Enforcing Temporal Consistency for Color Constancy in Video Sequences

Marco Buzzelli(✉), Claudio Rota, Simone Bianco, and Raimondo Schettini

Department of Informatics Systems and Communication,
University of Milano-Bicocca, Milan, Italy
marco.buzzelli@unimib.it

Abstract. This paper focuses on enhancing temporal color constancy in video sequences, ensuring that the result not only achieves color accuracy frame-by-frame but is also consistent over time. Our approach consists of a three-step process: per-frame illuminant estimation and correction, video stabilization to ensure temporal consistency, and consensus-based illuminant correction. By employing consensus-driven illuminant estimation over the result of temporal stabilization, we effectively mitigate spatial artifacts and concurrently enhance the overall stability of the sequence. Our method is tested using the Gray Ball and BCC datasets, showing the potential of integrating temporal stabilization with color correction processes to enhance the visual continuity of video content. While our primary objective is to reduce temporal flickering, a significant side effect of our approach is the improvement of color constancy accuracy across frames.

Keywords: Color constancy · Temporal consistency · Temporal stabilization · Automatic white balance. · Video sequences

1 Introduction

Color constancy is the perceptual property of the visual system that ensures the colors of objects remain relatively constant under varying illumination conditions [11]. This process is crucial in digital imaging and is typically composed of two major steps: illuminant estimation and illuminant correction [7]. Illuminant estimation involves determining the dominant light source color in an image, while illuminant correction adjusts the colors in the image to appear as they would under a neutral light source.

Historically, the application of color constancy has been primarily focused on still images. Today, with the increasing availability of devices able to acquire video, the application of color constancy to video sequences, known as temporal color constancy, presents both new opportunities and challenges.

The principal opportunity presented by temporal color constancy is that the abundance of frames in video sequences allows the problem to be more

R. Schettini et al. (Eds.): CCIW 2024, LNCS 15193, pp. 274–288, 2025.
https://doi.org/10.1007/978-3-031-72845-7_20

effectively constrained. Single frame color constancy is inherently an ill-posed problem, characterized by the potential for multiple viable solutions. Typically, additional information is exploited to constrain this set of solutions, using for example assumptions on the scene content [18], knowledge derived from training data [6], or the physical plausibility of different solutions [9]. A straightforward assumption often utilized is the "gray world hypothesis", which posits that the average reflectance in a scene is achromatic. In the case of temporal color constancy, the additional information crucial for determining illuminant conditions in a given input often comes from the availability of multiple adjacent frames depicting a very similar scene content within a video sequence. This inter-frame assistance enables a more robust estimation of the lighting conditions of individual frames, as frames with clearer illuminant indicators can compensate for neighboring frames where such cues are sparse, ambiguous, or even misleading.

As previously noted, transitioning to the domain of video introduces significant challenges, particularly regarding the consequences of incorrect illuminant estimation and correction in the individual frames. In video sequences, the impact of such errors is, in fact, compounded: not only does each frame suffer from inaccurate color correction, but inconsistencies in the corrections applied across adjacent frames can also lead to flickering. This phenomenon disrupts the visual continuity and creates a distracting and unnatural appearance in the video stream. Avoiding flickering is particularly important in scenarios such as social video sharing, film production, surveillance systems, and other multimedia applications where changes in color perception can be annoying, distracting or even misleading.

Image color constancy is a problem clearly defined as involving a single input image, that is analyzed and adjusted to produce one output image with corrected colors. This approach operates under the principle of one input corresponding to one output, focusing solely on the colors within that specific image. Temporal color constancy, on the other hand, is a much less mature field of research, and is consequently less clearly-defined. In this paper, we distinguish between two possible sub-categories of temporal color constancy, resulting from the different nature of existing datasets [8,17]:

- Video Color Constancy: In this scenario, each frame of a video is processed in consideration of its preceding frames. This method requires adjusting each frame potentially with a different illuminant, i.e. multiple inputs each with its own corrected output. For specific applications such as offline video processing, it is also possible to include subsequent frames in the elaboration of the current one, thus breaking the causality constraint.
- Burst Color Constancy: Unlike video color constancy, burst color constancy treats a sequence of frames as a cohesive whole. All frames collectively inform the correction process, aiming to achieve a uniform illuminant correction across the entire burst. This model considers multiple inputs but produces a singular, consistent output for the whole sequence.

Figure 1 offers a visual representation of the aforementioned concepts, with "per-frame color constancy" bridging the gap between image color constancy and temporal color constancy, at the expense of introducing flickering artifacts.

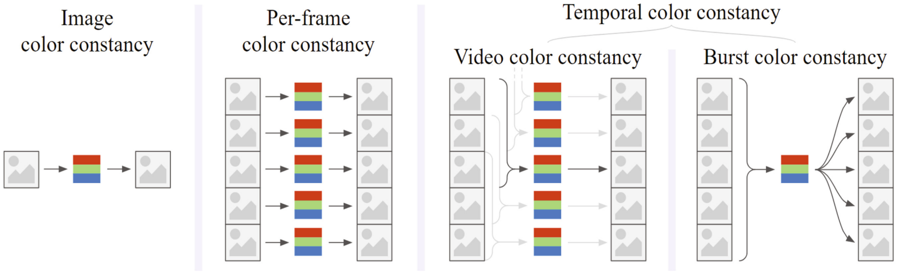

Fig. 1. Visual representation of different applications of computational color constancy to images and video sequences.

There have been some methodologies developed to address temporal color constancy, discussed in Sect. 2. By considering several frames together, these approaches inherently manage the aforementioned issue of flickering.

Alternatively, another approach involves applying image color constancy techniques to each individual frame of a video sequence, followed by a stabilization procedure to process the results and to enforce temporal consistency. Buzzelli and Erba [4] conducted research that delved into the applicability of image color constancy in the temporal domain. They characterized various single-frame methods with respect to their temporal stability. Building on this foundational work, we implement stabilization as a post-processing step to enforce temporal consistency. We evaluate the effectiveness of this approach on the Gray Ball and Burst Color Constancy datasets, and provide several considerations based on the outcomes of our analysis. The advantages of this approach are twofold. First, by decoupling the illuminant correction process from the video stabilization process, it is possible to select, and combine, the best algorithms from both tasks. This includes well-established methods for single-frame color constancy. Secondly, if we resort to a task-independent (agnostic) method for temporal consistency, the same method may also be used to stabilize other aspects of the processed video, correcting for example any artifacts also introduced by per-frame contrast enhancement algorithms.

2 Related Works

In this section we explore the state of the art for temporal color constancy methods, which explicitly account for the temporal dimension of video sequences to produce per-frame or per-sequence illuminant correction.

Wang et al. [19] enhanced illumination estimation for video sequences by exploiting correlations between adjacent frames. Their method segments videos into scenes, assuming a consistent illuminant for all frames within a scene. This scene-based approach aggregates illuminant data across frames to derive a more stable and accurate chromaticity estimate for each scene. Their experimental results show that this method outperforms traditional single-frame algorithms by effectively using multi-frame information to maintain illuminant consistency across video shots.

Barron et al. [2] introduced the Fast Fourier Color Constancy (FFCC), a novel algorithm that significantly improves upon the traditional methods of illuminant estimation by transforming it into a spatial localization task on a torus in the frequency domain. Specifically for temporal color constancy, they have adapted FFCC to handle video sequences more effectively. They incorporate a smoothing model inspired by the Kalman filter to mitigate errors in individual frame predictions. This enhancement allows FFCC to maintain consistency and accuracy across frames, addressing the common challenge of flickering in video caused by frame-to-frame illuminant estimation variability.

Qian et al. [15] introduced a novel approach to temporal color constancy by considering multiple preceding frames for illuminant estimation. Their method utilizes an end-to-end trainable network, RCC-Net, which incorporates convolutional LSTMs to capture compositional representations over time and space effectively. By adapting the SFU Gray Ball Dataset for a temporal context, they demonstrate that RCC-Net consistently surpasses both the traditional single-frame algorithms and their temporal adaptations in terms of performance. This success is attributed to the network's ability to exploit sequential frame information, enhancing the accuracy of color constancy across video sequences.

Later, Qian et al. [17] expanded the scope of color constancy research by introducing a benchmark for Burst Color Constancy (CC), with a method that uses multiple frames from a sequence to estimate the illumination color of a shot frame, challenging the traditional single-frame approach. Their benchmark comprises 600 real-world sequences captured with a high-resolution mobile phone camera, a fixed train-test split for consistent evaluation, and a baseline method that demonstrates high accuracy across this new and previous datasets. Their study also reports results for over 20 well-known color constancy methods, including recent state-of-the-art developments. Additionally, the method has been enhanced by integrating a more robust backbone network for semantic feature extraction and a 2D LSTM for more effective spatial recurrent processing.

Buzzelli and Erba [4] addressed the limitations of traditional evaluations of color constancy algorithms, which typically rely on angular error analysis of static images. Recognizing the growing use of video in consumer technology, they proposed an expanded evaluation framework that includes temporal and spatial stability. This approach assesses how well these algorithms perform under varying scene conditions unrelated to illuminant changes, like motion of subjects or the camera. Using stable sequences from the Gray Ball and Burst Color Constancy datasets, their analysis revealed that while some algorithms excel in minimizing angular errors, they may falter on stability metrics. This work emphasizes the need for multi-faceted evaluation criteria in the realm of color constancy, particularly for video applications.

Zini et al. [21] introduced COCOA, a novel strategy for combining color constancy algorithms through a compact neural network architecture. This architecture processes and amalgamates the illuminant estimations from various algorithms, each based on different assumptions about the input scene content. COCOA is versatile, applicable both to single-frame images and video sequences, the latter utilizing a Long Short-Term Memory (LSTM) module to manage varying-length sequences. The approach is tested using only learning-free algorithms that rely on simple image statistics, experimenting on the Shi-Gehler and NUS datasets for still images, and the Burst Color Constancy dataset for videos. Their results indicate that COCOA surpasses other combination strategies in performance, achieving illuminant estimation accuracy on par with more complex and computationally intensive methods. Additionally, the effectiveness of COCOA is maintained even with fewer training instances, and the study includes an assessment of each contributing method's impact on the final estimation accuracy.

3 Proposed Method

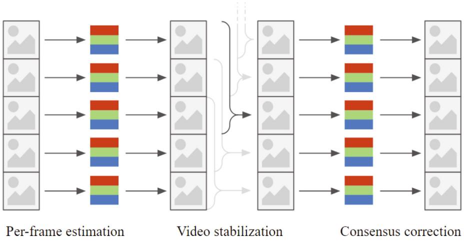

Fig. 2. Visual representation of the proposed three-step method to enforce temporal consistency for color constancy in video sequences.

Let $S = \{x_1, x_2, ..., x_i, ...\}$ be a sequence of frames x. By targeting video color constancy, our objective is to produce an illuminant estimation y_i for each frame, in the form of an RGB triplet. This relationship can be expressed through the function f, mapping the sequence of frames to their corresponding illuminant estimations:

$$E = f(S) = \{y_1, y_2, ..., y_i, ...\}. \tag{1}$$

We address the task through a three-step process: per-frame correction, temporal stabilization, consensus-based correction, as represented in Fig. 2.

In the first step, we generate a preliminary estimation y'_i for each frame, using a single-frame method g, and then we apply the corresponding illuminant correction with function h:

$$x'_i = h(x_i, g(x_i)) = h(x_i, y'_i) \quad \forall x_i \in S, \tag{2}$$

which produces a potentially-unstable sequence $S' = \{x'_1, x'_2, ..., x'_i, ...\}$. We refer the reader to Sect. 4.2 for the specific selection of methods g used in our experimental setup. As described by Eq. 2, before moving to the second step it is essential to apply the estimation as a correction on the original input x, because the temporal stabilization methods used in this work are designed to work with images (frames x'_i) rather than direct estimations (RGB triplets y'_i). The correction function h includes a von Kries-like diagonal transform [13], as well as gamma delinearization. These adjustments help produce images that more closely resemble sRGB outputs, which are better suited for the general-purpose temporal stabilization methods employed.

In the second step, we stabilize the sequence of corrected frames x'_i using a video stabilization method m:

$$S'' = m\left(S'\right) = \{x''_1, x''_2, ..., x''_i, ...\}. \tag{3}$$

We refer the reader to Sect. 4.3 for the specific selection of methods m used in our experimental setup. Theoretically, this process yields a stabilized sequence of frames. However, practical challenges may arise:

1. Local artifacts could be introduced by enforcing temporal consistency, potentially affecting the visual quality of the sequence.
2. The output from this two-step process cannot be directly compared against the estimations from the initial single-frame method using standard color constancy metrics.

To address these issues, in the third and final step we extract a per-frame stabilized illuminant by consensus [5]: we calculate the average illuminant for each frame by taking the ratio of the corrected frame x''_i to the original frame x_i and then averaging this ratio:

$$y_i = \frac{1}{|x_i|} \sum_{p \in x_i} \left(\frac{x''_i(p)}{x_i(p)} \right) \quad \forall x_i \in S. \tag{4}$$

Each original frame x_i can then be re-corrected globally using the stabilized illuminant y_i with the application of an adjusted von Kries diagonal transform.

4 Experimental Setup

4.1 Selected Datasets

Gray Ball. The Gray Ball dataset [8] is one of the few datasets potentially suitable for video color constancy. It comprises 11,346 images organized into 15 sequences, with many shots captured at close intervals. The dataset was recorded using a Sony VX-2000 digital video camera. Although it is an older dataset and lacks raw images, we follow Buzzelli and Erba's approach by processing a linear version of the dataset for illuminant estimation and correction. Concerning the presence of video cuts, the original sequences were manually curated into 337 smaller sequences to ensure only smooth transitions in scene content. From these, 168 sequences deemed stable were selected for further analysis, based on the extraction of MIC and STD metrics on the ground truth illuminants (see Eqs. 5 and 7). According to Buzzelli and Erba, the most and least stable color constancy methods on the Gray Ball dataset are found to be Second-order Gray Edge (GE2) [18] and the Grayness Index (GI) [16], respectively, from a benchmark of 11 methods. For the purposes of this paper, we will focus on evaluating the performance of these two extremes.

Burst Color Constancy (BCC). The BCC dataset [17], also known as the Temporal Benchmark dataset, was introduced by Qian et al. for burst color constancy. It comprises 600 sequences of varying lengths, ranging from 3 to 17 frames. These are divided into 400 sequences for training and 200 sequences designated for the test set, the latter of which we utilize in our analysis. The images were captured using a Huawei Mate 20 Pro mobile phone and are stored in a proprietary 16-bit RAW format. Additionally, the dataset includes 8-bit PNG images for each sequence. A SpyderCube calibration target was placed in the scene immediately after the sequence acquisition to provide an out-of-sequence reference shot, representing the entire video sequence. According to Buzzelli and Erba, the most and least stable methods for color constancy on the BCC dataset are Sensor-Independent illumination Estimation (SIIE) [1] and White Point (WP) [18], respectively. For the purposes of this paper, we will focus on evaluating the performance of these two extremes.

4.2 Selected Color Constancy Methods

Based on the preliminary experiments reported in Sect. 4.1, we have focused our experimentation on the following algorithms for single-frame computational color constancy.

White Point (WP). The White Point method for color constancy is based on the assumption that the brightest color present in a scene, is white under a neutral illuminant. Although this method is straightforward and computationally efficient, its performance can vary significantly depending on the presence of truly neutral colors in the scene.

Second-order Gray Edge (GE2). Developed by Van de Weijer et al. in 2007 [18], the Second-order Gray Edge method extends the gray world hypothesis by considering higher-order derivatives of color edges in the image. This approach posits that changes in edge colors in a scene are predominantly due to variations in the scene illumination. By analyzing the second-order derivatives, this method aims to more accurately estimate the global illuminant than simpler methods.

Grayness Index (GI). Introduced by Qian et al. in 2019 [16], the Grayness Index method focuses on identifying pixels within an image that appear to be achromatic or'gray' under a wide range of lighting conditions. These gray pixels are assumed to reflect the true color of the illuminant, as they exhibit minimal hue but varying intensities. By aggregating the colors of these gray pixels, the method estimates the scene's illuminant.

Sensor-independent Illumination Estimation (SIIE). Proposed by Afifi and Brown in 2019 [1], this method aims to provide a robust solution to the color constancy problem that is independent of camera sensors. The approach uses a deep neural network trained on a dataset comprising multiple cameras with diverse spectral sensitivities. By doing so, it learns to generalize illumination estimation across different imaging conditions without being biased towards any specific camera sensor characteristics.

4.3 Selected Temporal Consistency Methods

For the purpose of this paper, we will focus on two well-known algorithm for task-agnostic (blind) temporal consistency, whose publicly-available code allows us to experiment their applicability in the domain of color constancy.

Lai et al. 2018 Lai et al. [14] introduced an end-to-end solution utilizing a deep recurrent network to enforce temporal consistency across video sequences. Their method operates by processing both the original unprocessed video and the individually processed frames to output a video with enhanced temporal consistency. This approach is designed to be algorithm-agnostic, meaning it does not depend on the specifics of the image processing techniques applied to the original video. The network is trained to minimize both short-term and long-term temporal losses as well as perceptual loss, balancing temporal stability with perceptual similarity to the processed frames. Notably, their method operates without the need for optical flow computations, allowing for real-time performance even on high-resolution videos. The approach was tested on various tasks such as artistic style transfer, enhancement, colorization, image-to-image translation, and intrinsic image decomposition.

TDMS-Net. Zhou et al. [20] framed temporal consistency as a temporal denoising problem aimed at mitigating flickering in previously unstable pre-processed frames. To tackle this, they introduce a novel model called the Temporal Denoising Mask Synthesis Network (TDMS-Net). This network is designed to synthesize temporally consistent frames by jointly predicting a motion mask, soft optical flow, and a refining mask. The approach taken by Zhou et al. learns temporal consistency directly from the original video and its temporal features, which are then used to refine the output frames.

5 Results

5.1 Assessing Temporal Stabilization

Buzzelli and Erba [4] refer to a temporally stable sequence as a sequence that:

1. Does not contain video cuts.
2. Does not involve abrupt illuminant changes.
3. Does not span a wide set of illuminants (even if gradually changing).

Point 1 is addressed at dataset selection level: it involves manual processing for the Gray Ball dataset, as described in Sect. 4.1, while it is valid by assumption for the BCC dataset.

The identification of abrupt illuminant changes is achieved by quantifying the"Maximum Illuminant Change" (MIC) in the estimated illuminant $y = (y^{(R)}, y^{(G)}, y^{(B)})$ between consecutive frames. More precisely, for each pair of consecutive frames in a sequence S, we measure the angular error between their ground truth illuminants and then we select the maximum of such errors:

$$\text{MIC}\left(S\right) = \max\left(\text{err}\left(y_i, y_{i+1}\right)\right), \quad i = 1...|S| - 1. \tag{5}$$

As angular error, we refer to the recovery error [12], which computes the angle between the RGB vector representing the ground truth illuminant U with the RGB vector representing the estimated illuminant V:

$$\text{err}\left(U, V\right) = \arccos\left(\frac{U \cdot V}{||U||||V||}\right), \tag{6}$$

where "$\cdot$" is the dot product, and "$||...||$" the Euclidean norm. Alternatively, the reproduction error [10] may be also used.

To identify sequences spanning a large range of illuminants, the chromaticity Standard Distance (STD) is used as a metric for scatteredness. Specifically, we first convert the ground truth illuminants into Angle-Retaining Chromaticity (ARC) [3]: a 2-dimensional representation where Euclidean distances correspond to angular distances in the original RGB space. Then, we compute a 2-dimensional generalization of the concept of standard deviation, defined as:

$$\text{STD}(S) = \sqrt{\sum_{i=1}^{|S|} \frac{(\alpha_{xi} - \overline{\alpha_x})^2}{|S|} + \sum_{i=1}^{|S|} \frac{(\alpha_{yi} - \overline{\alpha_y})^2}{|S|}}, \tag{7}$$

where $(\alpha_{xi}, \alpha_{yi})$ are the ARC coordinates of the i-th illuminant of sequence S, and $(\overline{\alpha_x}, \overline{\alpha_y})$ is the average of each ARC coordinate for the sequence.

5.2 Quantitative Results

In Table 1 we report aggregate results for the enforcement and assessment of temporal consistency in the domain of temporal color constancy. Specifically, we focus on three groups of images: the Gray Ball dataset (split into stable and unstable sequences following the first criterion from Sect. 5.1), and the BCC dataset. For each, we analyze the effect on the most stable algorithms for computational color constancy as highlighted in Sect. 4.1 (GE2 [18] and GI [16] respectively for Gray Ball, and SIIE [1] and WP [18] respectively for BCC). We first extract statistics on these algorithms, labelled as "Original", measuring both color constancy accuracy via the angular error ('err' from Eq. 6), and baseline temporal consistency via STD and MIC from Eq. 7 and Eq. 5 respectively. Focusing only on temporal consistency, in fact, would provide an incomplete

Table 1. Average statistics, expressed in degrees, for the impact of temporal consistency in temporal color constancy. The lower, the better. Best results in bold.

		Most stable algorithm			Least stable algorithm		
		GE2 [18]			GI [16]		
		err	SID	MIC	err	SID	MIC
Gray Ball(stablesequences)	Original	**5.098**	**1.139**	**1.896**	7.542	4.178	10.117
	Lai et al. [14]	5.464	1.737	3.040	7.266	3.779	7.364
	TDMSNet [20]	5.194	1.373	2.241	**7.235**	**3.589**	**7.347**
	(Δ Lai et al.)	(+0.366)	(+0.598)	(+1.144)	(-0.276)	(-0.399)	(-2.752)
	(Δ TDMSNet)	(+0.096)	(+0.234)	(+0.345)	(-0.307)	(-0.590)	(-2.770)
		GE2 [18]			GI [16]		
		err	SID	MIC	err	SID	MIC
Gray Ball(unstablesequences)	Original	**6.628**	**2.258**	3.**550**	7.208	4.984	12.969
	Lai et al. [14]	7.201	2.668	5.009	6.990	4.795	10.506
	TDMSNet [20]	6.646	2.539	4.141	**7.205**	**4.616**	**10.194**
	(Δ Lai et al.)	(+0.573)	(+0.410)	(+1.459)	(-0.219)	(-0.189)	(-2.463)
	(Δ TDMSNet)	(+0.018)	(+0.281)	(+0.591)	(-0.003)	(-0.368)	(-2.775)
		SIIE [1]			WP [18]		
		err	SID	MIC	err	SID	MIC
BCC	Original	4.494	2.561	4.971	6.986	5.128	11.253
	Lai et al.	4.483	2.594	4.817	6.432	4.227	7.961
	TDMSNet	**4.389**	**2.433**	**4.479**	**6.098**	**3.767**	**7.081**
	(Δ Lai et al.)	(-0.011)	(+0.033)	(-0.154)	(-0.554)	(-0.901)	(-3.292)
	(Δ TDMSNet)	(-0.105)	(-0.127)	(-0.492)	(-0.888)	(-1.361)	(-4.172)

view of the overall performance, since an hypothetical "Do-nothing" algorithm would produce results that are perfect in terms of temporal consistency, but useless in terms of actual color rendition. From this baseline, we then evaluate the effect of enforcing temporal consistency through the method by Lai et al., and TDMSNet, following the methodology defined in Sect. 3.

Considering that, for all involved measures, a lower result is better, we can observe how temporal consistency is always improved for color constancy algorithms that were originally found to be extremely unstable. On the other hand, already-stable algorithms appear to be negatively affected (on average) by enforcing temporal consistency in the Gray Ball dataset, and produce mixed results on the BCC dataset. Finally, it may be observed that, despite not actively working to improve the angular error, temporal consistency methods Lai et al. and TDMSNet also have a beneficial impact on angular error itself as a byproduct of temporal stabilization: this is due to the inherent removal of outliers in the sequence of per-frame estimated illuminants.

5.3 Qualitative Results

In addition to the quantitative analysis, we report in the following some visualizations for a qualitative assessment of the enforcement of temporal consistency.

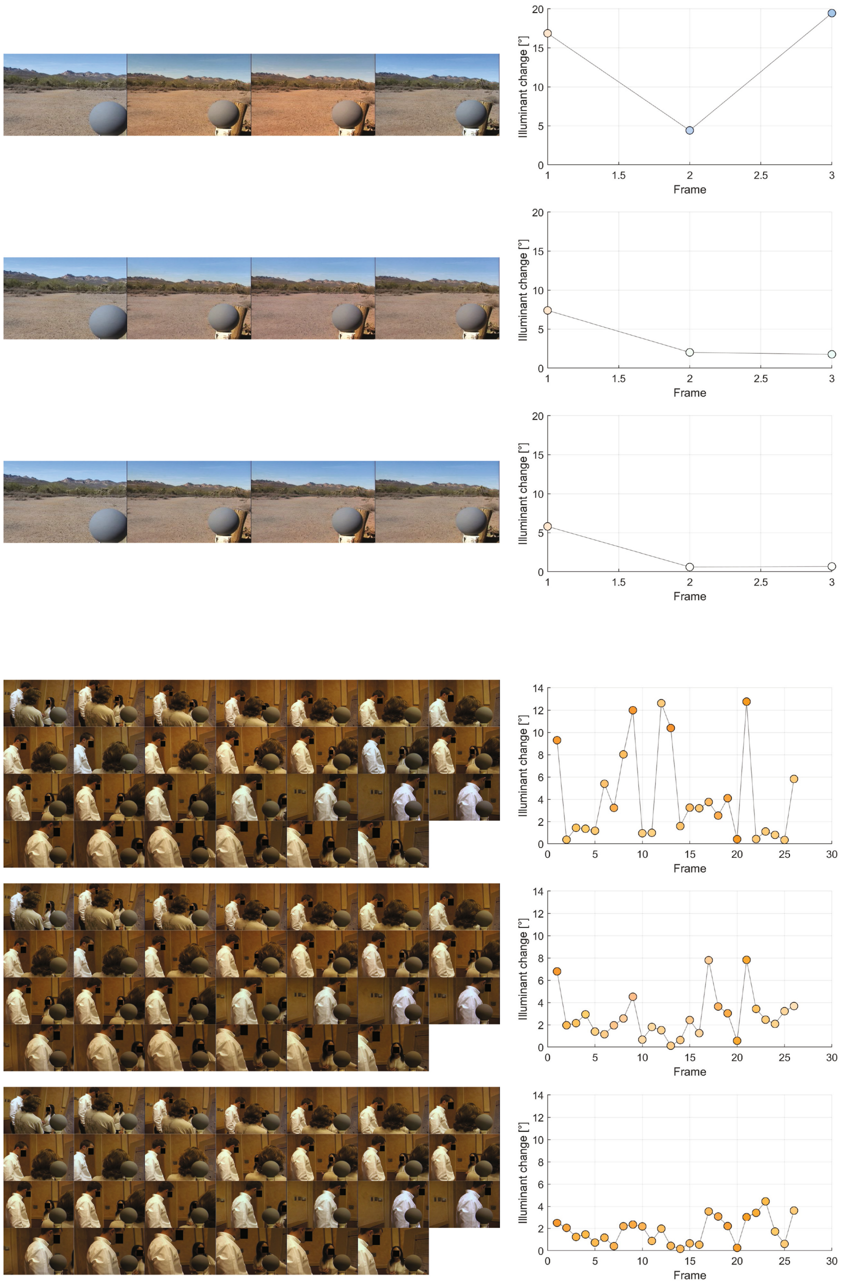

Fig. 3. Temporal stabilization on sample sequences. For each, we show the original sequence (top), the output as stabilized with Lai et al. (middle), and the output as stabilized with TDMSNet (bottom). We show the sequences on the left, and the plots of illuminant change (Eq. 5) on the right.

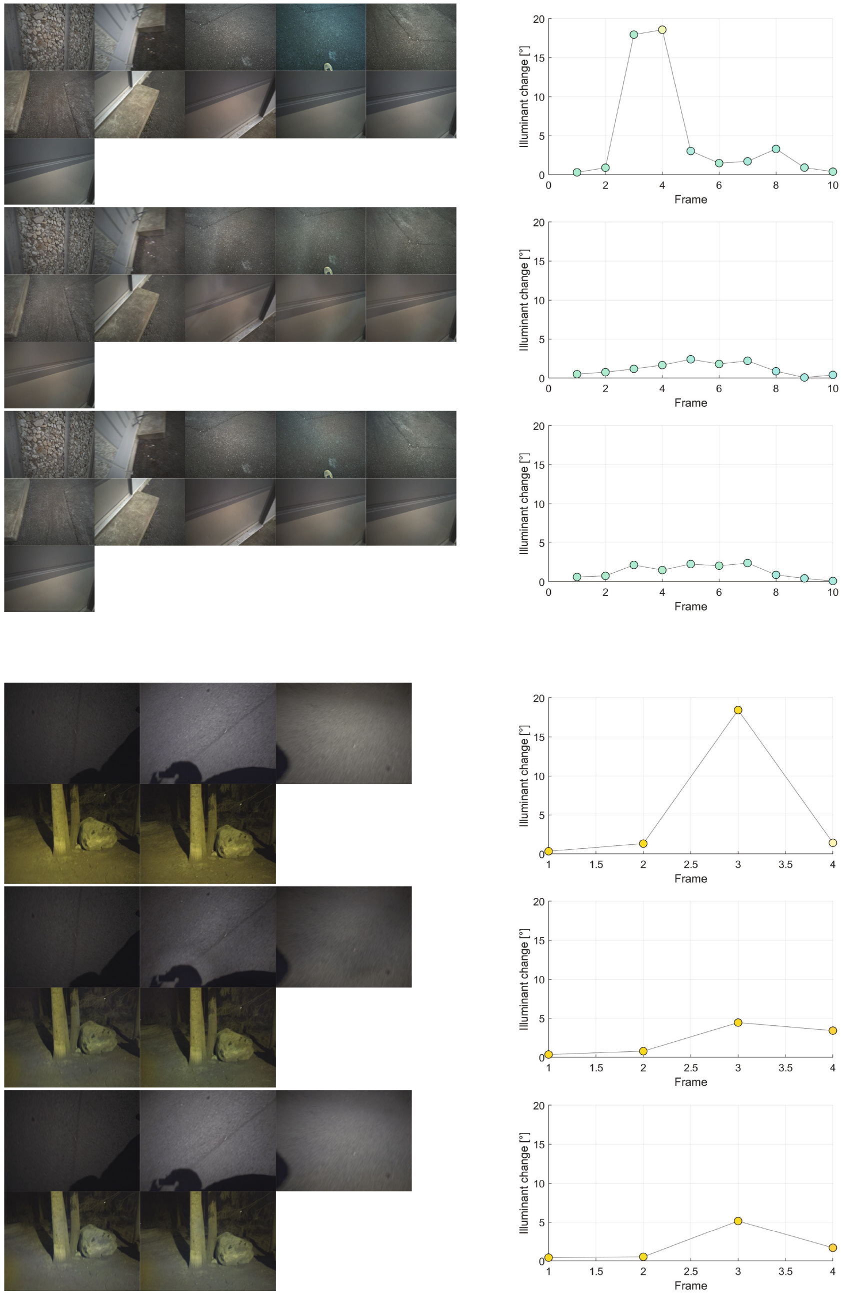

Fig. 4. Temporal stabilization on sample sequences. For each, we show the original sequence (top), the output as stabilized with Lai et al. (middle), and the output as stabilized with TDMSNet (bottom). We show the sequences on the left, and the plots of illuminant change (Eq. 5) on the right.

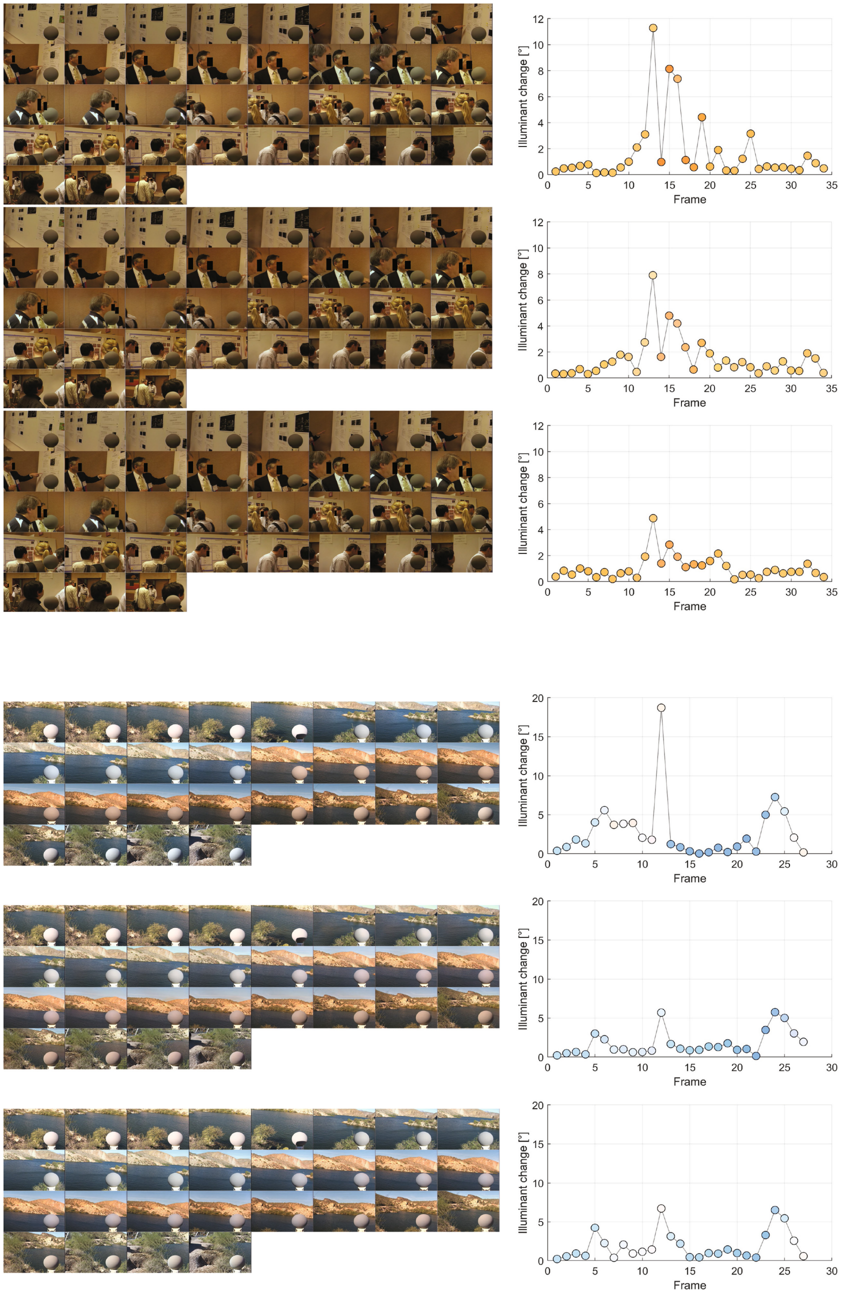

Fig. 5. Temporal stabilization on sample sequences. For each, we show the original sequence (top), the output as stabilized with Lai et al. (middle), and the output as stabilized with TDMSNet (bottom). We show the sequences on the left, and the plots of illuminant change (Eq. 5) on the right.

Figures 3, 4 and 5 present a selection of sample sequences on the left, and the corresponding plots of illuminant change between consecutive frames on the right (whose maximum corresponds to MIC). For each, we show the original sequence (top), the output as stabilized with Lai et al. (middle), and the output as stabilized with TDMSNet (bottom). The second sequence in Fig. 3 showcases a clear result where temporal stability is improved with respect to the original sequence, since the tint on the main subject's shirt is more stable. Nonetheless, some variability is still observable, as also documented by the corresponding illuminant change plot.

6 Conclusions

In this study, we have explored the challenges of achieving temporal color constancy in video sequences, a critical area in the advancement of automatic white balance and color correction. Our proposed method, which combines per-frame color correction with advanced temporal stabilization techniques, addresses the inherent variability and instability in video illuminant estimation. Through experiments conducted using the Gray Ball and BCC datasets, our approach demonstrated improvements in maintaining color consistency across video frames compared to traditional single-frame methods, localized to sequences that were originally very unstable. The use of consensus-driven illuminant correction ensured that no spatial artifacts are introduced by the temporal stabilization step.

Future work will focus on refining these techniques and exploring their applications in more diverse scenarios, including real-time video processing and integration into consumer video devices.

Acknowledgments. This work was partially supported by the MUR under the grant "Dipartimenti di Eccellenza 2023–2027" of the Department of Informatics, Systems and Communication of the University of Milano-Bicocca, Italy.

References

1. Afifi, M., Brown, M.S.: Sensor-independent illumination estimation for dnn models. arXiv preprint arXiv:1912.06888 (2019)
2. Barron, J.T., Tsai, Y.T.: Fast fourier color constancy. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 886–894 (2017)
3. Buzzelli, M., Bianco, S., Schettini, R.: Arc: Angle-retaining chromaticity diagram for color constancy error analysis. JOSA A **37**(11), 1721–1730 (2020)
4. Buzzelli, M., Erba, I.: On the evaluation of temporal and spatial stability of color constancy algorithms. JOSA A **38**(9), 1349–1356 (2021)
5. Buzzelli, M., Riva, R., Bianco, S., Schettini, R.: Consensus-driven illuminant estimation with gans. In: 13th ICMV. vol. 11605, pp. 578–584. SPIE (2021)
6. Buzzelli, M., Zini, S., Bianco, S., Ciocca, G., Schettini, R., Tchobanou, M.K.: Analysis of biases in automatic white balance datasets and methods. Color Res. Appl. **48**(1), 40–62 (2023)

7. Cheng, D., Abdelhamed, A., Price, B., Cohen, S., Brown, M.S.: Two illuminant estimation and user correction preference. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 469–477 (2016)
8. Ciurea, F., Funt, B.: A large image database for color constancy research. In: Color and Imaging Conference. pp. 160–164 (2003)
9. Ershov, E., Tesalin, V., Ermakov, I., Brown, M.S.: Physically-plausible illumination distribution estimation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12928–12936 (2023)
10. Finlayson, G.D., Zakizadeh, R.: Reproduction angular error: An improved performance metric for illuminant estimation. In: BMVC (2014)
11. Gehler, P.V., Rother, C., Blake, A., Minka, T., Sharp, T.: Bayesian color constancy revisited. In: 2008 IEEE Conference on CVPR. pp. 1–8. IEEE (2008)
12. Hordley, S.D., Finlayson, G.D.: Reevaluation of color constancy algorithm performance. JOSA A **23**(5), 1008–1020 (2006)
13. von Kries, J.: Theoretische studien über die umstimmung des sehorgans. Festschrift der Albrecht-Ludwigs-Universität pp. 145–158 (1902)
14. Lai, W.S., Huang, J.B., Wang, O., Shechtman, E., Yumer, E., Yang, M.H.: Learning blind video temporal consistency. In: Proceedings of the ECCV. pp. 170–185 (2018)
15. Qian, Y., Chen, K., Nikkanen, J., Kamarainen, J.K., Matas, J.: Recurrent color constancy. In: Proceedings of the IEEE ICCV. pp. 5458–5466 (2017)
16. Qian, Y., Kamarainen, J.K., Nikkanen, J., Matas, J.: On finding gray pixels. In: Proceedings of the IEEE/CVF Conference on CVPR. pp. 8062–8070 (2019)
17. Qian, Y., Käpylä, J., Kämäräinen, J.K., Koskinen, S., Matas, J.: A benchmark for burst color constancy. In: Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16. pp. 359–375. Springer (2020)
18. Van De Weijer, J., Gevers, T., Gijsenij, A.: Edge-based color constancy. IEEE Trans. Image Process. **16**(9), 2207–2214 (2007)
19. Wang, N., Funt, B., Lang, C., Xu, D.: Video-based illumination estimation. In: Computational Color Imaging: Third International Workshop, CCIW 2011, Milan, Italy, April 20-21, 2011. Proceedings 3. pp. 188–198. Springer (2011)
20. Zhou, Y., Xu, X., Shen, F., Gao, L., Lu, H., Shen, H.T.: Temporal denoising mask synthesis network for learning blind video temporal consistency. In: Proceedings of the 28th ACM international conference on multimedia. pp. 475–483 (2020)
21. Zini, S., Buzzelli, M., Bianco, S., Schettini, R.: Cocoa: combining color constancy algorithms for images and videos. IEEE Trans. Computat. Imaging **8**, 795–807 (2022)