IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 2, 2021, accepted February 27, 2021, date of publication March 8, 2021, date of current version March 16, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3064382

INTEL-TAU: A Color Constancy Dataset

FIRAS LAAKOM[1], JENNI RAITOHARJU[2], (Member, IEEE), JARNO NIKKANEN[3], ALEXANDROS IOSIFIDIS[4], (Senior Member, IEEE), AND MONCEF GABBOUJ[1], (Fellow, IEEE)

[1]Faculty of Information Technology and Communication Sciences, Tampere University, 33100 Tampere, Finland
[2]Programme for Environmental Information, Finnish Environment Institute, FI-40500 Jyväskylä, Finland
[3]Xiaomi Finland Oy, 33720 Tampere, Finland
[4]Department of Engineering, Aarhus University, 8200 Aarhus, Denmark

Corresponding author: Firas Laakom (firas.laakom@tuni.fi)

This work was supported by the NSF-Business Finland Center for Visual and Decision Informatics (CVDI) through the Project Advanced Machine Learning for Industrial Applications (AMALIA) 2019.

ABSTRACT In this paper, we describe a new large dataset for illumination estimation. This dataset, called INTEL-TAU, contains 7022 images in total, which makes it the largest available high-resolution dataset for illumination estimation research. The variety of scenes captured using three different camera models, namely Canon 5DSR, Nikon D810, and Sony IMX135, makes the dataset appropriate for evaluating the camera and scene invariance of the different illumination estimation techniques. Privacy masking is done for sensitive information, e.g., faces. Thus, the dataset is coherent with the new General Data Protection Regulation (GDPR). Furthermore, the effect of color shading for mobile images can be evaluated with INTEL-TAU dataset, as both corrected and uncorrected versions of the raw data are provided. Furthermore, this paper benchmarks several color constancy approaches on the proposed dataset.

INDEX TERMS Color constancy, dataset, illumination estimation, regression.

I. INTRODUCTION

The observed color of an object in a scene depends on its spectral reflectance and spectral composition of the illuminant. As a result, when the scene illuminant changes, the light reflected from the object also changes [1], [2]. The ability to filter out the color of the light source is called color constancy [1], [3]. executing this ability is critical for many image processing and computer vision applications. It results in better quality images. For a robust color-based system, the illumination effects of the light source need to be discounted, so that colors present in the image reflect the intrinsic properties of the objects in the scene [2]–[5]. This is important for many high level image or video applications. Without computational color constancy, colors would be an unreliable feature and inconsistent for object recognition, detection, and tracking. Thus, color constancy research, also called illumination estimation, has been extensively studied and several approaches have been proposed to tackle it [6]–[11].

One key assumption of classical computational color constancy approaches is that the illumination of a scene is uniform. Thus, the problem can be divided into two main steps. In the first step, the global illumination is estimated and, in the second step, all color pixels of the scene are normalized using the estimated illuminant color. As the second step is a straight-forward transformation, the computational color constancy problem is equivalent to illumination estimation. Typically, illumination estimation algorithms are divided into two main groups, namely unsupervised approaches and supervised approaches. The former involves methods requiring no training which are based on low-level statistics [12]–[16] and methods using physics-based dichromatic reflection model [17]–[19], while the latter involves data-driven approaches that learn to estimate the illuminant in a supervised manner using labeled data.

With the advancement of machine learning in general and deep learning in particular, many machine learning-based approaches have been proposed for color constancy [20]–[28]. However, machine learning-based approaches, especially methods relying on Convolutional Neural Networks (CNNs), usually have a large number of parameters that need to be optimized for solving the illumination estimation problem. Thus, training such models require a large amount of labeled data for training and evaluation. Moreover, The performance of such methods in the test scenario

The associate editor coordinating the review of this manuscript and approving it for publication was Sudhakar Radhakrishnan.

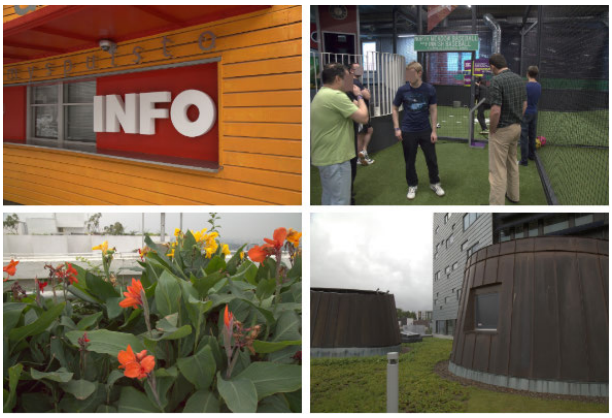

FIGURE 1. Samples from INTEL-TAU dataset.

heavily depends on the quality and the diversity of the data seen during the training process. However, acquiring labeled datasets for illumination estimation is a challenging task [29], as in order to extract the ground truth illumination of a scene, a ColorChecker chart needs to be included in the scene. In addition, after the introduction of General Data Protection Regulation (GDPR) act [30] in Europe, data privacy in datasets needs to be addressed and sensitive information needs to be masked.

In this paper, we propose a new INTEL-TAU dataset for color constancy research. The dataset contains 7022 high-resolution images and it is by far the largest publicly available high-resolution dataset for training and evaluation of color constancy algorithms. Furthermore, all recognizable faces, license plates, and other privacy sensitive information have been carefully masked. Thus, the dataset is now fully GDPR compliant. A subset of 1558 images of the current dataset was previously published as Intel-TUT dataset [31], but had to be retracted due to its GDPR non-compliance. Images in INTEL-TAU dataset were collected using three different cameras: Canon 5DSR, Nikon D810, and Mobile Sony IMX135. The images contain both field and lab scenes. The dataset has mainly real scenes along with some lab printouts with the corresponding white point information. The black level was subtracted from all the scenes and the saturation points, i.e., where the incident light at a pixel causes one of the color channels of the camera sensor to respond at its maximum value producing an undesirable artifact, were normalized. This dataset is suitable for scene and camera-invariance estimation of color constancy algorithms.

The rest of this paper is organized as follows. First, we review the available color constancy datasets in Section II. In Section III, we describe the proposed dataset and highlight its main novelties. We propose several protocols for using this dataset for illumination estimation research in Section V. In Section VI, we evaluate the performance of several baseline and recently proposed color constancy algorithms on the proposed dataset. We conclude the paper in Section VII.

II. PREVIOUSLY PUBLISHED COLOR CONSTANCY DATASETS

One of the most commonly used dataset in color constancy is the ColorChecker dataset [32]. It is composed of 568 high-resolution raw images acquired by two cameras: Canon 1D and Canon 5D. Shi [33] proposed a methodology to reprocess the original images and to recalculate the ground truth. The images are demosaiced and available as TIFF images. The location of the color chart and the saturated and clipped pixels are also provided with the database. Later, Finlayson *et al.* [34] raised a problem, regarding the computation of the ground truth, with the Shi reprocessed dataset. To remedy this problem, a Recommended ColorChecker dataset with an updated ground truth was introduced in [29], [35].

Another publicly available dataset is SFU HDR [36], [37] containing 105 high dynamic range images captured using a calibrated camera. Nine images per scene were captured in order to generate the high dynamic range images. For an accurate measure of global illumination, four color charts were used at different locations in the scene.

NUS-8 [38] has been one of the largest color constancy datasets. It contains 1736 raw images. Eight different camera models were used to capture the scenes of this dataset and a total of $\sim$ 210 images were captured by each camera. Although the dataset is relatively large, a commonly used protocol is to perform tests on each camera separately and report the mean of all the results. As a result, each experiment involves using only 210 images for both training and testing, which is not enough to appropriately train deep learning-based approaches.

Banic and Loncaric introduced the Cube dataset in [39]. This dataset is composed of 1365 RGB images. All dataset images are outdoor scenes acquired with a Canon EOS 550D camera in Croatia, Slovenia, and Austria. This dataset was also extended into Cube+ dataset [39]. This extension was enriched by an additional 342 images containing indoor and outdoor scenes. The overall distribution of illuminations in the Cube+ is similar to the ground truth distribution of the NUS-8.

Other hyperspectral datasets [40]–[43] are available for color constancy research. However, these dataset are relatively scarce and thus unsuitable for machine learning-based solutions with the exception of [43] which contains 11000 images. However, this dataset is actually composed of video frames and, as a result, most of the images are highly correlated and only 600 are not [32]. Moreover, this dataset has low-resolution images that were subject to correction.

A different research direction in computational color constancy is the multi-frame illumination estimation [44], [45]. Several video-based datasets have been proposed to incorporate the temporal information in the learning process [46]–[48]. Another noteworthy type of datasets are the multi-illumination datasets [49]–[52]. It has been argued that many inverse problems involving lighting and material

TABLE 1. Characteristics of different high-resolution color constancy datasets.

Dataset	ColorChecker (Gehler's Raw)	Cube	cube+	SFU HDR	NUS-8	INTEL-TUT	**INTEL-TAU**
Number of samples	568	1365	1707	105	1736	1558	**7022**
Number of camera models	2	1	1	1	8	3	**3**
Indoor and outdoor samples	✓	–	✓	✓	✓	✓	✓
GDPR compliance	–	–	–	–	–	–	✓
Publicly available	✓	✓	✓	✓	✓	–	✓

understanding remain severely ill-posed to be solved efficiently with single-illumination datasets [49], [53], [54]. However, capturing labeled samples of multiple illuminants is hard and, thus, the available datasets are typically small and collected in a controlled environment [49], which make them impractical for real-world applications. Thus, there is still a need for larger single-illumination datasets.

Intel-TUT was proposed in [31]. It contained a subset of 1558 images of the proposed INTEL-TAU dataset. Due to the aforementioned problems with GDPR regulations, it was recently retracted. Furthermore, a larger subset of 3420 images was used for experiments in [27] and [28], but similar privacy issues were encountered. The privacy masking, which we applied for the proposed INTEL-TAU dataset, resolves all the privacy issues, while preserving all the advantages of the previously published subsets and providing further benefits as described in the next section. Table 1 presents a comparison of different color constancy datasets.

III. INTEL-TAU DATASET DESCRIPTION

We introduce a new color constancy dataset, called INTEL-TAU, with the following properties. INTEL-TAU

- is currently the largest publicly available high-resolution color constancy dataset containing 7022 images with ground truth illumination,
- is available at http://urn.fi/urn:nbn:fi:att:f8b62270-d471-4036-b427-f21bce32b965,
- provides the training images without the color charts (i.e., there is no need for color chart masking),
- contains images taken with 3 different cameras to allow camera invariance evaluation,
- contains images grouped by scene type to allow scene invariance evaluation,
- contains mobile images before and after color shading[1] to allow studying the effect of shading,
- is fully GDPR compliant with privacy masking applied on all sensitive information,

INTEL-TAU contains both outdoor and indoor images captured in 17 different countries. There are 7022 1080p[2] images in total, captured using three different cameras: Canon 5DSR, Nikon D810, and Mobile Sony IMX135. The dataset has four folders per camera: *field_1_camera*, containing unique field images captured by the camera, *field_3_cameras* containing images of common[3] scenes captured by all cameras, *lab_printouts*, containing lab printouts, and *lab_realscenes* consisting of real lab scenes. Table 2 reports the numbers of images per category.

TABLE 2. INTEL-TAU composition.

	field_1_cameras	field_3_cameras	lab_printouts	lab_realscenes
Canon	1645	144	300	20
Nikon	2329	144	300	20
Sony	1656	144	300	20

When capturing the images, we avoided strong mixed illumination. Instead, we targeted the framing so that one illumination is dominating in the scene. To define the ground truth, there is one ground truth raw Bayer image associated with each raw Bayer image in the database. The ground truth image has a X-Rite ColorChecker Passport chart positioned in such way that it reflects the main illumination in the scene. The actual database image does not contain the chart, except for a handful of images in which it was intentionally inserted as image contents. The same ground truth image can be associated with multiple database images if the illumination is common in those images. We calculated the ground truth white point from grey patches #20 – #23, omitting the brightest grey patch #19, the darkest grey patch #24, and additional saturated patches if any. Noise was reduced by a 9×9 averaging filter before recording the color component values inside the center area of the grey patch. We manually checked the annotation for each image.

The associated .ccm was not calculated based on the ground truth image, but selected from a pre-calculated set of CCMs according to the estimate of the illumination (daylight, indoor fluorescent, indoor tungsten-halogen). Consequently, the .ccm should not be treated as an accurate color conversion matrix, but just for more convenient illustration. It can further serve as a means to guide the color shading correction that was applied on the Sony IMX135 images. Figure 2 presents an example of ground truth and database image pair as an illustration (this in not the actual raw Bayer content). Figure 3 presents the actual raw images of an example ground truth and database image pair as a reminder to the reader that the database has raw Bayer images. Different camera characteristics are presented in Table 3.

[1]The non-shaded mobile images are available in the additional resources discussed in Section IV

[2]We also provide the full resolution raw data, as described in Section IV.

[3]There are 144 scenes that were pictured by the three camera models. For each camera, the folder *field_3_cameras* contains these images of the scenes for the specific camera model.

TABLE 3. Characteristics of the camera models used in Intel-TAU.

	Canon EOS 5DSR	Nikon D810	Mobile
Resolution	52Mpix (8896H × 5920V)	36Mpix (7380H × 4928V)	8Mpix (3264H × 2448V)
Focal length	EF 24-105/4L @ 28mm (*)	AF-S 24-70/2.8G @ 28mm (*)	30.4mm (actual 4.12mm)
Aperature size	F8.0(**)	F8.0(**)	F2.4
Pixel size	4.14um	4.88um	1.12um
raw data bit depth	14bpp	14bpp	10bpp

(*): 28mm was the closest to the mobile device focal length that was easy to set consistently based on the markings on the objectives
(**): Smaller aperture was used in order to reduce the depth-of-field difference between the DSLRs and the mobile module

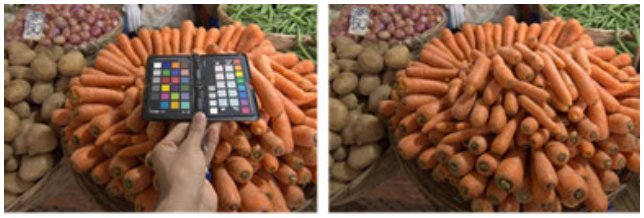

FIGURE 2. An example ground truth and database image pair (illustration, not actual raw Bayer content).

FIGURE 3. An example ground truth and database image pair (actual raw image).

Only the database images are made publicly available along with the ground truth illumination. The ground truth images, i.e., images with the color chart, are not published in this version of the dataset. Thus, no color chart masking needs to be done before evaluating color constancy approaches using the proposed INTEL-TAU dataset. In addition, the black level was subtracted, the saturation points were normalized, and all images were down-sampled to 1080p. The images are stored in TIFF format and the associated groundtruths in the normalized [R,G,B] coordinates. Following the GDPR regulations, we applied privacy masking for recognizable faces, license plates, and other privacy sensitive information. The color component values inside the privacy masking area were averaged.

IV. ADDITIONAL RESOURCES

We also provide the original raw version, i.e., raw Bayer, of the dataset.[4] Different image characteristics are presented in Table 4. The white points are stored as [R/G, B/G] coordinates. The spectral responses of the different camera models and the spectral power distributions of the lab light sources are also provided. It should be noted that the size of the raw image set is large, 290GB, compared to 50GB of the preprocessed dataset images. This variant of the dataset can be used to study the color shading effect as we provide the mobile images before and after color shading. Since raw Bayer images are provided, this dataset can also be used to study various imaging problems, such as demosaicing [55]. The downscaled 1080p version offers a more easily transferrable database size for those use cases that do not require the full resolution, such as color constancy. We also provide a preprocessed downscaled version of the dataset with isotropic scaling: in this variant, the resized images have a height of 1080 lines and the width was adjusted accordingly to preserve the original aspect ratio of the images.

[4]http://urn.fi/urn:nbn:fi:att:f8b62270-d471-4036-b427-f21bce32b965

TABLE 4. Image characteristics in the original raw images of Intel-TAU.

	Canon 5DSR	Nikon D810	Sony IMX135
Image width	8896	7380	3264
Light shielded pixels at left	160	0	0
Image height	5920	4928	2448
Light shielded pixels at top	64	0	0
Bayer order	RG_GB	RG_GB	GR_BG
Raw data bit depth(*)	14	14	10
Data pedestal/black level	2047	601	64
Saturation point (**)	15380	16383	1023

(*): The raw frames are stored as uint16 value per each pixel
(**): Note that the saturation point is not necessarily $2^{raw\_bpp} - 1$
(**): Some of the Sony IMX135 images are upside down

TABLE 5. Characteristics of the 10 subsets of the INTEL-TAU used for evaluation with the cross-validation protocol.

Subset IDs	Subset size	Camera	Country
01	464	Nikon	Finland
02	724	Canon	India
03	701	Canon	Diverse set of countries
04	684	Canon	Diverse set of countries
05	803	Nikon	Tenerife
06	700	Nikon	India and Finland
07	826	Nikon	Iceland and Finland
08	645	Sony	Malta
09	750	Sony	Diverse set of countries
10	725	Sony	Diverse set of countries

V. EXPERIMENTAL PROTOCOLS

We propose two experimental protocols for using the proposed INTEL-TAU dataset. The first protocol is for evaluating the camera invariance of the models. In the cross-validation protocol, a 10-fold non-random cross-validation experiment is conducted.

A. CAMERA INVARIANCE PROTOCOL

The proposed INTEL-TAU dataset can be used to evaluate the camera invariance of color constancy approaches, similarly to [31]. To this end, all scenes acquired by one camera are used

TABLE 6. Results using INTEL-TAU dataset using camera invariance protocols.

	$e_{recovery}$					$e_{reproduction}$				
Method	Best25%	Mean	Med.	Tri.	W.25%	Best25%	Mean	Med.	Tri.	W.25%
Grey-World [57]	0.9	4.7	3.7	4.0	10.0	1.1	5.7	4.6	4.9	11.9
White-Patch [58]	1.1	7.0	5.4	6.2	14.6	1.3	7.5	6.3	6.7	15.7
Grey-Edge [60]	1.0	5.3	4.1	4.5	11.7	1.2	6.2	4.9	5.2	13.4
2nd order Grey-Edge [60]	1.0	5.1	3.8	4.2	11.3	1.2	6.0	4.6	4.8	13.1
Shades-of-Grey [15]	0.7	4.0	2.9	3.2	9.0	0.8	4.8	3.6	3.9	10.9
Cheng et al. 2014 [38]	0.7	4.6	3.4	3.7	10.3	0.9	5.5	4.2	4.5	12.1
Weighted Grey-Edge [14]	0.9	6.0	4.2	4.8	14.2	1.1	6.8	5.0	5.5	15.6
Bianco [22]	0.8	3.4	2.5	2.7	7.2	1.0	4.3	3.2	3.4	9.3
C3AE [28]	0.9	3.4	2.7	2.8	7.0	1.1	3.9	3.3	3.5	8.8
BoCF [27]	0.9	2.9	2.4	2.5	6.1	0.9	3.6	2.8	2.9	7.5
FC^4 (VGG16) [23]	0.7	2.6	2.0	2.2	5.5	0.8	3.3	2.6	2.7	7.1

TABLE 7. Results of benchmark methods on INTEL-TAU dataset using cross-validation protocol.

	$e_{recovery}$					$e_{reproduction}$				
Method	Best25%	Mean	Med.	Tri.	W.25%	Best25%	Mean	Med.	Tri.	W.25%
Grey-World [57]	1.0	4.9	3.9	4.1	10.5	1.2	6.1	4.9	5.2	13.0
White-Patch [58]	1.4	9.4	9.1	9.2	17.6	1.8	10.0	9.5	9.8	19.2
Grey-Edge [60]	1.0	5.9	4.0	4.6	13.8	1.2	6.8	4.9	5.5	13.5
2nd order Grey-Edge [60]	1.0	6.0	3.9	4.8	14.0	1.2	6.9	4.9	5.6	15.7
Shades-of-Grey [15]	0.9	5.2	3.8	4.3	11.9	1.1	6.3	4.7	5.1	13.9
Cheng et al. 2014 [38]	0.7	4.5	3.2	3.5	10.6	0.9	5.5	4.0	4.4	12.7
Weighted Grey-Edge [14]	0.8	6.1	3.7	4.6	15.1	1.1	6.9	4.5	5.4	16.5
Yang et al. 2015 [12]	0.6	3.2	2.2	2.4	7.6	0.7	4.1	2.7	3.1	9.6
Color Tiger [39]	1.0	4.2	2.6	3.2	9.9	1.1	5.3	3.3	4.1	12.7
Greyness Index [59]	0.5	3.9	2.3	2.7	9.8	0.6	4.9	3.0	3.5	12.1
PCC_Q2 [16]	0.6	3.9	2.4	2.8	9.6	0.7	5.1	3.5	4.0	11.9
FFCC [7]	0.4	2.4	1.6	1.8	5.6	0.5	3.0	2.1	2.3	7.1
Bianco [22]	0.9	3.5	2.6	2.8	7.4	1.1	4.4	3.4	3.6	9.4
C3AE [28]	0.9	3.4	2.7	2.8	7.0	1.1	3.9	3.3	3.5	8.8
BoCF [27]	0.7	2.4	1.9	2.0	5.1	0.8	3.0	2.3	2.5	6.5
FC^4 (VGG16) [23]	0.6	2.2	1.7	1.8	4.7	0.7	2.9	2.2	2.3	6.1

for training, all scenes acquired by a second camera are used for validation, and all scenes acquired by a third camera are used for testing in three experiments:

1) Images acquired by Canon as a training set, images acquired by Nikon for validation, and Sony images for testing (training: 2109 images, validation: 2793 images, testing: 2120 images).
2) Images acquired by Nikon for training, images acquired by Sony for validation, and Canon images for testing, (training: 2793 images, validation: 2120 images, testing: 2109 images),
3) Images acquired by Sony for training, images acquired by Canon for validation, and Nikon images for testing (training: 2120 images, validation: 2109 images, testing: 2793 images).

Results are reported as the mean of the results over the three experiments. This test protocol evaluates the camera generalization of the approaches. However, it should be noted that the dataset has multiple samples with similar scene content under various illuminations. Thus, splitting the dataset per camera as illustrated above or using a random split can lead to a partition with a high correlation between the training set and the test set images. To avoid such scenario, we design a second evaluation protocol which is based on 10-fold non-random cross-validation.

B. CROSS-VALIDATION PROTOCOL

Similar to other color constancy datasets, INTEL-TAU contains samples from the same geographical location under different illuminations or using different cameras. Thus, random splitting of the dataset might result in a high correlation between the contents of the training and test set images. To avoid this problem, we propose a non-random 10-fold split of the data to be used for cross-validation. The proposed subset division is provided along with the dataset. In Table 5, we illustrate the characteristics of each subset. We further split the training set, i.e., formed by the nine remaining subsets, by randomly selecting 70% for training and 30% for validation. In total, we have ten experiments and the mean of the achieved results is reported as the final result.

Each subset has around 700 samples except for the first one which has 464 images. Using this evaluation scheme, we have more than 6200 samples for the training and validation in each split. Thus, it is by far the largest training set available for color constancy evaluation. This is extremely useful especially for the evaluation of deep learning-based methods. The results are reported as the mean value of these ten experiments.

VI. EXPERIMENTAL RESULTS

For all experiments, we report the mean of the top 25%, the mean value, the median, Tukey's trimean, and the mean of

TABLE 8. Median errors of $e_{recovery}$ per split for CNN-based approaches for the cross-validation protocol.

	Subset IDS									
Method	01	02	03	04	05	06	07	08	09	10
Bianco [22]	3.6	2.2	2.2	4.1	1.2	2.3	3.0	2.8	2.0	2.5
C3AE [28]	3.9	2.3	2.6	3.0	2.2	2.5	3.0	2.5	2.3	2.6
BoCF [27]	2.3	1.7	1.6	1.8	1.6	1.8	2.2	1.7	2.0	2.0
FC^4 (VGG16) [23]	2.0	1.5	1.5	1.8	1.7	1.8	2.0	1.5	1.7	1.9

the worst 25% of the 'Recovery angular error' [35] between the ground truth white point and the estimated illuminant defined as follows:

$$e_{recovery}(I^{gt}, I^{est}) = \cos^{-1}\left(\frac{I^{gt} I^{est}}{\|I^{gt}\| \|I^{est}\|}\right), \quad (1)$$

where I^{gt} is the ground truth illumination and I^{est} is the estimated illumination. In [56], another robust metric for evaluating the performance of a color constancy method called 'Reproduction angular error' was proposed. It is defined as follows:

$$e_{reproduction}(I^{gt}, I^{est}) = \cos^{-1}\left(\frac{I^{gt}/I^{est} \quad w}{\|I^{gt}/I^{est}\|\sqrt{3}}\right), \quad (2)$$

where / is the element wise division operator and w is defined as the unit vector, i.e., $w = [1, 1, 1]^T$. We also provide results using this error metric.

In our experiments, we considered the following static methods: Grey-World [57], White-Patch [58], Spatial domain [38], Shades-of-Grey [15], and Weighted Grey-Edge [14], Greyness Index 2019 [59], Color Tiger [39], PCC_Q2, and the method reported in [12]. Furthermore, we evaluated the performance of the following learning-based methods: Fast Fourier Color Constancy (FFCC) [7], Fully Convolutional Color Constancy With Confidence-Weighted Pooling (FC^4) [23], Bianco CNN [22], Color Constancy Convolutional Autoencoder (C3AE) [28], and Bag of Color Features [27].

A. CAMERA INVARIANCE PROTOCOL

Table 6 reports the results of several color constancy approaches using the camera invariance protocol. For the unsupervised approaches, we note high error rates for Grey-world, White-Patch, and the variants of Grey-Edge especially in terms of the mean and the worst 25%. Shades of Grey achieves the best results across all metrics for both error functions, $e_{recovery}$ and $e_{reproduction}$.

The supervised approaches yield lower error rates compared to the unsupervised methods especially in terms of the mean and worst 25%. For example, in terms of the worst 25% the top unsupervised method, i.e., Shades-of-Grey, achieves 9° in $e_{recovery}$ compared to 7.2° for the worst supervised method, i.e., Bianco. We note a similar analysis for the $e_{reproduction}$ error metric. For the supervised methods, FC^4 achieves the best performance, especially in terms of the median and the worst 25%.

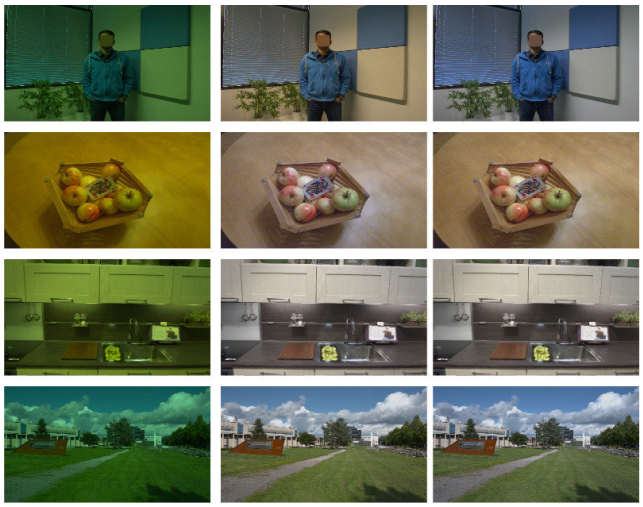

FIGURE 4. Visual results on INTEL-TAU using BoCF method [27]. From left to right, our input images from INTEL-TAU, our corrected images with BoCF method [27], and the ground truth image. .

B. CROSS-VALIDATION PROTOCOL

We perform more extensive experiments using the cross-validation protocol. Table 7 reports the results of different color constancy techniques. For the unsupervised approaches, we note high errors for both angular error metrics. The method in [12] achieves the best results across all the metrics except for the best 25% using $e_{recovery}$, where Grayness Index achieves the smallest errors. It should be noted that the worst 25% error rate is high for all these approaches (larger than 7°).

We note that the supervised methods largely outperform most of the unsupervised approaches, especially in terms of the worst 25%. FFCC, despite not being a deep learning-based approach, achieves competitive results compared to the Convolutional Neural Network (CNN) methods. This can be noted especially in terms of the trimean and median. FC^4 achieves the lowest error rates across all metrics expect for the median, where FFCC achieves the best results.

Figure 4 illustrates visual results on three INTEL-TAU image samples, along with the output of BoCF [27] and the associated ground truth. We note that for a different types of scenes, the model manages to recover the original colors of the scene and produce visually similar results to the output.

In Table 8, we report the median result of various CNN-based approaches, i.e., FC^4, Bianco, C3AE, and BoCF on the different folds of the split. We note high error rates for specific

folds: the first, forth, and seventh subsets. We also note that the second, and ninth sets are easy to learn and most CNN-based approaches achieve less than 2.3° median error.

VII. CONCLUSION

In this paper, a new color constancy dataset, namely INTEL-TAU, is presented. It is the largest available dataset and thus the most suitable for deep learning methods evaluation. The diversity of scenes and camera models makes the new database appropriate for evaluating the camera and scene invariance of different illumination estimation techniques. Privacy masking has been applied for sensitive information, e.g., faces, thus, rendering the dataset compliant with the new GDPR regulations. Furthermore, the effect of color shading for mobile images can be evaluated with INTEL-TAU, as it provides both corrected and uncorrected versions of the raw mobile data.

ACKNOWLEDGMENT

The authors thank Harish Essaky Sankaran, Uygar Tuna, and Lauri Suomela for participating in capturing the new images.

REFERENCES

[1] M. Ebner, *Color Constancy*, vol. 7. Hoboken, NJ, USA: Wiley, 2007.
[2] Y. Fu, Y. Zheng, L. Zhang, and H. Huang, "Spectral reflectance recovery from a single RGB image," *IEEE Trans. Comput. Imag.*, vol. 4, no. 3, pp. 382–394, Sep. 2018.
[3] A. D. Logvinenko, B. Funt, H. Mirzaei, and R. Tokunaga, "Rethinking colour constancy," *PLoS ONE*, vol. 10, no. 9, Sep. 2015, Art. no. e0135029.
[4] A. Halimi, J. M. Bioucas-Dias, N. Dobigeon, G. S. Buller, and S. McLaughlin, "Fast hyperspectral unmixing in presence of nonlinearity or mismodeling effects," *IEEE Trans. Comput. Imag.*, vol. 3, no. 2, pp. 146–159, Jun. 2017.
[5] M. R. Kellman, E. Bostan, N. A. Repina, and L. Waller, "Physics-based learned design: Optimized coded-illumination for quantitative phase imaging," *IEEE Trans. Comput. Imag.*, vol. 5, no. 3, pp. 344–353, Sep. 2019.
[6] 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.
[7] J. T. Barron and Y.-T. Tsai, "Fast Fourier color constancy," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 886–894.
[8] X.-S. Zhang, S.-B. Gao, R.-X. Li, X.-Y. Du, C.-Y. Li, and Y.-J. Li, "A retinal mechanism inspired color constancy model," *IEEE Trans. Image Process.*, vol. 25, no. 3, pp. 1219–1232, Mar. 2016.
[9] 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.
[10] M. Afifi and M. S. Brown, "Deep white-balance editing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1397–1406.
[11] 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
[12] 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.
[13] A. Rizzi, C. Gatta, and D. Marini, "Color correction between gray world and white patch," *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 4662, pp. 367–375, May 2002.
[14] A. Gijsenij, T. Gevers, and J. van de Weijer, "Physics-based edge evaluation for improved color constancy," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 581–588.
[15] G. D. Finlayson and E. Trezzi, "Shades of gray and colour constancy," in *Proc. Color Imaging Conf.*, vol. 1. Springfield, VA, USA: Society for Imaging Science and Technology, 2004, pp. 37–41.
[16] F. Laakom, J. Raitoharju, A. Iosifidis, U. Tuna, J. Nikkanen, and M. Gabbouj, "Probabilistic color constancy," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2020, pp. 978–982.
[17] H. C. Lee, "Method for computing the scene-illuminant chromaticity from specular highlights," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 3, no. 10, pp. 1694–1699, 1986.
[18] R. Tan, K. Nishino, and K. Ikeuchi, "Color constancy through inverse-intensity chromaticity space," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 21, no. 3, pp. 321–334, 2004.
[19] R. Lu, A. Gijsenij, T. Gevers, D. Xu, V. Nedovic, and J. M. Geusebroek, "Color constancy using 3D scene geometry," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep. 2009, pp. 1749–1756.
[20] 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.
[21] 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.
[22] 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.
[23] Y. Hu, B. Wang, and S. Lin, "FC^4: Fully convolutional color constancy with confidence-weighted pooling," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4085–4094.
[24] W. Shi, C. C. Loy, and X. Tang, "Deep specialized network for illuminant estimation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Springer, 2016, pp. 371–387.
[25] J. T. Barron, "Convolutional color constancy," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 379–387.
[26] 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.
[27] F. Laakom, N. Passalis, J. Raitoharju, J. Nikkanen, A. Tefas, A. Iosifidis, and M. Gabbouj, "Bag of color features for color constancy," *IEEE Trans. Image Process.*, vol. 29, pp. 7722–7734, Jul. 2020.
[28] F. Laakom, J. Raitoharju, A. Iosifidis, J. Nikkanen, and M. Gabbouj, "Color constancy convolutional autoencoder," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Dec. 2019, pp. 1085–1090.
[29] G. Hemrit, G. Finlayson, A. Gijsenij, P. Gehler, S. Bianco, B. Funt, M. Drew, and L. and Shi, "Rehabilitating the colorchecker dataset for illuminant estimation," in *Proc. Color Imag. Conf.*, 2018, pp. 350–353.
[30] P. Voigt and A. Von dem Bussche, "The eu general data protection regulation (GDPR)," *A Practical Guide*. Cham, Switzerland: Springer, 2017.
[31] 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.
[32] 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.
[33] L. Shi. (2000). *Re-Processed Version of the Gehler Color Constancy Dataset of 568 Images*. [Online]. Available: http://www.cs.sfu.ca/~color/data/
[34] 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.*, vol. 25. Springfield, VA, USA: Society for Imaging Science and Technology, 2017, pp. 64–69.
[35] S. D. Hordley and G. D. Finlayson, "Re-evaluating colour constancy algorithms," in *Proc. 17th Int. Conf. Pattern Recognit.*, 2004, pp. 76–79.
[36] B. Funt and L. Shi, "The effect of exposure on MaxRGB color constancy," *Proc. SPIE Hum. Vis. Electron. Imag.*, vol. 7527, Dec. 2010, Art. no. 75270Y.
[37] B. Funt and L. Shi, "The rehabilitation of MaxRGB," in *Proc. 18th Color Imag. Conf. Final Program Color Imag. Conf.*, vol. 1. Springfield, VA, USA: Society for Imaging Science and Technology, 2010, pp. 256–259.
[38] 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.
[39] N. Banić, K. Koščevic, and S. Lon čarié, "Unsupervised learning for color constancy," 2017, *arXiv:1712.00436*. [Online]. Available: http://arxiv.org/abs/1712.00436
[40] K. Barnard, L. Martin, B. Funt, and A. Coath, "A data set for color research," *Color Res. Appl.*, vol. 27, no. 3, pp. 147–151, Jun. 2002.
[41] S. M. Nascimento, F. P. Ferreira, and D. H. Foster, "Statistics of spatial cone-excitation ratios in natural scenes," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 19, no. 8, pp. 1484–1490, 2002.
[42] M. J. Vrhel, R. Gershon, and L. S. Iwan, "Measurement and analysis of object reflectance spectra," *Color Res. Appl.*, vol. 19, no. 1, pp. 4–9, Feb. 1994.

[43] F. Ciurea and B. Funt, "A large image database for color constancy research," in *Proc. Color Imag. Conf.*, vol. 1. Springfield, VA, USA: Society for Imaging Science and Technology, 2003, pp. 160–164.
[44] Y. Qian, K. Chen, J. Nikkanen, J.-K. Kamarainen, and J. Matas, "Recurrent color constancy," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 5458–5466.
[45] S. Beigpour, M. L. Ha, S. Kunz, A. Kolb, and V. Blanz, "Multi-view multi-illuminant intrinsic dataset," in *Proc. BMVC*, 2016, pp. 10.1–10.13.
[46] 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
[47] V. Prinet, D. Lischinski, and M. Werman, "Illuminant chromaticity from image sequences," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 3320–3327.
[48] J.-S. Yoo and J.-O. Kim, "Dichromatic model based temporal color constancy for AC light sources," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, p. 12.
[49] 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.
[50] Y. Aksoy, C. Kim, P. Kellnhofer, S. Paris, M. Elgharib, M. Pollefeys, and W. Matusik, "A dataset of flash and ambient illumination pairs from the crowd," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 634–649.
[51] Z. Hui, A. C. Sankaranarayanan, K. Sunkavalli, and S. Hadap, "White balance under mixed illumination using flash photography," in *Proc. IEEE Int. Conf. Comput. Photography (ICCP)*, May 2016, pp. 1–10.
[52] X. Hao, B. Funt, and H. Jiang, "Evaluating colour constancy on the new mist dataset of multi-illuminant scenes," in *Proc. Color Imag. Conf.*, vol. 1. Springfield, VA, USA: Society for Imaging Science and Technology, 2019, pp. 108–113.
[53] M. Afifi, B. Price, S. Cohen, and M. S. Brown, "When color constancy goes wrong: Correcting improperly white-balanced images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1535–1544.
[54] M. Afifi and M. Brown, "What else can fool deep learning? Addressing color constancy errors on deep neural network performance," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 243–252.
[55] N.-S. Syu, Y.-S. Chen, and Y.-Y. Chuang, "Learning deep convolutional networks for demosaicing," 2018, *arXiv:1802.03769*. [Online]. Available: http://arxiv.org/abs/1802.03769
[56] 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.
[57] J. Cepeda-Negrete and R. Sanchez-Yanez, "Gray-world assumption on perceptual color spaces," in *Proc. Pacific-Rim Symp. Image Video Technol.*, 2014, pp. 493–504.
[58] E. H. Land and J. J. McCann, "Lightness and Retinex theory," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 61, no. 1, pp. 1–11, 1971.
[59] Y. Qian, J.-K. Kamarainen, J. Nikkanen, and J. Matas, "On finding gray pixels," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8062–8070.
[60] 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.

FIRAS LAAKOM received the Engineering degree from the Tunisia Polytechnic School (TPS), in 2018. He is currently pursuing the Ph.D. degree with Tampere University, Finland. His research interests include deep learning, computer vision, and computational intelligence.

JENNI RAITOHARJU (Member, IEEE) received the Ph.D. degree from the Tampere University of Technology, Finland, in 2017. She currently works as a Senior Research Scientist with the Finnish Environment Institute, Jyväskylä. She also leads two research projects funded by the Academy of Finland focusing on automatic taxa identification. She has coauthored 24 international journal articles and 34 papers in international conferences. Her research interests include machine learning and pattern recognition methods along with applications in biomonitoring and autonomous systems. From 2019 to 2021, she is the Chair of the Young Academy Finland.

JARNO NIKKANEN received the M.Sc. and Dr.Sc.Tech. degrees in signal processing from the Tampere University of Technology, in 2001 and 2013, respectively. He has over 20 years of industry experience in digital imaging topics, including Nokia Corporation, where he developed and productized many digital camera algorithms, Intel Corporation, where he worked as a Principal Engineer and an Imaging Technology Architect, and Xiaomi Technologies, where he is currently working as a Senior Director and the Head of Xiaomi Finland Camera Research and Development. He holds international patents for many digital camera related inventions that have been productized in various types of camera devices.

ALEXANDROS IOSIFIDIS (Senior Member, IEEE) is currently an Associate Professor with Aarhus University, Denmark. He has contributed in more than twenty Research and Development projects financed by EU, Finnish, and Danish funding agencies and companies. He has coauthored 78 articles in international journals and 96 papers in international conferences proposing novel Machine Learning Techniques and their application in a variety of problems. He served as an Officer for the Finnish IEEE Signal Processing-Circuits and Systems Chapter.

MONCEF GABBOUJ (Fellow, IEEE) received the M.S. and Ph.D. degrees in electrical engineering from Purdue University, in 1986 and 1989, respectively. He is currently a Professor of signal processing with the Department of Computing Sciences, Tampere University, Finland. His research interests include big data analytics, multimedia analysis, artificial intelligence, machine learning, pattern recognition, nonlinear signal processing, and video processing and coding. He is a member of the Academia Europaea and the Finnish Academy of Science and Letters.