https://doi.org/10.2352/issn.2694-118X.2021.LIM-63

Image understanding for color constancy and vice versa

***Simone Bianco*[1] *, Marco Buzzelli*[1]**
[1] ***University of Milano-Bicocca, Milano, Italy***

Abstract

In this article we show the change in paradigm occurred in color constancy algorithms: from a pre-processing step in image understanding, to the exploitation of image understanding and computer vision results and techniques. Since color constancy is an ill-posed problem, we give an overview of the assumptions on which classical color constancy algorithms are based in order to solve it. Then, we chronologically review the color constancy algorithms that exploit results and techniques borrowed from the image understanding research field in order to exploit assumptions that could be met in a larger number of images.

Introduction

The observed color of the objects in the scene depends on the surface spectral reflectance of the object, on the illumination, on their relative positions, and on the observer characteristics. Many computer vision problems both in the image and video domain where color is an important feature for distinguishing objects such as visual recognition [16], surveillance [22], image manipulation detection [35] etc., make or can make use of color constancy processing as a pre-processing step in order to guarantee that the recorded color of the objects in the scene does not change when the illumination conditions vary.

Despite its apparent simplicity, being able to estimate the color of the illuminant in the scene solely from the image data is a very challenging problem for both human and computer vision systems [26, 21], in fact many algorithms are explicitly inspired by the mechanisms of human vision [39, 23, 1, 24]. Computational color constancy is a two-stage operation: the former is devoted to the estimation of the color of the scene illuminant from the image data, the latter corrects the image on the basis of this estimate to generate a new image of the scene as if it was taken under a reference illuminant. While the second one is generally performed using the diagonal von Kries model [30], the first one addresses a severely ill-posed problem as its solution lacks uniqueness or stability. To cope with this problem, many color constancy solutions in the state of the art exploit some assumptions about the statistical properties of the expected illuminants and/or of the objects reflectances in the scene.

With the aim of providing more robust assumptions, in the last years more and more color constancy algorithms have started to exploit computer vision techniques and results.

The aim of this paper is to provide a chronological overview of this last category of algorithms showing the different computer vision techniques exploited to obtain robust color constancy algorithms. A graphical representation of the timeline representing the change in the paradigm of the color constancy algorithms, from the use of color constancy for image understanding to the use of image understanding techniques for the design of more robust color constancy algorithms is reported in Figure 1.

Color constancy for image understanding

As previously said in the Introduction, illuminant estimation is a severely ill-posed problem, and traditional color constancy algorithms need to exploit some assumptions in order to find a solution to this problem. In this section we will review the assumptions on which the so-called statistics-based approaches are based.

Statistics-based approaches

Early color constancy algorithms, often referred to as statistics-based approaches, exploit assumptions about the statistical properties of the expected illuminants and/or of the object reflectances in the scene. For example, the Grey World (GW) algorithm [13], is based on the assumption that the average reflectance in a scene is achromatic, and estimates the color of the illuminant in the scene as the deviation from this hypothesis.

The White Point (WP) algorithm [31], also known as Maximum RGB, is based on the assumption that the maximum reflectance in a scene is achromatic. Therefore it estimates the color of the illuminant in the scene as the maximum value in the image in the three channels separately.

The Shades of Gray (SoG) algorithm [20], is based on the assumption that the $n-th$ Minkowski norm of a scene is achromatic.

The General Grey World (gGW) algorthm [2], is based on the assumption that the $n-th$ Minkowski norm of a scene after the application of a smoothing filter is achromatic.

The first order Gray Edge (GE1) algorithm [37], is based on the assumption that the $n-th$ Minkowski norm of the first order derivative in a scene is achromatic.

The second order Gray Edge (GE2) algorithm [37], is based on the assumption that the th Minkowski norm of the second order derivative in a scene is achromatic.

Image understanding for color constancy

As seen in the previous section, the hypotheses on which traditional color constancy algorithms rely are simple, and when they are not met, the algorithms can badly fail. In order to create more robust algorithms, some authors have started to employ computer vision results and techniques in order to exploit assumptions that could be met in a larger number of images. This section chronologically introduces such algorithms.

Exploiting high level visual information

Van De Weijer et al. [38] in 2007 proposed the use of high-level visual information to select the most plausible illuminant out of a set of possible illuminants. They achieved this by restating the problem in terms of semantic interpretability of the image. The idea is that the illuminant which results in the most likely image interpretation is the more likely to be correct. Image interpretation is meant in terms of semantics: an image where the sky is blue and located in the top of the image, and the road is grey and located in the bottom, can be considered more likely than an image with purple grass surrounding a reddish cow.

In their work they applied to the input image several color constancy methods in order to generate a set of illuminant hypotheses. For each generated illuminant hypothesis, they correct

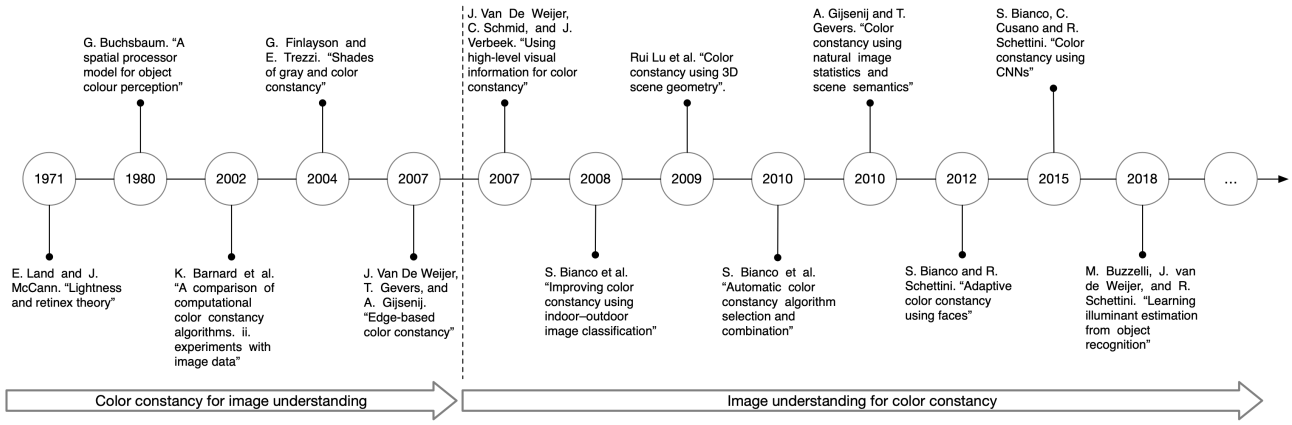

Figure 1. *Timeline representing the change in the paradigm of the color constancy algorithms: from the use of color constancy for image understanding to the used of image understanding techniques for the design of more robust color constancy algorithms.*

the image, and evaluate the likelihood of the semantic content of the corrected image. Finally, the most likely illuminant color is selected.

In addition, they extend the set of illuminant hypotheses with a set of top-down hypotheses based on the assumption that the average reflectance of semantic classes in an image is equal to the average reflectance of the semantic topic in the database. For each of the semantic classes present in the image they compute the illuminant which transforms the pixels assigned to this class in such a way that the average reflectance is in accordance with the average color of the class in the database. For example, a patch of grass which turned reddish in the evening light, will correctly hypothesize a red illuminant, since such an illuminant will transform it to green under white light.

Exploiting indoor-outdoor image classification

Bianco et al. [12] in 2008 investigated the idea that the effectiveness of automatic illuminant estimation techniques may be improved if information about the content of the images is taken into account. To this end, they designed an illuminant estimation approach which exploits the information provided by an image classifier. As suggested by Szummer and Picard [36], they considered the indoor and outdoor classes, which correspond to categories of images with different content, usually taken under different illumination conditions. Therefore their assumption is that these two classes of images may require different color processing procedures.

They describe each image by a set of low-level features related to color, texture, and edge distribution. The extracted features are stacked in a feature vector and fed to a decision forest trained to distinguish between indoor and outdoor images. Information about color distribution is captured by spatial color moments: they transform the image into the YCbCr color space, divide it into seven horizontal bands, and compute the mean and the standard deviation of each of the three color bands. Since the YCbCr color space decorrelates luminance and chrominance components, it is commonly used in image classification tasks. The subdivision in horizontal bands adequately describes some characteristics which are very useful for indoor/outdoor classification (images with blue sky in the upper part, or green grass in the lower part). Color moments are less useful when the bands contain heterogeneous color regions. Therefore, a global color histogram has been selected as a second color feature. The RGB color space has been subdivided in 27 bins by a uniform quantization of each component in three ranges. To describe the most salient edges they used an 18 bin edge direction histogram (ten degrees for each bin): the gradient of the luminance image is computed using Gaussian derivative filters tuned to retain only the major edges. Only the points for which the magnitude of the gradient exceeds a set threshold contribute to the histogram. Texture information is extracted computing a set of features based on a multiresolution analysis. A three level wavelet transform of the luminance image is computed, yielding to ten different sub-bands. For each band, they compute the average absolute value of the coefficients and their standard deviation. Therefore each image is described by a feature vector of 107 components. For classification, they used decision trees built according to the CART methodology.

Exploiting image 3D stage geometry classification

Lu et al. [32] in 2009 exploited 3D geometry models to determine which color constancy method to use for the different geometrical regions found in images. To this end, they first classify images into stages, i.e. rough 3D geometry models. According to the stage models, images are divided into different regions using hard and soft segmentation. After that, the best color constancy algorithm is selected for each geometry segment. As a result, light source estimation is tuned to the global scene geometry. Their algorithm opens the possibility to estimate the remote scene illumination color, by distinguishing nearby light source from distant illuminants.

From the typical 3D scene geometries, i.e. stages, proposed in [33], the authors selected 13 different stages corresponding to the typical 3D geometrical models of the scenes occurring in the considered color constancy dataset.

The image description is based on the Bag-of-words representation, using as a visual descriptor a color modification of the SIFT descriptor. scene categorization. The classifier adopted is a 1-vs-all SVM classifier with a χ^2 kernel.

Directly exploiting image features

In order to bypass the manual definition of the image classes that may need a different correction, Bianco et al. [11] and Gijsenij et al. [25] independently tried in 2010 to directly predict the best algorithm to use on the input image.

More in detail, they investigated if it is possible to automatically derive the suitability of an illuminant estimation algorithm for a given image by analyzing a set of visual features. Given a set of illuminant estimation algorithms, their frameworks determine how the estimation of the illuminant of a given image should be computed.

In [11] the prediction of the suitability of each algorithm is carried out by an image classifier based on an ensemble of decision trees trained to identify the best algorithm in the considered set, on the basis of the values of a set of low-level visual features. These features for the most part are general-purpose features taken from the pattern recognition and image analysis fields. The remaining features were instead specifically designed for the illuminant estimation problem.

Since an image conveys information at different levels, in order to capture most of the image information, they use different features at the same time in order to have multiple representations which characterize the content from different perspectives. To describe the image content they considered two groups of low level features: general-purpose features and problem-dependent features. The general-purpose features are features that can be used on a large range of applications since they do not capture characteristic of the images that are problem specific. The features in this category that they selected are: color histogram, edge direction histogram, statistics on the wavelet coefficients, and color moments. The problem-dependent features they chose are: the number of different colors contained in the image, the percentage of color components that are clipped to the highest or lowest value that can be represented in the image color space, a cast index representing the extent of the presence of a color cast in the image, and the magnitudes of the edges.

Exploiting face detection

Inspired by the robustness and efficiency of face detectors Bianco and Schettini [10] in 2012 proposed a fully automatic method to exploit the skin color extracted from detected faces to estimate the illuminant in the scene. The method was further extended in 2014 [9] to cope also with multiple illuminants. Their method is based on three assumptions:

- skin colors form a sufficiently compact cluster in the color space in order to represent a valid clue for illuminant estimation [27];
- the illumination on each face is uniform;
- the illumination estimated on the faces properly samples the illumination distribution in the scene.

They conducted a preliminary analysis showing that by generating different queries on Flickr (http://www.flickr.com/) using very generic tags such as cameras and mobile phones manufacturers, among 30% and 60% of the returned images were portraits or included faces. More specific queries such as party, family, birthday, holiday, etc. usually contain a much higher number of faces. They also showed that using skin color for illuminant estimation is statistically equivalent to having a neutral patch in the image.

The first step in their method consists in running the face detector module on the input image to detect any faces. If no faces are detected, the input image may be processed with any other state-of-the-art illuminant estimation algorithm. If one or more faces are detected, a skin detection module [8] is run on the detected faces to filter out any non-skin and unreliable pixels. A local illuminant estimation is performed on the detected skin pixels of each face. If the maximum distance among the estimations on the different faces is lower than a fixed threshold, the local estimates are combined into a unique global illuminant estimate; otherwise the single face estimates are propagated to the rest of the image to give a local illuminant estimate for each pixel of the image.

Exploiting deep learning

Bianco et al. [6] in 2015 inspired by the success obtained by deep neural networks outperforming state-of-the-art approaches on various computer vision tasks, were the first to apply such techniques to color constancy.

They tried two different approaches: in the first one they extracted the 4096-dimensional feature vector from each image using a CNN trained on a large dataset (ILSVRC 2012) with image-level annotations to classify images into 1000 different classes; the extracted features are then used as input to a linear Support Vector Regressor (SVR) to estimate the illuminant color for each image. In the second approach they designed an ad-hoc CNN working on small image patches. This particularity led to its extension to estimate multiple illuminants [7]. Interestingly, in their analysis they showed that some neurons of their trained network behave like already existing methods in the state of the art: for instance, some neurons seem to fire on image edges (feature exploited for example by [37]), some neurons on highlights (feature related to the assumption exploited in [31]), some neurons on sky and bluish texture, some on vegetation and greenish texture (reminding the features exploited in [38]), and some on skin and orange/reddish texture (reminding the use of faces and skin in [10, 9]).

More recent approaches try to exploit the power of different deep learning architectures, such as for example Generative Adversarial Networks (GANs) [17, 15], or use different modules, such as for example Attention [29, 40].

Exploiting object recognition

Buzzelli et al. [14] in 2018 investigated to what extent one can learn illuminant estimation as a byproduct of an independent auxiliary task. In their work they consider object recognition.

In their work they describe a training process which uses only the labels for the auxiliary task but no illuminant ground truth whatsoever. During training, the input image first passes through an illuminant estimation network, which estimates the scene illuminant and corrects the image accordingly. The illuminant corrected image is then processed by an object classification network, which produces an estimation of the classes that are present in the image. By training both networks in an end-to-end fashion, they show how they can effectively train the illuminant estimation network without any illuminant ground truth. After the training is complete, this illuminant estimation network can then be independently applied to any other dataset, such as standard color constancy benchmarks, simply removing the object recognition network. Their method is particularly interesting since it is the first learned method for illuminant estimation which does not require illuminant annotations, an idea that has been further developed in [5].

Other approaches

Some researchers have reformulated the problem of illuminant estimation as a 2D spatial localization task in a chromaticity space [3, 4], thereby allowing them to apply techniques from object detection and structured prediction to the color constancy problem.

It is also worth mentioning that in parallel with the development of methods that employ computer vision results and techniques to obtain robust illuminant estimates, some authors designed alternative methods reaching competitive results. For example [19] obtained good results using a fixed bias correction of simple statistics-based approaches.

Future directions

In this article we showed the change in paradigm that occurred in color constancy algorithms. At the beginning color constancy was commonly used as a pre-processing step to help improving the performance of image understanding tasks. More recently instead, color constancy has started to exploit the successful results obtained in image understanding and computer vision in general to create more robust color constancy algorithms, as is it also evident from the trends of the methods that participated in very recent illuminant estimation challenges [18].

The development of robust color constancy algorithms able to cope with all the conditions that may occur in digital photography is very important, since color constancy is also a crucial step in the processing pipeline of digital cameras.

Furthermore we believe that this is a very promising research direction, that in the future will continue to exploit the recent successes in the image understanding field, such as for example the depth and 3D information, that have been experimentally proven to enable a better color constancy performance in human observers [28].

Another research direction that could be investigated is the exploitation of video understanding techniques to extend color constancy also to the video domain [34], and to the multispectral domain.

References

[1] Arash Akbarinia and C Alejandro Parraga. "Colour constancy beyond the classical receptive field". In: *IEEE transactions on pattern analysis and machine intelligence* 40.9 (2017), pp. 2081–2094.

[2] Kobus Barnard et al. "A comparison of computational color constancy algorithms. ii. experiments with image data". In: *IEEE transactions on Image Processing* 11.9 (2002), pp. 985–996.

[3] Jonathan T Barron. "Convolutional color constancy". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2015, pp. 379–387.

[4] Jonathan T Barron and Yun-Ta Tsai. "Fast fourier color constancy". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 886–894.

[5] Simone Bianco and Claudio Cusano. "Quasi-unsupervised color constancy". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 12212–12221.

[6] Simone Bianco, Claudio Cusano, and Raimondo Schettini. "Color constancy using CNNs". In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2015, pp. 81–89.

[7] Simone Bianco, Claudio Cusano, and Raimondo Schettini. "Single and multiple illuminant estimation using convolutional neural networks". In: *IEEE Transactions on Image Processing* 26.9 (2017), pp. 4347–4362.

[8] Simone Bianco, Francesca Gasparini, and Raimondo Schettini. "Adaptive skin classification using face and body detection". In: *IEEE Transactions on Image Processing* 24.12 (2015), pp. 4756–4765.

[9] Simone Bianco and Raimondo Schettini. "Adaptive color constancy using faces". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36.8 (2014), pp. 1505–1518.

[10] Simone Bianco and Raimondo Schettini. "Color constancy using faces". In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 2012, pp. 65–72.

[11] Simone Bianco et al. "Automatic color constancy algorithm selection and combination". In: *Pattern recognition* 43.3 (2010), pp. 695–705.

[12] Simone Bianco et al. "Improving color constancy using indoor–outdoor image classification". In: *IEEE Transactions on image processing* 17.12 (2008), pp. 2381–2392.

[13] Gershon Buchsbaum. "A spatial processor model for object colour perception". In: *Journal of the Franklin institute* 310.1 (1980), pp. 1–26.

[14] Marco Buzzelli, Joost van de Weijer, and Raimondo Schettini. "Learning illuminant estimation from object recognition". In: *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE. 2018, pp. 3234–3238.

[15] Marco Buzzelli et al. "Consensus-driven illuminant estimation with GANs". In: *Thirteenth International Conference on Machine Vision*. Vol. 11605. International Society for Optics and Photonics. 2021, p. 1160520.

[16] Yu-Hsiu Chen et al. "Filter-invariant image classification on social media photos". In: *Proceedings of the 23rd ACM international conference on Multimedia*. 2015, pp. 855–858.

[17] Partha Das et al. "Color constancy by GANs: An experimental survey". In: *arXiv preprint arXiv:1812.03085* (2018).

[18] Egor Ershov et al. "Illumination estimation challenge: The experience of the first 2 years". In: *Color Research & Application* 46.4 (2021), pp. 705–718.

[19] Graham D Finlayson. "Corrected-moment illuminant estimation". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2013, pp. 1904–1911.

[20] Graham D Finlayson and Elisabetta Trezzi. "Shades of gray and colour constancy". In: *Color and Imaging Conference*. Vol. 2004. 1. Society for Imaging Science and Technology. 2004, pp. 37–41.

[21] David H Foster. "Color constancy". In: *Vision research* 51.7 (2011), pp. 674–700.

[22] Hiren Galiyawala et al. "Person retrieval in surveillance video using height, color and gender". In: *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE. 2018, pp. 1–6.

[23] Shao-Bing Gao et al. "Color constancy using double-opponency". In: *IEEE transactions on pattern analysis and machine intelligence* 37.10 (2015), pp. 1973–1985.

[24] Shao-Bing Gao et al. "Combining bottom-up and top-down visual mechanisms for color constancy under varying illumination". In: *IEEE Transactions on Image Processing* 28.9 (2019), pp. 4387–4400.

[25] Arjan Gijsenij and Theo Gevers. "Color constancy using natural image statistics and scene semantics". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.4 (2010), pp. 687–698.

[26] Arjan Gijsenij, Theo Gevers, and Joost Van De Weijer. "Computational color constancy: Survey and experiments". In: *IEEE transactions on image processing* 20.9 (2011), pp. 2475–2489.

[27] Thorsten Hansen et al. "Memory modulates color appearance". In: *Nature neuroscience* 9.11 (2006), pp. 1367–1368.

[28] Monika Hedrich, Marina Bloj, and Alexa I Ruppertsberg. "Color constancy improves for real 3D objects". In: *Journal of Vision* 9.4 (2009), pp. 16–16.

[29] Karlo Koščević, Marko Subašić, and Sven Lončarić. "Attention-based convolutional neural network for computer vision color constancy". In: *2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA)*. IEEE. 2019, pp. 372–377.

[30] J von Kries. *Chromatic Adaptation, Festschrift der Albercht-Ludwig-Universität*. 1902.

[31] Edwin H Land and John J McCann. "Lightness and retinex theory". In: *Josa* 61.1 (1971), pp. 1–11.

[32] Rui Lu et al. "Color constancy using 3D scene geometry". In: *2009 IEEE 12th International Conference on Computer Vision*. IEEE. 2009, pp. 1749–1756.

[33] Vladimir Nedovic et al. "Depth information by stage classification". In: *2007 IEEE 11th International Conference on Computer Vision*. IEEE. 2007, pp. 1–8.

[34] Yanlin Qian et al. "A benchmark for burst color constancy". In: *European Conference on Computer Vision*. Springer. 2020, pp. 359–375.

[35] Christian Riess and Elli Angelopoulou. "Scene illumination as an indicator of image manipulation". In: *International Workshop on Information Hiding*. Springer. 2010, pp. 66–80.

[36] Martin Szummer and Rosalind W Picard. "Indoor-outdoor image classification". In: *Proceedings 1998 IEEE International Workshop on Content-Based Access of Image and Video Database*. IEEE. 1998, pp. 42–51.

[37] Joost Van De Weijer, Theo Gevers, and Arjan Gijsenij. "Edge-based color constancy". In: *IEEE Transactions on image processing* 16.9 (2007), pp. 2207–2214.

[38] Joost Van De Weijer, Cordelia Schmid, and Jakob Verbeek. "Using high-level visual information for color constancy". In: *2007 IEEE 11th International Conference on Computer Vision*. IEEE. 2007, pp. 1–8.

[39] Javier Vazquez-Corral et al. "Color constancy by category correlation". In: *IEEE Transactions on image processing* 21.4 (2011), pp. 1997–2007.

[40] Tian Yuan and Xueming Li. "Full Convolutional Color Constancy with Attention". In: *Chinese Conference on Image and Graphics Technologies*. Springer. 2020, pp. 114–126.

Author Biography

Simone Bianco obtained the PhD in Computer Science at DISCo (Dipartimento di Informatica, Sistemistica e Comunicazione) of the University of Milano-Bicocca, Italy, in 2010. He obtained the BSc and the MSc degree in Mathematics from the University of Milano-Bicocca, Italy, respectively in 2003 and 2006. He is currently Associate Professor and his research interests include computer vision, machine learning, optimization algorithms, and color imaging.

Marco Buzzelli obtained a bachelor degree and masters degree in Computer Science respectively in 2012 and 2014, focusing on image processing and computer vision tasks. He received a PhD in Computer Science in 2019 at the University of Milano, Bicocca, Italy, where he is currently employed as a post-doctoral researcher. His main topics of research include characterization of digital imaging devices, and object recognition in complex scenes.