were divided into three categories: a) images with **full light source estimation**, where the cube was illuminated by all the main light sources in the scene, b) **incorrect** images where the cube does not allow to determine the illumination rating consistent with the scene, and c) the rest i.e. images with **partial light source estimation**, see Fig. 8.

As a result, about 400 images were marked as incorrect and removed from the dataset, about 524 images were marked as difficult images with a partial light source estimation, and the rest of the images were marked as good.

Additionally, the fiber on the top of the cube may fall on a cube or remain on the image after cropping out the color target. To prevent it, the fibers were glued to the cube or just cut off on most images. All of the images are captured horizontally without the use of a camera flash.

C. GROUND-TRUTH EXTRACTION

The ground-truth extraction was performed on raw images. First, a simple debayering has been performed by transforming each RGGB Bayer pattern square into a single pixel. The red and blue channels of the pixel color were obtained directly from the R and B components of the pattern, while the green was obtained by averaging the two G values. No interpolation was performed and therefore the number of image rows and columns was halved. Next, the oversaturated pixels were masked out, and then the black level of 2^{11} was subtracted from all pixels. Finally, the ground-truth illumination values were extracted by calculating the average chromaticity of the manually annotated areas of the SpyderCube triangles.

Four chromaticities were calculated for every image. They correspond to white and gray triangles on the left and right cube faces. Note that on the brightly illuminated cubes, the white triangles may have oversaturated areas that cannot be properly used. On the contrary, the gray triangles chromaticities on the darker images may not be stable due to the black level noise. It is important to note no image contains saturated grey edges, while some of the images contain saturated white edges and in such cases, a corresponding mark is provided.

The illuminations for a triangle were calculated as the mean illumination of its area after 50% downscale to the barycenter. The value of 50% is selected as a simple empirical trade-off. Namely, a full-size triangle may contain non-triangle pixels because of unfocused cube or markup inaccuracies, while a tiny triangle would contain too few pixels and would be affected by noise.

D. SEMANTIC MARKUP

When developing and testing an algorithm for illumination estimation in a scene, it is useful to be able to analyze the structure of errors. The average error over the entire dataset will often not help to reveal whether e.g. the accuracy of the method for indoor images is much less accurate than for outdoor images. To enable performing such and similar checks faster and easier, additional information about the scene and shooting conditions were added to each image in the dataset. In addition to the information available during the shooting, this also includes the following manual annotation:

- Time of day (field **daytime**, with values day/night/unknown).

- The presence of objects with known coloration (**has_known_objects** field with values true/false).
- Scene illumination type (**illumination** field with values artificial/natural/unknown). It is worth noting that there are no flash photos in the dataset.
- Image sharpness (**is_sharp** field with values true/false).
- The presence of light sources in the scene (the **light_objects** multiple choice field with the values lamp/sky/sun/none).
- The place where the image was captured (the **place** field with the values outdoor/indoor/unknown).
- Scene richness (field **richness** with values rich/simple).
- The presence of shadows in the scene (**shadows** field with values yes/no/unknown)
- The cube illumination by all the main light sources in the scene (**estimation** field with values full/partial)

Finally, it is important to note that none of the fields had a preset default value. In that way, the value of every field had to be explicitly set by an annotating person. Namely, if some default field values were to be set in advance, it could increase the annotation bias.

V. THE PROPOSED DATASET

Having in mind all of the concerns and motivation from the previous section, a new dataset named Cube++ is proposed that continues on the previous Cube+ dataset. The dataset download link, the accompanying code, and the technical file description are available at https://github.com/Visillect/CubePlusPlus/.

The Cube++ dataset contains 4890 images. It includes only 1359 of the 1707 images from the Cube+ dataset and only 330 of the 363 images from the 1st Illumination Estimation Challenge (IEC#1) test set [76]. Other images were excluded because they may go against respecting privacy by containing personal data such as faces and license plates or they may be problematic for ground-truth extraction. The remaining 3201 images are brand new.

Cube++ has diverse scene illumination cases as demonstrated by Fig 9. There it can be seen that the chromaticity coverage area is wider than in e.g. Cube+. In other words, the illumination variability has been significantly improved.

The ground-truth illumination distribution for Cube++ and its parts can also be seen from another point of view by taking a look at Fig. 10.

This figure shows that Cube+, which includes Cube, and Cube++ have somewhat similar distributions, which in turn means that a lot of images were taken under outdoor daylight illumination.

One of the important features of the proposed dataset is the fact that it contains two ground-truth illumination records per image, one for each side of the SpyderCube instance. Even though in many of the images there is effectively only one dominant illumination in the scene, Fig. 11 helps to better understand the relation between the two recorded illuminations over the dataset images. Currently, the average angular
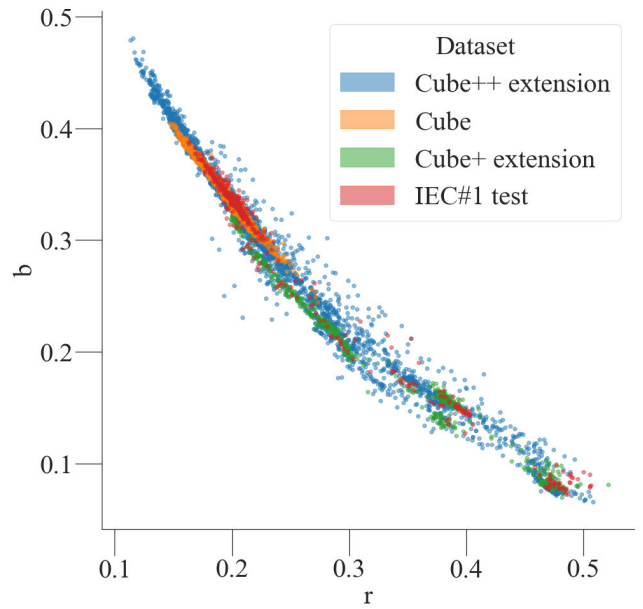

FIGURE 9. Scatter plot of ground truth illumination chromaticities captured by the SpyderCube gray faces.
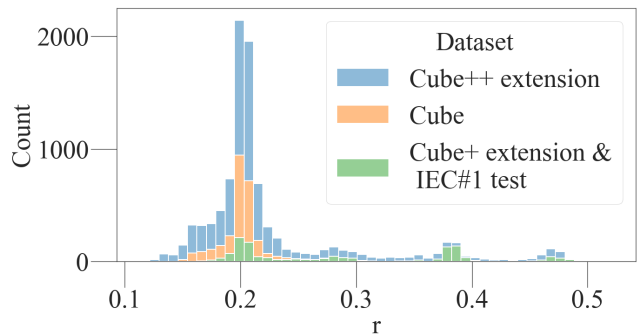
FIGURE 10. Stacked histogram of the red chromaticity values $r = R/(R + G + B)$ of Cube++ ground-truth illuminations.
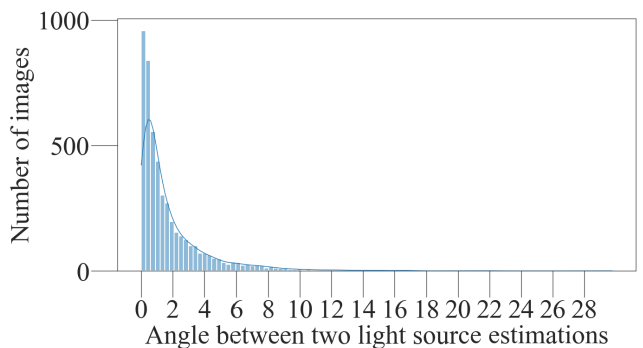
FIGURE 11. Histogram of angular differences between SpyderCube's left and right gray faces for Cube++.

error of state-of-the-art illumination estimation methods is arguably somewhere between 1° and 2°. With that in mind, all images with larger angular difference between their illumination records can be treated as two-illumination cases.

TABLE 2. Feature statistics for various Cube++ subsets.

| Feature		Cube++ parts					Cube++ subsets	
		$Cube^8$	$Cube+^8$ extension	$IEC\#1^8$ test	Cube++ extension	**Total Cube++**	Full estimation	**Simple Cube++**
estimation	full	916	306	301	2843	**4366**	4366	**2234**
	partial	115	22	29	358	**524**	0	**0**
daytime	day	1007	24	176	2302	**3509**	3155	**1615**
	night	1	1	1	71	**74**	36	**5**
	unknown	23	303	153	828	**1307**	1175	**614**
place	indoor	12	281	27	454	**774**	653	**379**
	outdoor	1017	46	217	2536	**3816**	3432	**1738**
	unknown	2	1	86	211	**300**	281	**117**
illumination	artificial	2	9	16	306	**333**	233	**109**
	mixed	0	0	1	27	**28**	13	**7**
	natural	995	41	164	2311	**3511**	3171	**1611**
	unknown	34	278	149	557	**1018**	949	**507**
light_objects	none	560	326	287	1505	**2678**	2434	**1246**
	flash	0	0	0	3	**3**	3	**2**
	lamp	1	0	1	125	**127**	58	**29**
	lamp&sky	0	0	0	21	**21**	7	**5**
	sky	470	2	42	1543	**2057**	1861	**952**
	sky&sun	0	0	0	3	**3**	3	**0**
	sun	0	0	0	1	**1**	0	**0**
shadows	yes	264	203	61	1282	**1810**	1573	**696**
	no	767	125	259	1879	**3030**	2751	**1519**
	unknown	0	0	10	40	**50**	42	**19**
richness	simple	3	8	94	337	**442**	400	**198**
	rich	932	152	233	2840	**4157**	3704	**1860**
	unknown	96	168	3	24	**291**	262	**176**
has_known_objects	True	767	49	76	1555	**2447**	2185	**1138**
	False	264	279	254	1646	**2443**	2181	**1096**
is_sharp	True	1002	259	309	3006	**4576**	4088	**2088**
	False	29	69	21	195	**314**	278	**146**
Total		1031	328	330	3201	**4890**	4366	**2234**

Another important feature of the proposed dataset that has to be stressed additionally is that it contains semantic data for each image. All semantic information features are shown in Table 2. Different features provided in the semantic data can be helpful for algorithm tuning and profiling as they give potentially useful information about each image individually.

A. TECHNICAL DESCRIPTION

The dataset consists of several parts. First, there are the raw images with only simple linear debayering performed that are stored in 16-bit PNG format. Next, there are CSV files with ground-truth illumination values and CSV files with additional related properties. Furthermore, there are JPEG images generated by using the dcraw open-source tool.[7] Finally, there are also additional files for storing auxiliary information. All these files are automatically built from sources included in the dataset by running a script that is also

[7] https://www.dechifro.org/dcraw/

provided. The sources contain the original CR2 images from the camera and JSON files with the manual annotation data.

The original camera JPEG images are not included as their generation depends on cameras' settings, which means that they cannot be recreated simply or even accurately [77].

1) PNG AND JPEG IMAGES

The main 16-bit PNG images are generated from the original CR2 files in three steps. First, the CR2 files are decoded by using the dcraw tool with the options $-D$ -4 $-T$. This generates a 16-bit 1-channel TIFF image. Second, the $[10, 10 + 5184] \times [4, 4 + 3456]$ rectangle was cropped, to have the same area as the default camera JPEG, which comes with certain advantages. Finally, a naive debayering is applied so that every $\begin{smallmatrix} R, & G_1, \\ G_2, & B \end{smallmatrix}$ pattern is converted to a pixel of color $\left(R, \frac{G_1+G_2}{2}, B\right)$. After that the size of the generated PNG images is $2592 \times 1728 = 2^5 3^4 \times 2^6 3^3$. Even though the color channel values have 16 bits of storage, in practice their

maximal value is always below $2^{14} - 1$. The black level that can be used for every dataset image is $2048 = 2^{11}$.

For visualization purposes, the modified versions of JPEG images generated by the `dcraw` tool are included as well. The modification includes cropping and downscaling in order for the JPEG images to have the same size as the PNG images. Downscaling is required because JPEG images generated by `dcraw` are not downscaled like the PNG images. On the other hand, JPEG images generated by the camera were not included because they depend on camera settings and the camera's white balancing algorithm, which is proprietary, not fully documented, and it may differ for Canon 550D and 600D cameras that have been used for image capturing. Because of that, they can not be recreated reliably.

2) THE GROUND-TRUTH

The ground-truth illumination records are stored in the `gt.csv` file. Ground-truth illuminations are calculated as described in Section IV-C. The columns are: *image* and for each of the 4 triangles (left, right, left bottom, right bottom) it contains three columns *r, g, b* with the corresponding RGB illumination chromaticities so that $r + g + b = 1$. The triangles brightness values are given in the `properties.csv` file.

Usually, computational color constancy datasets contain only a single ground-truth illumination vector, which represents the dominant illumination in the scene. In the Cube++ dataset such illumination is not given, because the precise single illuminant estimation may require specialist annotation. Moreover, some images have two significantly different illuminations, which makes it harder to select the dominant one. If only a single ground-truth illumination is required and the possible errors that it leads to are acceptable, then one of the following methods can be used to obtain it:

- sample images with relatively similar left and right ground-truth illuminations (the sugested answers for such images are denoted in *properties.csv*);
- select from the left and right sides the brighter one; then select the white triangle for a dark image, and grey triangle for a bright image.

Note that the difference between the sensitivities of the white faces is greater then the difference between the sensitivities of the gray ones (see Section IV-A). Additionally, since the white faces are more often overexposed than the gray ones, using the gray faces should be preferred. On the contrary, using white faces may be better on dark images as mentioned in Section IV-C.

We also estimated if the ground truth values are distorted by the pixels with the clipped values. The images with overexposed grey triangles were removed from the dataset. The images with the clipped values on white triangles are present in the dataset, but the overexposed triangles are marked in *properties.csv*.

3) RELEVANT META-INFORMATION

The `properties.csv` file contains the most relevant meta-information about the Cube++ images. It includes the average triangle brightness $\frac{R+B+G}{3}$, manual annotation data, information about overexposed triangles, and a carefully selected subset of EXIF data fields.

The EXIF data was extracted from CR2 files using the `PyExifTool` library.[8] All the extracted values can be found in the corresponding JSON files. The properties table contains only a few selected ones. The EXIF data format slightly differs between the Canon EOS 550D and 600D cameras: there are 312 common fields, 2 in 550D only, and 21 in 600D only. All the selected EXIF fields are common.

The `cam_estimation.csv` file contains the EXIF fields of the camera that contains the camera's light source estimation

B. IMAGE PREPARATION

Finally, it is important to clearly specify how to properly prepare the provided Cube++ images before handing them over to illumination estimation methods that are to be tested.

There are three main steps that have to be taken.

1) BLACK LEVEL SUBSTRACTION

The first step is to subtract the approximate black level of 2048 from all image pixel color components. In some cases this can result in negative values, but such values should then be set to 0.

2) SATURATION DETECTION

The second step is to calculate the maximum value *m* for all pixels across all color channels. After that all pixels that have a value greater than or equal to $m - 50$ in any of their channels should have all their channel values set to 0 This would remove most of the incorrect pixels with clipped values. Nevertheless, it would leave some rare overexposed pixels, because demosaicing procedure may mix them with the normal ones. To get precise information about saturated pixels it is recommended to analyse images before demosaicing (the last one can be extracted from CR2 files).

3) COLOR TARGET MASKING

The last step is to mask out the lower right rectangle of the image that contains the color target to remove any potential bias and thus to have a relatively fair testing. The size of this rectangle is 700×1000 for all images. The rectangle is masked out by setting all channel values of all its pixels to 0, i.e. by making it black.

C. INTENDED DATASET USAGE

With all its features, especially the two ground-truth illumination records, Cube++ is appropriate for several illumination estimation use cases. All datasets mentioned in Section II, except for maybe TCC dataset, are designed for the most widely used classical illumination estimation problem: estimation of the single light source in the scene. Therefore, each image is provided with only single light source ground-truth,

---

[8]https://pypi.org/project/PyExifTool/

even in cases when the scene is obviously under the influence of multiple illuminations. In contrast to these datasets, Cube++ allows to work on following problem statements:

1) Estimate one and only dominant lighting in a scene;
2) Estimate two dominant light sources in a scene;
3) Estimate at least one dominant light source in the scene.

For each of the listed problem statements we propose the following rules to filter the Cube++ images that are suitable for it. To form the dataset subset for the first problem, one needs to select all images where the angular differences between its two extracted ground-truth illuminations is below 1° (except partially light source estimation part, see section IV-B). For the two light source estimation problem, one needs to do the opposite, i.e. to select all images that are not selected for the first problem (except partially light source estimation part). Finally, to work on the third problem, all images can be chosen.

FIGURE 12. Example of chromatic adaptation based on illumination extracted from a) left and b) right gray face of the SpyderCube calibration object placed in the scene.

Here it is important to mention that the proposed rules are arbitrary, that they will result in some of the images being inappropriately selected, and that they may be improved. One example where these rules fail is shown in Fig. 12. There the extracted ground-truth illuminations differ significantly, but practically the whole scene is mostly under the illumination captured by the right cube face. This means that even though the scene is effectively under uniform illumination, the mentioned rules will result in the opposite conclusion based on the difference between the extracted illuminations on the two cube faces.

Also it is important to mention what at the moment partially light source estimation part of the dataset is not provided with subjective single illumination estimation choice. The plan is to solve this in future work.

D. SimpleCube++ DATASET
In addition to the main 200GB Cube++ dataset, a 2GB-small and simpler version of it is prepared. The small dataset contains 4x downscaled images that have less than 1° difference between ground-truth illuminations of SpyderCube's left and right grey faces. It includes only images with a single illumination source, and consequently, the ground-truth file contains only one ground-truth per image. This ground-truth was extracted in the following manner: firstly, average values for both gray faces were calculated as in the main

Cube++ dataset; secondly, they were normalized by using l_1-norm ($r + g + b = 1$ for both gray faces); finally, obtained ground-truth values were averaged and normalized again by using l_1-norm. This dataset has two main advantages: small weight (around 2GB) and a single answer per image.

SimpleCube++ contains PNG and JPG files, `gt.csv` with ground-truth data, and `properties.csv` with manual annotation data. In addition, this dataset was divided into train and test parts. Each image was independently assigned to the test set with a probability of 20%.

VI. DISCUSSION
Having another high-quality illumination estimation dataset such as the one proposed in this paper is certainly beneficial to the interested research community as well as the industrial sector and there should probably be no discussion about that. However, proposing a new dataset is still only an incremental move in terms of the overall paradigm of illumination estimation research since this has been done on numerous occasions while the dataset usage has remained relatively unchanged.

A much more constructive and necessary discussion that is rarely taken forward should be about the direction of how to better use or not use the datasets to achieve better progress in illumination estimation research. In terms of that, one of the burning issues is that the results in most illumination estimation papers are unverifiable and thus questionable. Therefore, for the sake of improving the state of the illumination estimation research, it would be quite useful to further discuss this problem as well as the potential solutions to it in more detail.

A. QUESTIONABLE RESEARCH PROGRESS
Obtaining low illumination estimation errors on a benchmark dataset is a regularly used approach when trying to demonstrate the superiority of a proposed illumination estimation method. For all well-known datasets the ground-truth illumination used during the test phase is publicly available and the actual error statistics calculation is usually performed by the authors themselves and published in their papers. However, this introduces several problems with the most serious being data dredging, i.e. *p*-hacking and erroneous reporting.

The problem with data dredging in illumination estimation is that in cases when a model selection is required, the final results that are reported were not always obtained through nested cross-validation [78]. Instead, the reported results are the ones that were used to select the model in the first place. By using these results, a method's true performance on new unknown data may be masked and unfairly shown to be better than it actually is. This can prevent or slow down the progress in illumination estimation research by giving misleading clues about the validity of the method's underlying assumptions.

In the area of visual odometry similar problems with e.g. the KITTI dataset [48] have been prevented by simply

---

[8]Not including Cube, Cube+, IEC#1 test part images, removed from the Cube++ dataset because of GDPR restrictions or possible problems with ground-truth extraction.

keeping the ground-truth for the test secret. By having the evaluation of the results on the test set carried out by the dataset administrators, any serious attempts of *p*-hacking have been prevented.

Another problem that can be prevented if the evaluation is carried out by a third party is erroneous reporting. For example, in [18] the results of the proposed illumination estimation method on several datasets were allegedly all obtained by using the same value of a hyperparameter. However, trying to re-implement the method fails to produce the same results and only after checking the associated webpage [79] it becomes clear that the hyperparameter value has to be changed for each dataset to fully reproduce the published results.

A somewhat similar example is the 2007 paper by van de Weijer *et al.* [23]. In an erratum published in 2008 [80] it was explained how testing was inadequately performed, which consequently resulted in reporting of erroneous error statistics.

Finally, any doubts in the validity of some reported error statistics could be reduced or fully eliminated if they were calculated not by the authors themselves, but by a reliable third party. This would also help the overall research progress.

B. ILLUMINATION ESTIMATION CHALLENGES

Inspired by the ideas mentioned in the previous subsection, two international illumination estimation challenges have already taken place [76], [81]. The challenges provided the participants with thousands of training images and their respective ground-truth illuminations, while for the test set only the images were provided and the ground-truth remained secret until the end of the challenge. Because of that the error statistics for the illumination estimations sent over by the authors were calculated by the challenge organizers, which prevented a lot of problems described in the previous section. The results were thus more trustworthy and they have shown e.g. high errors for some methods that were previously reported to be highly accurate. Additionally, the challenges helped to recognize additional problems such as training a method to obtain excellent values for a given error metric [82], which results in issues related to the so called Goodhart's law [83].

C. BENCHMARK

While the described international illumination estimation challenges have shown the advantages of having a reliable third-party calculate the error statistics, they were fixed in time and they cannot be repeated on the same images anymore. Therefore, the next step would be to create a benchmark dataset similar to the KITTI dataset with an online user interface for submitting the results at any given time. This would surely represent a significant contribution to the illumination estimation research since it would simultaneously provide the researchers with trustworthy results and also eliminate many of the serious problems that were described earlier in this paper.

For the above reasons, creating such a benchmark is already underway at the time of writing this paper. At present time we are working on the question of benchmark creation. Possible benchmark will be based on the images that were taken during the same time as the rest of the Cube++ images, but that were excluded from its final version. Because of that, in this paper there are purposely no error statistics obtained on the Cube++ dataset by any of the illumination estimation methods. The error statistics will be published online and they will be based on the first version of the benchmark test set. This aims to avoid providing any results obtained on the Cube++ images with known ground-truth illumination. Namely, the idea is to separate the testing and the associated problems from the dataset and to relegate it to the benchmark. Therefore, the overall goal of this paper is to provide high quality training data without any testing. The role of testing data is to be assumed by the future benchmark.

VII. CONCLUSION

A new illumination estimation dataset named Cube++ has been proposed. Unlike similar existing illumination estimation datasets, it provides rich, reliable, and verifiable data on scene illumination with specific care being given to precise calibration. For every one of its 4890 images, there are two ground-truth illumination records as well as a multitude of semantic information and it is GDPR-compliant. Furthermore, a wide variety of scene content is covered, and numerous illuminations are captured. Cube++ contains images taken with several instances of the same model of the camera sensor. In addition to that, a centralized versioning control system for Cube++ has been established to simplify and document possible future changes in the dataset and error handling. By having these properties and novelties, Cube++ is technically superior to most similar illumination estimation datasets. One of the future steps that should also be significant progress in the overall illumination estimation research is to create an online illumination estimation benchmark based on the infrastructure that was used to create the Cube++ dataset.

REFERENCES

[1] M. Ebner, *Color Constancy* (IS&T Series in Imaging Science and Technology). Hoboken, NJ, USA: Wiley, 2007.

[2] G. D. Finlayson, M. S. Drew, and B. V. Funt, "Color constancy: Enhancing von kries adaption via sensor transformations," in *Proc. 4th Hum. Vis., Vis. Process., Digit. Display, Int. Soc. Opt. Photon.*, vol. 1913, Sep. 1993, pp. 473–484.

[3] D. Cheng, B. Price, S. Cohen, and M. S. Brown, "Beyond white: Ground truth colors for color constancy correction," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 298–306.

[4] D. P. Nikolaev and P. P. Nikolayev, "On spectral models and colour constancy clues," in *Proc. 21st Eur. Conf. Modeling Simulation (ECMS)*, Prague, Czech Republic: Citeseer, 2007, pp. 318–323.

[5] P. P. Nikolaev, "Gybridnaya spectralnaya model v zadache constantnosty tsveta. 1. Zonalniye okraski pri gaussovskom osveshenii," *Sensory Syst.*, vol. 22, no. 4, pp. 287–308, 2008.

[6] K. Barnard, V. Cardei, and B. Funt, "A comparison of computational color constancy algorithms. I: Methodology and experiments with synthesized data," *IEEE Trans. Image Process.*, vol. 11, no. 9, pp. 972–984, Sep. 2002.

[7] E. H. Land, "The retinex theory of color vision," *Sci. Amer.*, vol. 237, no. 6, pp. 108–129, 1977.

[8] B. Funt and L. Shi, "The rehabilitation of MaxRGB," in *Proc. Color Imag. Conf., Soc. Imag. Sci. Technol.*, no. 1, 2010, pp. 256–259.

[9] N. Banic and S. Loncaric, "Using the random sprays Retinex algorithm for global illumination estimation," in *Proc. 2nd Croatian Comput. Vis. Workshopn (CCVW)*. Zagreb, Croatia: Univ. of Zagreb, Faculty of Electrical Engineering and Computing, Sep. 2014, pp. 3–7.

[10] N. Banic and S. Loncaric, "Color rabbit: Guiding the distance of local maximums in illumination estimation," in *Proc. 19th Int. Conf. Digit. Signal Process.*, Aug. 2014, pp. 345–350.

[11] N. Banic and S. Loncaric, "Improving the white patch method by subsampling," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 605–609.

[12] G. Buchsbaum, "A spatial processor model for object colour perception," *J. Franklin Inst.*, vol. 310, no. 1, pp. 1–26, Jul. 1980.

[13] G. D. Finlayson and E. Trezzi, "Shades of gray and colour constancy," in *Proc. Color Imag. Conf., Soc. Imag. Sci. Technol.*, no. 1, 2004, pp. 37–41.

[14] J. van de Weijer, T. Gevers, and A. Gijsenij, "Edge-based color constancy," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2207–2214, Sep. 2007.

[15] A. Gijsenij, T. Gevers, and J. van de Weijer, "Improving color constancy by photometric edge weighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 5, pp. 918–929, May 2012.

[16] H. R. V. Joze, M. S. Drew, G. D. Finlayson, and P. A. T. Rey, "The role of bright pixels in illumination estimation," in *Proc. Color Imag. Conf., Soc. Imag. Sci. Technol.*, no. 1, 2012, pp. 41–46.

[17] Y. Qian, S. Pertuz, J. Nikkanen, J.-K. Kämäräinen, and J. Matas, "Revisiting gray pixel for statistical illumination estimation," in *Proc. 14th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl. (VISAPP)*, 2019, pp. 36–46.

[18] D. Cheng, D. K. Prasad, and M. S. Brown, "Illuminant estimation for color constancy: Why spatial-domain methods work and the role of the color distribution," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 31, no. 5, pp. 1049–1058, 2014.

[19] N. Banic and S. Loncaric, "Blue shift assumption: Improving illumination estimation accuracy for single image from unknown source," in *Proc. 14th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl. (VISAPP)*, 2019, pp. 191–197.

[20] N. Banic and S. Loncaric, "Green stability assumption: Unsupervised learning for statistics-based illumination estimation," *J. Imag.*, vol. 4, no. 11, p. 127, Oct. 2018.

[21] D. A. Tarasov and O. B. Milder, "Mathematics and practice of color space invariants by the example of determining the gray balance for a digital printing system," *Comput. Opt.*, vol. 44, no. 1, pp. 117–126, Feb. 2020.

[22] V. C. Cardei, B. Funt, and K. Barnard, "Estimating the scene illumination chromaticity by using a neural network," *JOSA A*, vol. 19, no. 12, pp. 2374–2386, 2002.

[23] J. van de Weijer, C. Schmid, and J. Verbeek, "Using High-Level Visual Information for Color Constancy," in *Proc. IEEE 11th Int. Conf. Comput. Vis. (ICCV)*, Oct. 2007, pp. 1–8.

[24] A. Gijsenij and T. Gevers, "Color constancy using natural image statistics," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[25] P. V. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp, "Bayesian color constancy revisited," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[26] D. Hernandez-Juarez, S. Parisot, B. Busam, A. Leonardis, G. Slabaugh, and S. McDonagh, "A multi-hypothesis approach to color constancy," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2270–2280.

[27] A. Chakrabarti, K. Hirakawa, and T. Zickler, "Color constancy with spatio-spectral statistics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 8, pp. 1509–1519, Aug. 2012.

[28] N. Banic and S. Loncaric, "Color cat: Remembering colors for illumination estimation," *IEEE Signal Process. Lett.*, vol. 22, no. 6, pp. 651–655, Jun. 2015.

[29] N. Banic and S. Loncaric, "Using the red chromaticity for illumination estimation," in *Proc. 9th Int. Symp. Image Signal Process. Anal. (ISPA)*, Sep. 2015, pp. 131–136.

[30] N. Banic and S. Loncaric, "Color Dog–Guiding the global illumination estimation to better accuracy," in *Proc. 10th Int. Conf. Comput. Vis. Theory Appl. (VISAPP)*, 2015, pp. 129–135.

[31] G. D. Finlayson, "Corrected-moment illuminant estimation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1904–1911.

[32] D. Cheng, B. Price, S. Cohen, and M. S. Brown, "Effective learning-based illuminant estimation using simple features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1000–1008.

[33] J. T. Barron, "Convolutional color constancy," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 379–387.

[34] J. T. Barron and Y.-T. Tsai, "Fast Fourier color constancy," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jul. 2017, pp. 886–894.

[35] S. Bianco, C. Cusano, and R. Schettini, "Color constancy using CNNs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 81–89.

[36] W. Shi, C. C. Loy, and X. Tang, "Deep specialized network for illuminant estimation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 371–387.

[37] Y. Hu, B. Wang, and S. Lin, "FC4: Fully convolutional color constancy with confidence-weighted pooling," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4085–4094.

[38] S. W. Oh and S. J. Kim, "Approaching the computational color constancy as a classification problem through deep learning," *Pattern Recognit.*, vol. 61, pp. 405–416, Jan. 2017.

[39] K. Koščevic, N. Banic, and S. Loncaric, "Color beaver: Bounding illumination estimations for higher accuracy," in *Proc. 14th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2019, pp. 183–190.

[40] A. Akbarinia and C. A. Parraga, "Colour constancy beyond the classical receptive field," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 9, pp. 2081–2094, Sep. 2018.

[41] S.-M. Woo, S.-H. Lee, J.-S. Yoo, and J.-O. Kim, "Improving color constancy in an ambient light environment using the phong reflection model," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1862–1877, Apr. 2018.

[42] K.-F. Yang, S.-B. Gao, and Y.-J. Li, "Efficient illuminant estimation for color constancy using grey pixels," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2254–2263.

[43] S.-B. Gao, K.-F. Yang, C.-Y. Li, and Y.-J. Li, "Color constancy using double-opponency," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 10, pp. 1973–1985, Oct. 2015.

[44] N. Banic and S. Loncaric, "Unsupervised learning for color constancy," in *Proc. 13th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl. (VISAPP)*, 2018, pp. 181–188.

[45] M. Afifi and M. S. Brown, "Sensor-independent illumination estimation for DNN models," 2019, *arXiv:1912.06888*. [Online]. Available: http://arxiv.org/abs/1912.06888

[46] A. Gijsenij, T. Gevers, and J. van de Weijer, "Computational color constancy: Survey and experiments," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2475–2489, Sep. 2011.

[47] J. T. Barron, *Personcal Correspondence*. 2019.

[48] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3354–3361.

[49] F. Ciurea and B. Funt, "A large image database for color constancy research," in *Proc. Soc. Imag. Sci. Technol., Color Imag. Conf.*, no. 1, 2003, pp. 160–164.

[50] S. Bianco, G. Ciocca, C. Cusano, and R. Schettini, "Improving color constancy using indoor–outdoor image classification," *IEEE Trans. Image Process.*, vol. 17, no. 12, pp. 2381–2392, Dec. 2008.

[51] S. Bianco, G. Ciocca, C. Cusano, and R. Schettini, "Automatic color constancy algorithm selection and combination," *Pattern Recognit.*, vol. 43, no. 3, pp. 695–705, Mar. 2010.

[52] B. F. L. Shi. (May 2020) *Re-Processed Version of the Gehler Color Constancy Dataset of 568 Images*. [Online]. Available: https://www2.cs.sfu.ca/~colour/data/shi_gehler/

[53] S. E. Lynch, M. S. Drew, and G. D. Finlayson, "Colour constancy from both sides of the shadow edge," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Dec. 2013, pp. 899–906.

[54] G. D. Finlayson, G. Hemrit, A. Gijsenij, and P. Gehler, "A Curious Problem with Using the Colour Checker Dataset for Illuminant Estimation," in *Proc. Color Imag. Conf., Soc. Imag. Sci. Technol.*, no. 25, 2017, pp. 64–69.

[55] G. Hemrit, G. D. Finlayson, A. Gijsenij, P. Gehler, S. Bianco, B. Funt, M. Drew, and L. Shi, "Rehabilitating the colorchecker dataset for illuminant estimation," in *Proc. Soc. Imag. Sci. Technol. Color Imag. Conf.*, no. 1, 2018, pp. 350–353.

[56] N. Banic, K. Koscevic, M. Subasic, and S. Loncaric, "The past and the present of the color checker dataset misuse," in *Proc. 11th Int. Symp. Image Signal Process. Anal. (ISPA)*, Sep. 2019, pp. 366–371.

[57] F. Laakom, J. Raitoharju, A. Iosifidis, J. Nikkanen, and M. Gabbouj, "INTEL-TAU: A color constancy dataset," 2019, *arXiv:1910.10404*. [Online]. Available: http://arxiv.org/abs/1910.10404

[58] C. Aytekin, J. Nikkanen, and M. Gabbouj, "A data set for camera-independent color constancy," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 530–544, Feb. 2018.

[59] R. Zakizadeh, M. S. Brown, and G. D. Finlayson, "A hybrid strategy for illuminant estimation targeting hard images," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop (ICCVW)*, Dec. 2015, pp. 16–23.

[60] Y. Qian, J. Käpylä, J.-K. Kämäräinen, S. Koskinen, and J. Matas, "A benchmark for temporal color constancy," 2020, *arXiv:2003.03763*. [Online]. Available: http://arxiv.org/abs/2003.03763

[61] J.-M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders, "The Amsterdam library of object images," *Int. J. Comput. Vis.*, vol. 61, no. 1, pp. 103–112, Jan. 2005.

[62] M. Bleier, C. Riess, S. Beigpour, E. Eibenberger, E. Angelopoulou, T. Troger, and A. Kaup, "Color constancy and non-uniform illumination: Can existing algorithms work?" in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Nov. 2011, pp. 774–781.

[63] A. Rizzi, C. Bonanomi, D. Gadia, and G. Riopi, "YACCD2: Yet another color constancy database updated," in *Proc. 18th Color Imag. Displaying, Process., Hardcopy, Appl.*, vol. 8652, Feb. 2013, p. 86520.

[64] A. Smagina, E. Ershov, and A. Grigoryev, "Multiple light source dataset for colour research," in *Proc. 12th Int. Conf. Mach. Vis. (ICMV)*, Jan. 2020, Art. no. 114332.

[65] L. Murmann, M. Gharbi, M. Aittala, and F. Durand, "A dataset of multi-illumination images in the wild," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 4080–4089.

[66] D. Shepelev, A. Tereshin, P. P. Nikolaev, and E. Ershov, "Stereo correspondence problems in terms of linear theory of spectral stimulus formation," *Sensory Syst.*, vol. 32, no. 2, pp. 150–160, 2017.

[67] N. Banic and S. Loncaric, "A perceptual measure of illumination estimation error," in *Proc. 10th Int. Conf. Comput. Vis. Theory Appl.*, 2015, pp. 136–143.

[68] S. Bianco and R. Schettini, "Face-based illuminant estimation," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2012, pp. 623–626.

[69] S. B. Knorr and D. Kurz, "Real-time illumination estimation from faces for coherent rendering," in *Proc. IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, Sep. 2014, pp. 113–122.

[70] M. Males, A. Hedi, and M. Grgic, "Colour balancing using sclera colour," *IET Image Process.*, vol. 12, no. 3, pp. 416–421, Mar. 2018.

[71] S. Bianco and C. Cusano, "Quasi-unsupervised color constancy," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 12212–12221.

[72] S. Koskinen, D. Yang, and J.-K. Kämäräinen, "Cross-dataset color constancy revisited using sensor-to-sensor transfer," in *Proc. Brit. Mach. Vis. Conf.*, 2020.

[73] G. D. Finlayson, R. Zakizadeh, and A. Gijsenij, "The reproduction angular error for evaluating the performance of illuminant estimation algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 7, pp. 1482–1488, Jul. 2017.

[74] G. D. Finlayson, S. D. Hordley, and P. Morovic, "Colour constancy using the chromagenic constraint," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2005, pp. 1079–1086.

[75] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*. [Online]. Available: http://arxiv.org/abs/1804.02767

[76] (Sep. 2020) *Illumination Estimation Challenge*. [Online]. Available: https://www.isispa.org/illumination-estimation-challenge

[77] S. Joo Kim, H. Ting Lin, Z. Lu, S. Süsstrunk, S. Lin, and M. S. Brown, "A new in-camera imaging model for color computer vision and its application," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 12, pp. 2289–2302, Dec. 2012.

[78] M. P. Deisenroth, A. A. Faisal, and C. S. Ong, *Mathematics for Machine Learning*. Cambridge, U.K.: Cambridge Univ. Press, 2020.

[79] (Aug. 2020) *On Illuminant Detection*. [Online]. Available: http://cvil.eecs.yorku.ca/projects/public_html/illuminant/illuminant.html

[80] B. F. L. Shi. (Aug. 2020) *Using High-Level Visual Information for Color Constancy*. [Online]. Available: http://lear.inrialpes.fr/people/vandeweijer/papers/vandeweijer_iccv07.pdf

[81] (Sep. 2020). *Illumination Estimation Challenge*. [Online]. Available: http://chromaticity.iitp.ru/

[82] A. Savchik, E. Ershov, and S. Karpenko, "Color cerberus," in *Proc. 11th Int. Symp. Image Signal Process. Anal. (ISPA)*, Sep. 2019, pp. 355–359.

[83] M. Strathern, "'Improving ratings': Audit in the British University system," *Eur. Rev.*, vol. 5, no. 3, pp. 305–321, 1997.

EGOR ERSHOV received the Ph.D. degree from the Faculty of Radio Engineering and Computer Technology, in 2019. He is currently a Senior Researcher with the Vision Laboratory, Institute of Information Transmission Problems (IITP) (Kharkevic Institute), Russian Academy of Science (RAS). He is also Senior Lecturer with the Computer Science Faculty, Higher School of Economy. He is an Assistant Manager of the Cathedra at Computer Science Faculty, Higher School of Economy. His main areas of research interests include image processing and analysis, particularly color computer vision, color reproduction technologies, colorimetry, human vision systems, Hough transform. He was a recipient of several awards for his scientific and professional work.

ALEXEY (ALEX) SAVCHIK graduated from the Department of Mechanics and Mathematics, Moscow State University, in 2014, and from the Yandex School of Data Analysis, in 2012. He is currently with the Vision Systems Laboratory, Institute for Information Transmission Problems (Kharkevich Institute) of the Russian Academy of Sciences (IITP RAS), Moscow, Russia. His research interests include deep learning and computer vision.

ILLYA (ILYA) SEMENKOV received the bachelor's degree from the National Research University Higher School of Economics (NRU HSE), in 2019, where he is currently pursuing the master's degree with the Faculty of Computer Science. He is also with the Vision Systems Laboratory, Institute for Information Transmission Problems (Kharkevich Institute) of the Russian Academy of Sciences (IITP RAS), Moscow, Russia. His research interests include deep learning, computer vision, statistical learning, and optimal transport.
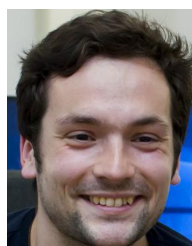
NIKOLA BANIĆ (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in computer science in 2011, 2013, and 2016, respectively. He is currently working as a Senior Computer Vision Engineer at Gideon Brothers, Croatia. He has worked in real-time image enhancement for embedded systems, digital signature recognition, people tracking and counting, and image processing for stereo vision. His research interests include image enhancement, color constancy, image processing for stereo vision, and tone mapping.
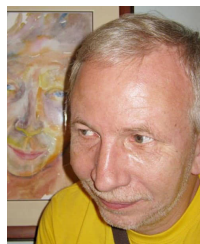
ALEXANDER BELOKOPYTOV graduated from the Moscow Institute of Physics and Technology (MIPT), with specialization in microwave engineering, in 1982.

He is currently with the Vision Systems Laboratory, Institute for Information Transmission Problems (Kharkevich Institute) of the Russian Academy of Sciences (IITP RAS), Moscow, Russia. While at IITP, he participated in research in peripheral and color human vision, underwater photography. His research interests include human vision (color and peripheral), optoelectronics, and scientific data visualization.

DARIA SENSHINA received the bachelor's degree from the Moscow Institute of Physics and Technology (MIPT), in 2020, where she is currently pursuing the master's degree with Phystech School of Applied Mathematics and Informatics. Simultaneously, she is pursuing the M.Sc. degree in industrial and applied maths with the Université Grenoble Alpes, Grenoble, France. She is also with the Vision Systems Laboratory, Institute for Information Transmission Problems (Kharkevich Institute) of the Russian Academy of Sciences (IITP RAS), Moscow, Russia. Her research interests include computer vision and image processing.

KARLO KOŠČEVIĆ (Student Member, IEEE) received the B.Sc. and M.Sc. degrees in computer science in 2016 and 2018, respectively. He is currently pursuing the Ph.D. degree in the technical sciences in the scientific field of computing Ph.D. Program at the Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia. His research interests include image processing, image analysis, and deep learning. His current research interest includes the area of color constancy with a focus on learning-based methods for illumination estimation.
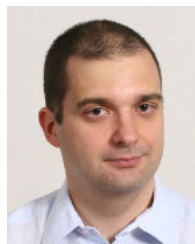
MARKO SUBAŠIĆ (Member, IEEE) received the Ph.D. degree from the Faculty of Electrical Engineering and Computing, University of Zagreb, in 2007. Since 1999, he has been working with the Department for Electronic Systems and Information Processing, Faculty of Electrical Engineering and Computing, University of Zagreb, where he is currently as an Associate Professor. He teaches several courses at graduate and undergraduate levels. His research interests include image processing and analysis and neural networks, with a particular interest in image segmentation, detection techniques, and deep learning. He is a member of the IEEE Computer Society, the Croatian Center for Computer Vision, the Croatian Society for Biomedical Engineering and Medical Physics, and the Centre of Research Excellence for Data Science and Advanced Cooperative Systems.

SVEN LONČARIĆ received the B.Sc., M.Sc., and Ph.D. degrees in 1985, 1989, and 1994, respectively. After earning his Ph.D. degree, he continued his academic career at the Faculty of Electrical Engineering and Computing, University of Zagreb, where he is currently a Full Professor. He was an Assistant Professor with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, NJ, USA, from 2001 to 2003. His main areas of research include image processing and analysis. He is the Director of the Center for Artificial Intelligence, and the Co-Director of the National Center of Research Excellence for Data Science and Advanced Cooperative Systems. Together with his students and collaborators, he has published more than 200 publications in scientific peer-reviewed journals and has presented his work at international conferences. He was a recipient of several awards for his scientific and professional work.

● ● ●