Analysis of Biases in Automatic White Balance Datasets

Simone Bianco [1], Marco Buzzelli [1,*], Gianluigi Ciocca [1], Raimondo Schettini [1], Mikhail Tchobanou [2] and Simone Zini [1]

[1] University of Milano - Bicocca; simone.bianco@unimib.it, marco.buzzelli@unimib.it, gianluigi.ciocca@unimib.it, raimondo.schettini@unimib.it, s.zini1@campus.unimib.it
[2] Moscow Research Center, Huawei Technologies Co. Ltd; Mikhail.Tchobanou@huawei.com
* Corresponding author: marco.buzzelli@unimib.it

Abstract

Learning-based methods for Automatic White Balance (AWB) are trained on properly-annotated datasets, where each image is associated to a ground truth illuminant. The intrinsic characteristics of such datasets, therefore, play a fundamental role in the generalization capability of the resulting AWB model. In this paper we analyze the biases of commonly-used datasets for Automatic White Balance: ColorChecker, Cube+, Gray Ball, INTEL-TAU, and NUS from National University of Singapore. We describe each dataset in terms of employed cameras, distribution of the illuminants, shooting parameters, and image content. The resulting analysis highlights the individual shortcomings of each dataset, as well as the type of image that is under-represented by all analyzed datasets, such as artificial-light and low-light scenarios.

Keywords: *Automatic white balance, illuminant estimation, color constancy, dataset analysis.*

INTRODUCTION

Automatic White Balance (AWB), sometimes referred to as computational color constancy, is the task of correcting a digital image as if the scene was captured under some reference illumination conditions. The development and benchmarking of AWB algorithms is based on the availability of datasets of images whose illuminant ground truth is given, or which can be easily estimated from a color target placed in the scene. In this paper we analyze the most commonly used datasets in order to verify if there is a bias in the illuminant chromaticity distributions, in shooting conditions, and in the semantic image content distribution. Image content is considered important as it is often exploited to various extents by different AWB algorithms, such as Gijsenij et al. (2010), Bianco et al. (2008) and (2012), and Buzzelli et al. (2018).

We consider our investigation necessary since the presence of a bias in the image datasets used in the AWB research not only may invalidate the evaluation of the methods available in the state of the art, but it is increasingly critical as almost all of the new methods are based on machine learning algorithms as shown by Ershov et al. (2021). The dependencies, cross-talks and deficiencies that we have found across all datasets suggest possible directions for the design of novel AWB methods and for future data collection.

ANALYZED DATASETS

For the analysis of this paper, we focused on five datasets considered relevant in the state of the art research for Automatic White Balance. Their main characteristics are reported in Table 1.

Title	Year	Cameras	Images	Reference target
ColorChecker (Gehler et al.)	2008	2	568	24-patch Macbeth Color Checker
Cube+ (Banić et al.)	2014	1	1707	Datacolor SpyderCUBE
Gray Ball (Ciurea and Funt)	2003	1	11346	Gray sphere
INTEL-TAU (Laakom et al.)	2019	3	7022	X-Rite ColorChecker Passport chart
NUS (Cheng et al.)	2014	9	1853	24-patch Macbeth Color Checker

Table 1: Analyzed datasets for Automatic White Balance, with their main characteristics.

- **ColorChecker** by Gehler et al. (2008).
  The dataset is composed of 568 images. Of these, 86 were shot with a Canon EOS-1DS, and the remaining 482 with a Canon EOS 5D. All images include in the frame a 24-patch Macbeth Color Checker target. Multiple versions of the ground truth of this dataset have been proposed through the years, such as the 'reprocessed' version by Shi and Funt (2012), and the more recent 'recommended' version by Hemrit et al. (2018), which has been used in this work.
- **Cube+** by Banić et al. (2017).
  This dataset contains all 1365 images of previous the "Cube" dataset and additional 342 images, for a total 1707 images, shot using a Canon EOS 550D camera. In the lower right corner of each image, the SpyderCube calibration object is placed. Its two neutral 18% gray faces were used to determine the ground-truth illumination for each image, thus providing two annotations per image, with one manually selected by the original authors as better representative for the corresponding scene.
- **Gray Ball** by Ciurea and Funt (2003).
  The dataset contains 11,346 images divided into 15 sequences, with many shots acquired at close interval one from another. Many of the images depict people, and include both indoor and outdoor scenarios, the latter taken in two different locations. The dataset was collected using a Sony VX-2000 digital video camera, and every shot includes the eponymous gray ball color target for ground truth annotation in the bottom-right corner. The images are provided in non-linear 8bit RGB format.
- **INTEL-TAU** by Laakom et al. (2019).
  This dataset is a set of 7022 images in total, captured using three different camera models: Canon 5DSR, Nikon D810, and Sony IMX135. It is the updated version of INTEL-TUT, which was released and later withdrawn due to non-compliance with privacy regulations. The ground truth of each image is defined through a separate set of images, captured in the same scenes but not included in the actual database, depicting s a X-Rite ColorChecker Passport chart reflecting the main illumination source.
- **NUS** by Cheng et al. (2014).
  Collected by the National University of Singapore (NUS), this dataset is composed of a total 1853 images, shot with 9 different cameras: Canon EOS-1Ds Mark III (259 images) Canon EOS 600D (200 images) Fujifilm X-M1 (196 images) Nikon D40 (117 images) Nikon D5200 (200 images) Olympus E-PL6 (208 images) Panasonic Lumix DMC-GX1 (203 images) Samsung NX2000 (202 images) and Sony SLT-A57 (268 images). All images include a 24-patch Macbeth Color Checker target.

ILLUMINANT CHROMATICITY DISTRIBUTIONS

In order to compare existing datasets on equal ground, it is necessary to consider the different spectral sensitivities of the sensors involved in their acquisition, i.e., to map the dataset illuminants into a device-independent color space by normalizing them to a reference white, for which we chose CIE D55. INTEL-TAU is the only dataset providing a colorimetric characterization of the employed cameras, for which we could accurately compute the corresponding reference white using the D55 spectrum from the Light Spectral Power Distribution Database by Roby and Aubé (2015). For the remaining datasets, we referred to a definition by Luo (2016), according to which “illuminant D55 can be assumed to represent the SPD for (direct) sunlight provided that the sun is not too low in the sky”. We thus handpicked for each camera an image that best-represents the adopted definition, relying on metadata such as timestamp and location to reinforce our selection, and chose the corresponding annotation as reference white in RAW-specific RGB form. These values were finally used to normalize all the datasets illuminants using a Von Kries-like transform, and were eventually plotted in the Angle-Retaining Chromaticity diagram by Buzzelli et al. (2020) to avoid any representation-specific distortion of the data. The result is presented in Figure 2, with reference whites reported in the legend as R/G and B/G chromaticity pairs, as well as non-normalized points indicated by an × symbol.

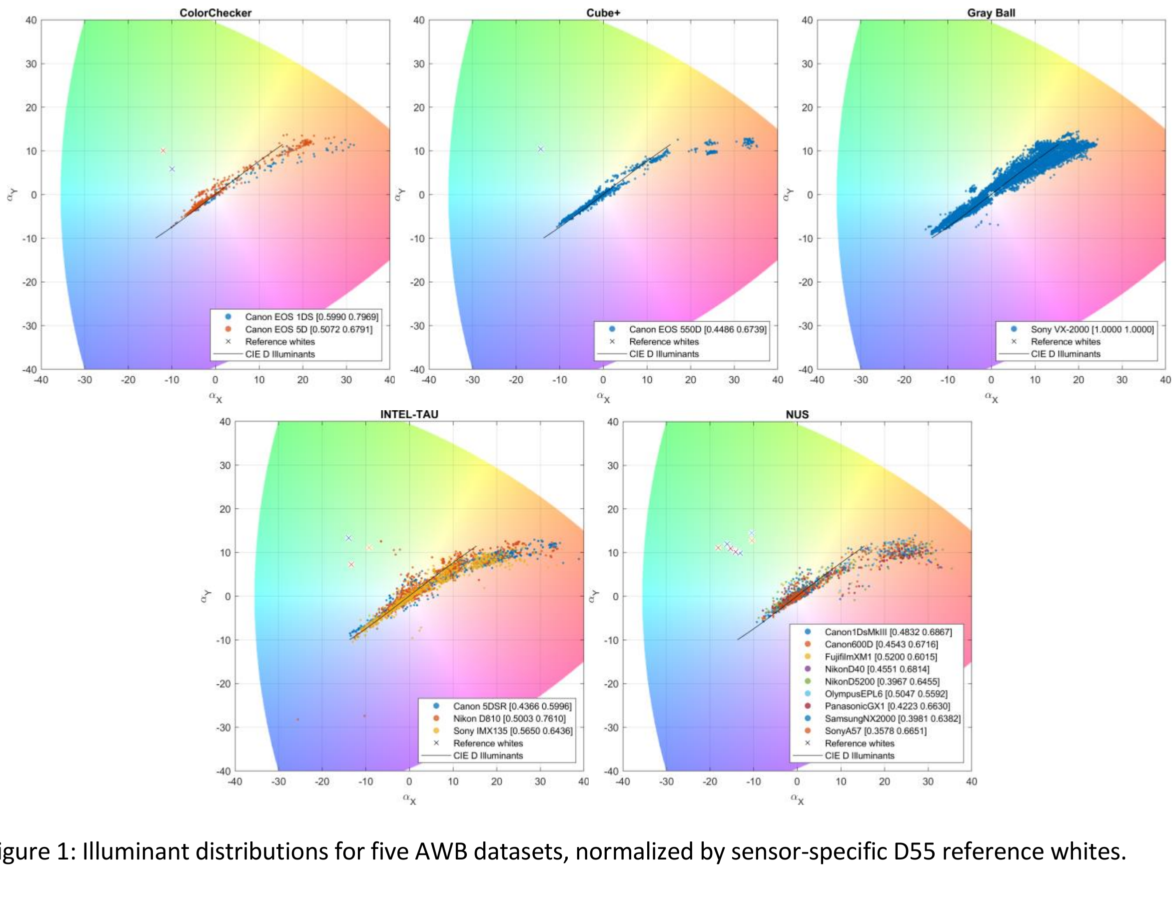

Figure 1: Illuminant distributions for five AWB datasets, normalized by sensor-specific D55 reference whites.

For further reference, CIE series D illuminants from D40 to D150 have also been reported in each plot. It is possible to observe how all analyzed datasets roughly follow the distribution of daylight illuminants, eventually contributing with additional data points at low Correlated Color Temperatures (CCT), typically found in indoor scenarios illuminated by incandescent light sources. The area of high CCTs, commonly covered by outdoor in-shadow surfaces, appears to be better represented by the Gray Ball dataset, and by the more recent INTEL-TAU dataset. Finally, the direction orthogonal to the axis defined by CIE series D illuminants, i.e. ranging from greenish to magenta-ish lights, is poorly represented by all the analyzed datasets. Collecting data from such range of non-natural light sources could be potentially useful in conducting research on human perception.

SHOOTING PARAMETERS AND ILLUMINATION LEVELS

We have characterized the five AWB datasets in terms of illumination levels, i.e. low-light to bright-light conditions, following two complementary approaches. With the first approach, we extracted shooting parameters from EXIF data, whose distributions are reported in Figure 2 (please, note that EXIF data are not available for the sRGB-encoded Gray Ball dataset, and that no ISO information was found for images from the Sony IMX135 camera of the INTEL-TAU dataset). Under the assumption of properly-set camera parameters, these pieces of information can be exploited to produce an estimation of the scene illumination according to Le et al. (2019):

$$I_{measure} = \log_{10}\left(\frac{aperture^2}{exposure\ time}\right) + \log_{10}\left(\frac{250}{ISO}\right). \quad (1)$$

The resulting distribution is shown in Figure 3 (left). Alternatively, the scene illumination conditions can be inferred by classifying the images into discrete labels, such as: “Highlight”, “Lowlight”, and “Sunset/Sunrise”. This information, shown in Figure 3 (right), has been obtained by training an unpublished Convolutional Neural Network (CNN) on a proprietary dataset of annotated images, and applying it to sRGB-rendered white-balanced images from the AWB datasets.

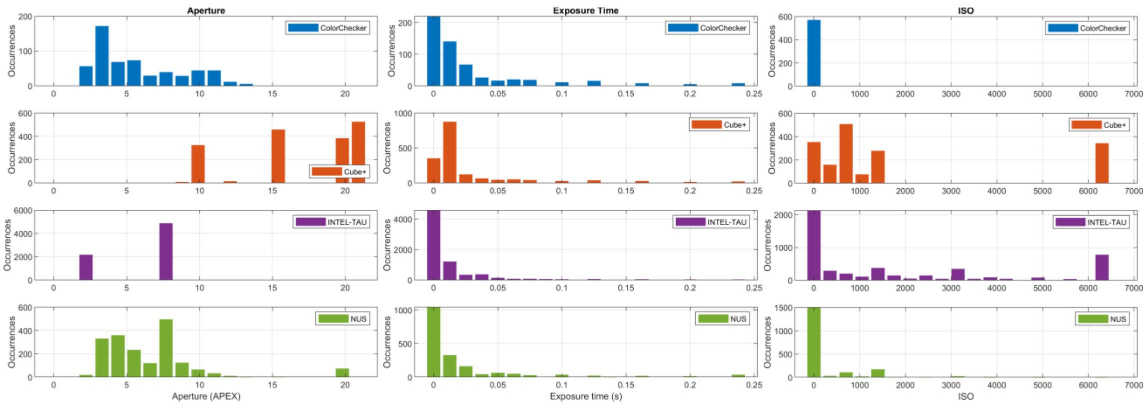

Figure 2: Shooting parameters distributions for the analyzed AWB datasets: Aperture, Exposure Time, and ISO.

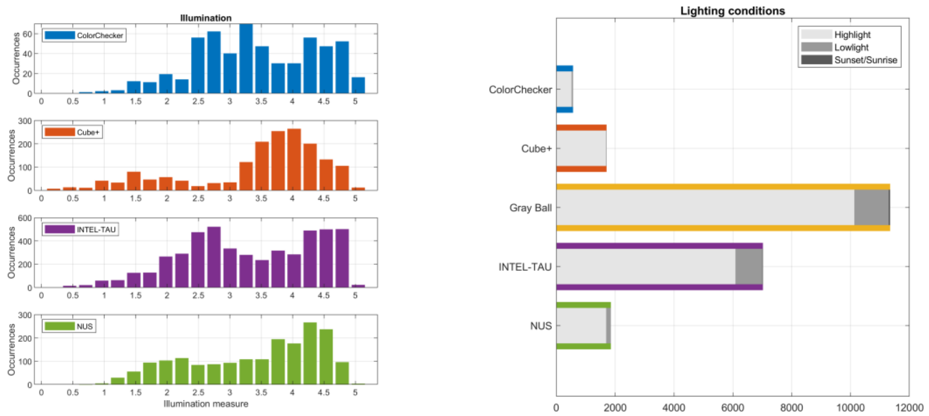

Figure 3: Illumination conditions computed by means of shooting parameters (left) and image analysis (right).

Without loss of generality, it is possible to observe that both independent analyses show a general predominance for bright light scenarios. Middle-to-low light scenes are partially represented by the INTEL-TAU dataset, by the Gray Ball dataset (only supported by classification-based analysis), and to some extent by the ColorChecker dataset, although its low number of total images must also taken into consideration. Finally, the Sunset/Sunrise scenario is scarcely covered across all datasets. We argue that such extreme imaging conditions are particularly relevant for real-life applications, as consumer devices equipped with AWB modules are found to often produce sub-optimal results in these set-ups.

IMAGE CONTENT

We have further analyzed the datasets in terms of image content. Following the same procedure described in the previous Section, we have trained three CNNs on a proprietary annotated dataset, and we have performed inference on white-balanced images from the five AWB datasets. The aggregated results are presented in Figure 4.

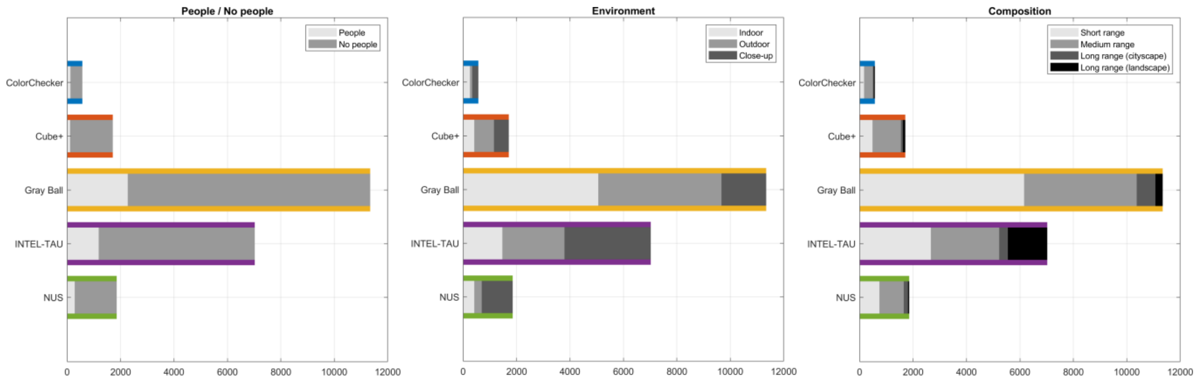

Figure 4: Image content distribution of the AWB datasets, according to three different sets of classes.

The amount of images depicting human subjects is relatively small, a problem partially related to privacy-protecting instruments such as the General Data Protection Regulation (GDPR). Note that, due to automatic annotation, some false positives (related to statues, posters, etc.) might occur, such as in the Cube+ dataset which is known not to contain any proper human subject. The INTEL-TAU dataset depicts a relevant number of people, however their faces are censored with an average-color mask, thus preventing the application of some semantic-based AWB methods, such as the one developed by Bianco et al. (2012). Finally, in terms of environments (indoor, outdoor, and close-ups) and composition (close-to-long range), all datasets appear to be well-balanced.

CONCLUSIONS

We have compared the most popular datasets for Automatic White Balance in terms of illuminant distribution, shooting parameters, and image content. We have highlighted the individual shortcomings of each dataset, thus suggesting the opportunity to merge multiple datasets into a more-complete set of images. For sensor-dependent AWB methods, this type of fusion can only be exploited by bringing the datasets into a common representation. To this extent, a set of reference white points has been compiled for each involved sensor and shared within this manuscript.

Deficiencies have been found across all datasets, such as a lack of images illuminated with artificial light sources and/or low-light images, indicating a direction for future data collection.

An analysis on image content has also been provided. As a future development, it would be valuable to extract similar statistics on non-AWB datasets that well-represent common user photographic collections, in order to compare their content distributions.

ACKNOWLEDGEMENTS

This work is part of a research collaboration between the Imaging and Vision Laboratory from University of Milano - Bicocca, and the Moscow Research Center from Huawei Technologies Co. Ltd.

REFERENCES

Banić, N., K. Koščević, and S. Lončarić. 2017. Unsupervised learning for color constancy. *arXiv preprint arXiv:1712.00436*.

Bianco, S., G. Ciocca, C. Cusano, and R. Schettini. 2008. Classification-based color constancy. In: *International Conference on Advances in Visual Information Systems*, Springer, Berlin, Heidelberg, pp. 104-113.

Bianco, S., and R. Schettini. 2012. Color constancy using faces. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 65-72.

Buzzelli, M., J. van de Weijer, and R. Schettini. 2018. Learning illuminant estimation from object recognition. In: *2018 25th IEEE International Conference on Image Processing*, IEEE, pp.3234-3238.

Buzzelli, M., S. Bianco, and R. Schettini. 2020. ARC: Angle-Retaining Chromaticity diagram for color constancy error analysis. *JOSA A, 37*(11), 1721-1730.

Cheng, D., D. K. Prasad, and M. S. Brown. 2014. Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution. *JOSA A*, *31*(5), 1049-1058.

Ciurea, F., and B. Funt. 2003. A large image database for color constancy research. In: *Color and Imaging Conference*, Society for Imaging Science and Technology, Vol. 2003, No. 1, pp. 160-164.

Ershov, E., A. Savchik, I. Semenkov, N. Banić, K. Koščević, M. Subašić, A. Belokopytov, A. Terekhin, D. Senshina, A. Nikonorov, Z. Li, Y. Qian, M. Buzzelli, R. Riva, S. Bianco, R. Schettini, J. T. Barron, S. Lončarić, and D. Nikolaev. 2021. Illumination estimation challenge: The experience of the first 2 years. *Color Research & Application*, 46(4), 705-718.

Gehler, P. V., C. Rother, A. Blake, T. Minka, and T. Sharp. 2008. Bayesian color constancy revisited. In: *2008 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 1-8.

Gijsenij, A., and T. Gevers. 2010. Color constancy using natural image statistics and scene semantics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4), 687-698.

Hemrit, G., G. D. Finlayson, A. Gijsenij, P. Gehler, S. Bianco, B. Funt, M. Drew, and L. Shi. 2018. Rehabilitating the colorchecker dataset for illuminant estimation. In: *Color and Imaging Conference*, Society for Imaging Science and Technology, Vol. 2018, No. 1, pp. 350-353.

Laakom, F., J. Raitoharju, J. Nikkanen, A. Iosifidis, and M. Gabbouj. 2021. Intel-tau: A color constancy dataset. *IEEE Access*, *9*, 39560-39567.

Le, Q. T., P. Ladret, H. T. Nguyen, and A. Caplier. 2019. Large Field/Close-Up Image Classification: From Simple to Very Complex Features. In: *International Conference on Computer Analysis of Images and Patterns*, Springer, Cham, pp. 532-543.

Luo, M. R. (Ed.). 2016. *Encyclopedia of color science and technology*. Springer New York.

Roby, J., and M. Aubé. 2015. *LSPDD | Light Spectral Power Distribution Database*. Online: https://lspdd.org/app/en/lamps?lampuse=Standard%20Illuminant. Last accessed: June 21, 2021.

Shi, L., and B. Funt. 2012. *Shi's Re-processing of Gehler's Dataset*. Online: https://www2.cs.sfu.ca/~colour/data/shi_gehler/. Last accessed: June 23, 2021.