https://doi.org/10.2352/CIC.2025.33.1.20

An Adaptive Per-Patch Weighted Regression for Target-Based Color Correction of sRGB Images under Uncontrolled Outdoor Illumination

Asel Esra Özyılmaz[1], Damien Muselet[2], Alain Trémeau[2], Mathis Cordier[3], Cindy Torres[3], Olivier Robert[3], Cyril Dambrine[3], Ali Boudjedra[3]
[1] Norwegian University of Science and Technology, Norway
[2] Laboratoire Hubert Curien, Jean Monnet University, France
[3] Vilmorin – Mikado / Limagrain Vegetable Seeds, France
Corresponding author: asel.esra.123@gmail.com

Abstract

Color-accurate digital imaging is critical for agricultural phenotyping, but the scientific literature predominantly assumes the availability of linear RAW sensor data. In some commercial workflows, only standard 8-bit sRGB JPEG images are available, which poses a significant challenge due to their non-linear encoding and information loss from in-camera processing. This paper presents a robust color correction pipeline designed specifically for such non-linear images captured in uncontrolled outdoor environments.

Our core contribution is a novel, per-patch adaptive weighting scheme for least-squares color correction. Instead of deriving a single global correction, our method generates a unique transformation for each patch on an in-frame ColorChecker. This is achieved through a leave-one-out approach where, for each target patch, a model is trained on the remaining 23 patches. Crucially, this training is guided by a weighting matrix customized for the target patch. This adaptive process allows a simple linear model to outperform more complex polynomials. Through systematic evaluation on an unseen test set, we demonstrate this method reduces the mean color error ΔE_{00} *from 11.23 to 3.79, providing a practical and effective solution for real-world agricultural imaging.*

Introduction

In precision agriculture and plant phenotyping, color is a primary indicator for assessing plant health, diagnosing nutrient stress, and estimating canopy cover. The reliability of these automated analyses depends on accurate and consistent color data that reflect biological reality, not artifacts of shifting illumination. As lighting conditions change, perceived colors can be altered to a degree that critically impacts the analysis. Therefore, color correction is an unavoidable step to ensure that color variations are due to plant characteristics and not external factors such as light conditions [1, 2].

Most commonly used target-based color correction methods are designed for linear RAW image data, where a direct mapping between sensor response and scene colorimetry can be established. However, in many large-scale commercial operations, such as those of our industrial partner Vilmorin–Mikado, imaging workflows rely exclusively on the 8-bit sRGB JPEG format for operational efficiency. These images have undergone irreversible in-camera processing, including gamma correction and compression, making recovering of true scene colors a theoretically ill-posed problem. This creates three core constraints:

- **Data-format constraint:** The non-linear nature of JPEG data violates the assumptions of most standard color correction models. Due to irreversible in-camera processing, any post-capture correction can only approximate true scene colors, as essential image information is permanently lost [3].
- **Ground-truth constraint:** Without RAW data, the only reliable color reference is to use a ColorChecker placed within the scene.
- **Illumination constraint:** Outdoor agricultural images often feature extreme local variations in lighting, such as direct sunlight and deep shadows within the same frame. Thus, applying a single global correction across the entire image becomes insufficient for addressing these mixed-light scenarios [4].

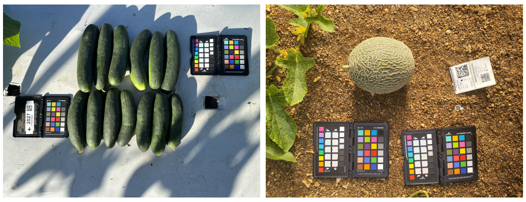
This paper addresses these challenges by introducing a color correction method centered on a per-patch adaptive weighting strategy. Our primary contribution is a novel approach where, instead of calculating a single global transform, we generate a unique correction for each color patch on an in-frame target. This is achieved via a Leave-One-Out Cross-Validation (LOOCV) framework. For each patch held out for testing, a weighted least-squares model is trained on the remaining patches. The weighting scheme is customized for the test patch, dynamically reducing the impact of training samples based on their colorimetric distance to the target. This targeted approach is designed to be robust against the challenging, mixed-lighting conditions common in agricultural settings while operating directly on non-linear sRGB JPEG data. We validate the method's effectiveness and probe its spatial limitations using a novel two-checker transfer test.

The following section details the dataset, experimental setup, and baseline comparisons that establish the foundation for our weighted regression framework.

Methodology

Dataset Description and Split Rationale

Our methodology was developed using a specially commissioned dataset designed to overcome the challenge of having no whole-image ground truth. While the standard workflow of the industrial partner, Vilmorin, leading this study produced thousands of images with a single ColorChecker, we requested a dataset of approximately 300 images that each contain two Passport Mini ColorCheckers. This two-checker configuration was intentional: it enabled us to validate whether corrections derived from one checker could transfer to the other under different local illumination conditions.



(a) Non-homogeneous light (b) Almost homogeneous light

(c) Homogeneous light

Figure 1: Example images from the two-checker dataset, showcasing a variety of subjects and challenging, mixed-lighting conditions encountered in the field.

The dataset is **confidential** because it was captured exclusively for this project. While the full images cannot be released, representative examples are provided in Fig. 1 to illustrate the experimental setup, checker placement, and typical outdoor illumination conditions. To support reproducibility, all model parameters and weighting functions are reported in this paper so that the method can be re-implemented on other datasets.

The images consist of fruit and vegetable subjects (e.g., cucumbers, peppers, melons) photographed under uncontrolled outdoor conditions. They were acquired using a range of modern consumer devices, including an Apple iPhone 13, Samsung Galaxy S22 (SM-S901U1), Samsung Galaxy Tab S9 FE 5G, and Samsung Galaxy Tab S6 Lite (SM-P613), with all files stored in JPEG format only.

For computational feasibility, we selected a **38-image development set** for model selection and tuning. This subset was chosen to capture diverse scenarios such as direct sunlight, deep shadow, and mixed illumination. The remaining **13 images** were held back as an independent test set for final evaluation. Many unused images are near-duplicates captured under similar conditions, so the 38/13 split was sufficient to represent the dataset while remaining tractable. We ensured that both homogeneous and non-homogeneous illumination cases were represented (see Fig. 1).

Patch coordinates are obtained with an automatic ColorChecker detector (existing in company pipeline before); Fig. 2 illustrates the input image, detected checker mask, and the zoomed target used with extracted patches.

All processing was performed using sRGB values extracted from the images and normalized to the [0, 1] range.

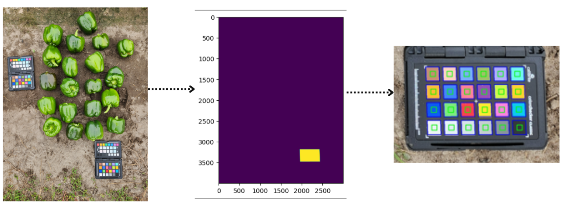


Figure 2: Automatic ColorChecker detection and patch extraction. **(a)** Original field image. **(b)** Binary mask of a detected checker region. **(c)** Zoomed target with the extracted patches

Establishing the Optimal Baseline Model

Our first objective was to identify the most effective correction architecture from a set of common target-based methods. We performed a comparative study on our 38-image development set, evaluating the following approaches:

- **Homography-based methods:** Alternating Least Squares (ALS) and RANSAC.
- **Regression-based methods:** Simple Ordinary Least-Squares (LS) and Root-Polynomial (RP) regression.

Following established practice in the literature [5], this initial evaluation was conducted in a linear color space (CIE XYZ). From this analysis, a 2nd-degree root-polynomial model emerged as the top-performing architecture.

Having selected the optimal model architecture, we next investigated the choice of working color space. The standard practice assumes linear sensor data, whereas our workflow is constrained to non-linear sRGB JPEGs. This discrepancy prompted a second, more focused experiment to determine if a linear space was truly optimal for our data. We took the winning 2nd-degree RP model and benchmarked its performance in two additional domains: a linearized RGB space (*via* an inverse-gamma transformation) and the camera's native sRGB space.

This second evaluation revealed that operating directly in the sRGB space yielded a further reduction in the mean ΔE_{00} error. Consequently, we established our definitive baseline as a **2nd-degree root-polynomial model applied directly in sRGB space**. While this represented the best result achievable with conventional methods, its residual errors were still too high for reliable field use, which directly motivated the development of the more robust, weighted regression framework detailed below.

Adaptive Weighted Least-Squares Formulation

Building on this baseline, our adaptive framework is a regularized weighted least-squares regression. It is built on the hypothesis that by strategically weighting the contribution of each color patch, we can improve the model's fit even with noisy, real-world data. This choice is motivated by an ongoing debate on patch selection: some studies report that small subsets (e.g., primaries or even mostly achromatics) can approach the performance of using all 24 patches, while others emphasize the need for chromatic diversity under varying illuminants [1, 7]. Rather than committing to a fixed subset, we use *all* patches and let the data modulate their influence via adaptive weights—practical for mixed-illuminant, JPEG-only settings where patch reliability can change with scene content [9, 8]. This approach allows us to systematically optimize model degree, regularization, and patch-weighting simultaneously. The general form of this model is:

$$\hat{\mathbf{M}} = \left(\mathbf{X}^T\mathbf{W}\mathbf{X} + \lambda\mathbf{I}\right)^{-1}\mathbf{X}^T\mathbf{W}\mathbf{Y}, \tag{1}$$

where the terms are defined as follows:

- $\hat{\mathbf{M}}$ is the estimated $k \times 3$ transformation matrix, where k is the number of polynomial terms.
- $\mathbf{X}$ is the $N \times k$ design matrix containing the input sRGB values from N color patches, expanded into a polynomial feature space.
- $\mathbf{Y}$ is the $N \times 3$ matrix of corresponding reference sRGB values for the N patches.
- $\mathbf{W}$ is the $N \times N$ diagonal matrix containing the weights assigned to each of the N patches.

- $\lambda\mathbf{I}$ is the Tikhonov regularization term, where $\mathbf{I}$ is the identity matrix. This term improves the numerical stability of the solution by adding a positive value to the diagonal of the $\mathbf{X}^T\mathbf{W}\mathbf{X}$ matrix.

The model's final configuration is determined by a set of hyperparameters—including the polynomial degree of $\mathbf{X}$, the regularization parameter λ, and the parameters defining the patch-weighting scheme—which were optimized through grid search as detailed in the following section.

Patch Weighting and Optimization Strategy

The patch weights in $\mathbf{W}$ are crucial for moderating the influence of unreliable patches. We developed a comprehensive grid search strategy to find the optimal weighting scheme based on the features and functions that best minimized the mean ΔE_{00} error. This involved two stages:

1. **Individual Feature Tuning:** We first tested four distinct weighting features in isolation:
   - ΔE in L*a*b* space (perceptual distance)
   - RGB Euclidean distance
   - Exposure (relative luminance)
   - Clipping severity (to detect channel saturation)

   For each feature, we evaluated multiple weighting functions. The two functions that formed our final model were an **inverse model** and an **exponential decay model**, defined as:

   $$w_{inverse} = \frac{1}{1+d^p} \tag{2}$$

   $$w_{exp\_decay} = \exp(-p \cdot d) \tag{3}$$

   where w is the calculated weight, d is the value of the input feature (e.g., color distance), and p is a tunable hyperparameter controlling the scale or rate of the function. A grid search was performed across a range of values for p for each function.
2. **Combination Stage:** We then tested combinations of the best-performing individual features, using methods such as weighted sum and geometric blending to find the final patch weight.

The goal of the grid search was to find the configuration of the model degree, regularization, and weighting scheme that minimized the average ΔE_{00} error under the Leave-One-Out Cross-Validation (LOOCV) protocol.

Results and Discussion

Optimal Model Configuration

The extensive grid search yielded a surprising result: the best-performing model was not a higher-order RPCC, but a **degree-1 model**. This simplifies the approach to a weighted linear least-squares regression. The final optimized configuration, detailed in Table 1, shows that the weighting was entirely driven by colorimetric distance (90% RGB distance, 10% LAB distance), while the weights for exposure and clipping were optimized to zero. This indicates that for this sRGB-to-sRGB correction task, identifying and down-weighting patches that are already colorimetrically distant is more effective than attempting to model physical phenomena like exposure.

Table 1: Final Optimized Model Configuration

Parameter	**Value**
Polynomial Degree	1 (Linear)
Regularization (λ)	0.0001
Weighting Blend	90% RGB distance, 10% LAB distance
RGB Weight Function	Exp. Decay (rate $p = 10.0$)
LAB Weight Function	Inverse (scale $p = 9.0$)

Quantitative Performance and Generalization

The weighted linear method demonstrated strong and, crucially, generalizable performance across both the training (38 images) and test (13 unseen images) sets. The absolute error reductions are detailed in Table 2, showing that on the unseen test set, the mean ΔE_{00} fell from 11.23 to 3.79, while the mean RGB error dropped from 0.24 to 0.07.

Translating these absolute values into relative improvements reveals the model's consistency. On the training set, the final weighted model improved the mean ΔE_{00} by 71.1% and the mean RGB error by 72.3%. This strong performance was mirrored on the unseen test set, which saw comparable improvements of 66.2% and 69.7% for the same metrics, respectively. Furthermore, the contribution of the weighting scheme itself was consistent across both datasets, providing an additional improvement of approximately 32% over the unweighted baseline. This remarkable consistency across different datasets and metrics confirms that the model generalizes effectively and is not overfit, providing a reliable solution for practical field use.

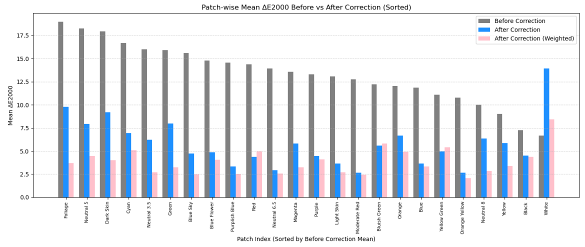
Table 2: Mean ΔE_{00} and mean RGB error on train and test sets

Scenario	**Train Set (n=38)**	**Test Set (n=13)**
Mean ΔE Before	13.36	11.23
After (Unweighted)	5.79	5.60
After (Weighted)	**3.87**	**3.79**
Mean RGB Error Before	0.27	0.24
After (Unweighted)	0.11	0.10
After (Weighted)	**0.07**	**0.07**

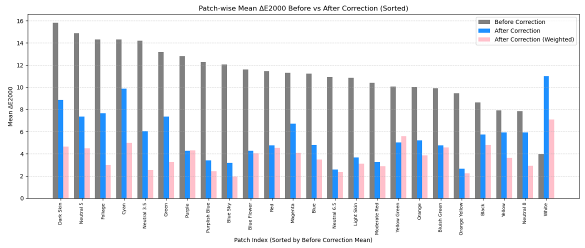
Analysis of Per-Patch Performance

Fig. 3 reports patch-wise mean ΔE_{00} for the development and test sets, comparing the uncorrected images, the unweighted baseline, and the proposed weighted model. Improvements concentrate on chromatic patches (e.g., foliage, skin, blues), while a few hues (yellow-green / bluish-green / reddish) show smaller gains or slight regressions.

While the overall performance improved significantly, analysis of patch-wise errors reveals that the effect was not uniform across all colors. Correction was most effective for chromatic patches such as foliage, skin tones, and blues, which often had the highest initial errors. However, patch-specific behavior revealed some exceptions. Some patches—notably in the bluish-green, yellow-green, and reddish-tone regions—showed a slight increase in error after weighting. We hypothesize that this may be due to the complex interaction between these specific colors and the non-linearities of the sRGB color space, or an artifact of having limited training data for those specific hues. This behavior highlights that even with optimized weighting, certain colors remain challenging to correct and warrant further investigation.



(a) Development set (n=38)



(b) Test set (n=13)

Figure 3: Patch-wise mean ΔE_{00} before correction, after the unweighted baseline, and after the proposed weighted regression. Bars are sorted by pre-correction error; tick labels follow ColorChecker names.

Two-Checker Transfer Test: Probing the Limits of Global Correction

As introduced in methodology, each image contains two ColorCheckers captured under potentially different local illumination.

To probe whether corrections generalize across different local illumination, we exploited the two-checker configuration. For each image, one checker (Checker A) was designated for training and the other (Checker B) for testing. Specifically, for each held-out patch on Checker A, we trained a weighted least-squares model on its remaining 23 patches using LOOCV, then applied the resulting transformation to the corresponding patch on Checker B. Errors were averaged across all patches to produce the transfer performance.

This experiment revealed a clear performance gap. Training and testing on the same checker reduced the mean ΔE_{00} error from 14.41 to 3.99. When the correction matrix was transferred to the second checker, the error only fell from 15.29 to 6.38. In relative terms, the correction was 72.3% effective on the source checker but dropped to 58.3% on the second. Degradation was most severe when the two checkers were under strongly contrasting illumination (e.g., one in sunlight, the other in shadow).

Despite this drop, the weighted model still achieved a substantial improvement on the second checker. This result validates that the model generalizes beyond the patches it was trained on, but also highlights the fundamental limitation of global correction in scenes with mixed illumination. These findings directly motivate our ongoing work on spatially-aware shadow handling.

Qualitative Results on Full-Image Subjects

Since no pixel-wise ground truth exists for the full images, we complement the quantitative checker-based evaluation with qualitative results on the actual subjects. For this, the main fruit or vegetable was segmented from the image (see Fig. 4), and the corresponding subject-specific color correction matrix was applied. This demonstrates how the proposed per-patch weighting scheme, tuned according to the subject's dominant color, improves the overall visual consistency.

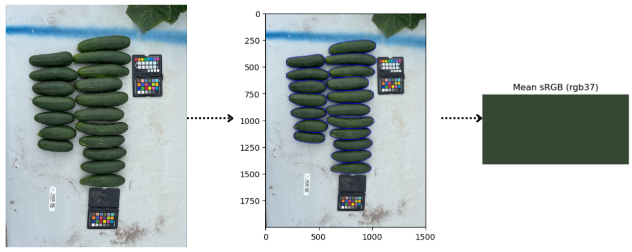


Figure 4: Example of subject segmentation used for qualitative evaluation. The fruit mask (center) is extracted from the original image (left) and the mean sRGB color swatch from the fruit mask is shown (right).

Representative before/after examples are shown in Fig. 5. On these unseen test images, the correction reduces strong illumination-induced shifts and brings fruit colors closer to their expected appearance. For instance, the cucumbers in (a)–(b) become less desaturated after correction, while the melon in (c)–(d) better matches the neutral background. These qualitative results confirm that the weighting strategy effectively adapts corrections toward the subject of interest, even when global pixel-wise ground truth is unavailable.

(a) Before (Image 1)

(b) After correction (Image 1)

(c) Before (Image 2)

(d) After correction (Image 2)

Figure 5: Qualitative evaluation on full-image subjects. Left: original uncorrected images. Right: images corrected using the proposed weighted regression with subject-adaptive weighting.

Discussion

The superior performance of the weighted linear model over more complex polynomials is a key finding of our study. We hypothesize that for non-linear JPEG data with only 24 reference patches, higher-order models are prone to overfitting to noise and in-camera artifacts [5, 6]. The adaptive weighting scheme, by down-weighting unreliable or distant patches, allows a simpler linear model to achieve a better global fit.

Patch-level analysis (Fig. 3) confirmed that the method systematically improves chromatic patches such as foliage, skin tones, and blues, which are often the most error-prone. At the same time, certain hues—notably yellow-green, bluish-green, and reddish tones—showed smaller or even negative gains. This highlights a limitation of relying solely on patch-based weighting: while the method is effective overall, specific spectral regions remain more difficult to correct.

The two-checker transfer test provided further insight into these limitations. When both checkers were under similar illumination, the correction transferred well, but when one checker was in direct light and the other in shadow (as in Fig. 1 (a)), the accuracy degraded substantially. This indicates that a purely global model cannot fully account for strong local illumination differences. These findings directly motivate ongoing work on spatially-aware shadow handling and non-homogeneity modeling.

Finally, although deep learning approaches have been shown to be powerful for color calibration [2], we deliberately did not adopt them in this work. The lack of pixel-wise ground truth, combined with the JPEG-only constraint, makes supervised training impractical. Instead, we prioritized interpretable regression-based methods that can be reliably evaluated with available reference patches. Once a dedicated dataset with RAW images or shadow masks is constructed, benchmarking against modern learning-based methods will become a natural next step.

Conclusion & Future Directions

This paper presented a color correction pipeline tailored to the practical constraints of agricultural field imaging, where only non-linear sRGB JPEG files are available. By combining an in-frame ColorChecker with an adaptive per-patch weighting scheme, a weighted linear least-squares model reduced the mean perceptual error (ΔE_{00}) from 11.23 to 3.79 on unseen images. This demonstrates that even with limited and noisy patch data, a simple linear model, guided by adaptive weights, can outperform more complex polynomial regressions.

Patch-level analysis revealed that the largest improvements occurred in chromatic patches such as foliage, skin tones, and blues, while some hues (yellow-green, bluish-green, reddish) remained challenging. These uneven results highlight both the robustness and the limits of the weighting scheme. The two-checker transfer test further demonstrated that global correction degrades in strongly non-homogeneous lighting, underscoring the need for spatially-aware methods. Such cases, particularly those involving shadowed versus sunlit checkers, directly motivate our future exploration of shadow modeling and illumination-adaptive correction.

Deep learning methods were not applied here due to the absence of suitable ground truth and the JPEG-only constraint. However, this work establishes a strong regression-based baseline. Future directions will include (i) developing shadow-aware, spatially adaptive corrections, (ii) a more detailed investigation of patch-specific anomalies, and (iii) benchmarking against learning-based models once supervised datasets become available.

References

[1] S. Sunoj, C. Igathinathane, N. Saliendra, J. Hendrickson, and D. Archer, “Color calibration of digital images for agriculture and other applications,” *ISPRS J. Photogram. Remote Sens.*, 146, 221–234 (2018).

[2] A. Abdalla, H. Cen, E. Abdel-Rahman, L. Wan, and Y. He, “Color calibration of proximal sensing RGB images of oilseed rape canopy via deep learning combined with K-means algorithm,” *Remote Sens.*, 11(24), 3001 (2019).

[3] G. Simone, M. Gaiani, A. Bellabeni, and A. Rizzi, “Problems in image target-based color correction,” *Electron. Imaging*, 34(15), 142-1–142-47 (2022).

[4] I. Boyadzhiev, K. Bala, S. Paris, and F. Durand, “User-guided white balance for mixed lighting conditions,” *ACM Trans. Graph.*, 31(6), 200:1–200:10 (2012).

[5] G. D. Finlayson, M. Mackiewicz, and A. Hurlbert, “Color correction using root-polynomial regression,” *IEEE Trans. Image Process.*, 24(5), 1460–1470 (2015).

[6] A. Kucuk, G. D. Finlayson, R. Mantiuk, and M. Ashraf, “Comparison of regression methods and neural networks for colour correction,” *Imaging Conference*, Vol. 2022, 117-1–117-5 (2022).

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

[8] M. Buzzelli, S. Zini, S. Bianco, G. Ciocca, R. Schettini, and M. K. Tchobanou, “Analysis of biases in automatic white balance datasets and methods,” *Color Res. Appl.*, 48(1), 40–62 (2023).

[9] M. N. M. Afifi, “Image color correction, enhancement, and editing,” PhD Dissertation, York University (2021).