

# Efficient illuminant correction in the Local, Linear, Learned ( $L^3$ ) method

Francois G. Germain<sup>a</sup>, Ireteiyao A. Akinola<sup>b</sup>, Qiyuan Tian<sup>b</sup>, Steven Linsel<sup>c</sup>, and Brian A. Wandell<sup>b,d</sup>

<sup>a</sup>Center for Computer Research in Music and Acoustics,  
Stanford University, Stanford, CA-94305, USA;

<sup>b</sup>Electrical Engineering Department, Stanford University, Stanford, CA-94305, USA;

<sup>c</sup>Olympus Corporation of the Americas, Sunnyvale, CA-94085, USA;

<sup>d</sup>Psychology Department, Stanford University, Stanford, CA-94305, USA.

## ABSTRACT

To speed the development of novel camera architectures we proposed a method,  $L^3$  (Local, Linear and Learned), that uses camera simulation to create an optimized image processing pipeline. The  $L^3$  method assigns each pixel into one of about 400 classes and uses camera simulation and machine learning to derive a linear transform for the data in each class. Each linear transform maps the sensor data from a pixel and its neighbors into the target output (e.g., image XYZ rendered under a D65 illuminant). The 400 transforms are stored in a table and used for image rendering. The sensor and rendering illuminant can be equal (same-illuminant table) or different (cross-illuminant table). In the original implementation, we derived many cross-illuminant tables, and the collection of tables can be relatively large. We find, however, that a single same-illuminant table (D65) effectively converts sensor data for many different same-illuminant conditions. Hence, we propose to render the data by applying the same-illuminant D65 table to the sensor data, followed by a color correction transform. The mean color reproduction error using the same-illuminant table is on the order of  $4\Delta E$  units, which is only slightly larger than the cross-illuminant table errors. This approach reduces table storage requirements significantly without substantially degrading color reproduction accuracy.

**Keywords:** Local Linear Learned ( $L^3$ ), image processing pipeline, illuminant correction

## 1. INTRODUCTION

The  $L^3$  algorithm automatically generates a standard image processing pipeline (sensor correction, demosaicking, illuminant correction, noise reduction) for novel camera architectures.<sup>1-4</sup> The  $L^3$  image processing pipeline comprises two parts.<sup>1</sup> First, the pipeline classifies each pixel (based on a statistical analysis of the responses at the pixel and its neighborhood) into one of many possible classes. Second, the processing pipeline linearly transforms the data of the pixel and its neighbors to the target output space.

The  $L^3$  training builds a table of linear transforms from a camera model that is implemented using the Image Systems Engineering Toolbox (ISET).<sup>5,6</sup> The simulation converts many samples of natural scenes to sensor data. The desired rendering, i.e. CIE XYZ values for consumer photography, are computed from the known natural scene radiance. Through multiple simulations, the  $L^3$  algorithm learns a table of linear transforms that convert sensor data to the target CIE XYZ values.

## 2. MOTIVATION: ILLUMINANT CORRECTION

Illuminant correction is an essential part of the image processing pipeline that is necessary to account for the perceptual phenomenon of *color constancy*.<sup>7</sup> The visual system adapts to changes in the ambient light level and color, and this adaptation has the general effect of preserving the color appearance of an object (e.g. a white

---

Send correspondence to Francois G. Germain. E-mail: FG (fgermain@stanford.edu), IA (iakinola@stanford.edu), QT (qytian@stanford.edu), SL (steven.linsel@olympus.com), BW (wandell@stanford.edu)

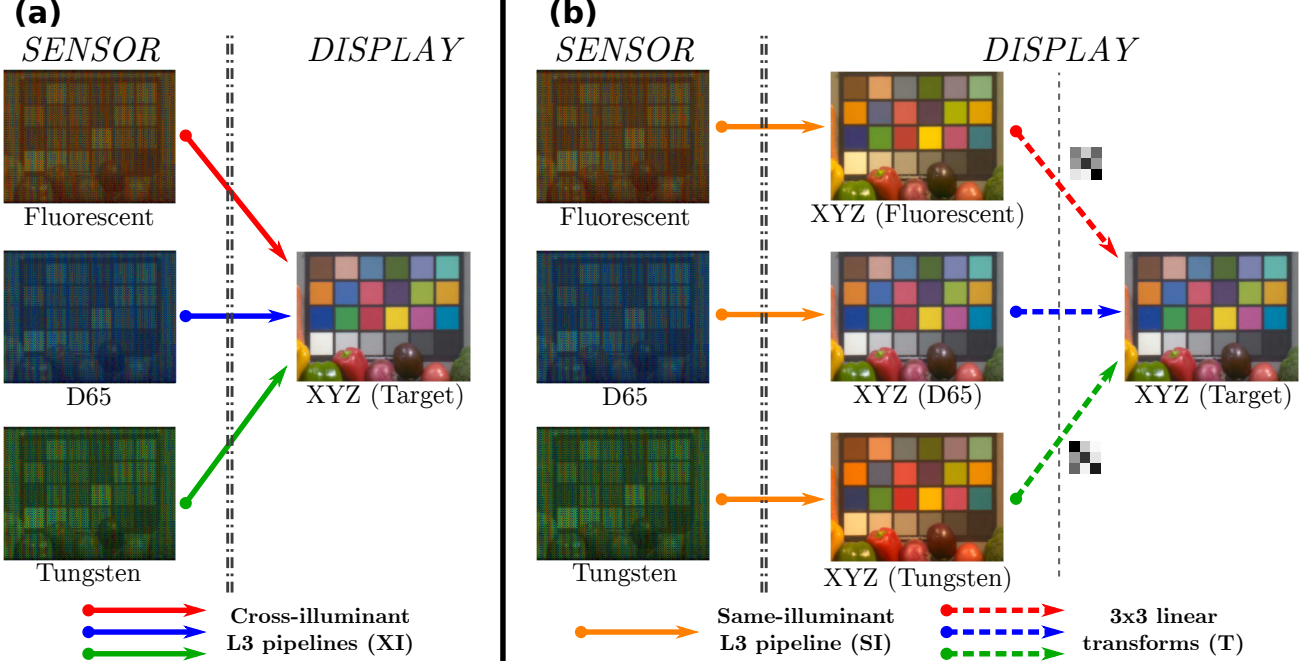


Figure 1. **Illuminant correction with  $L^3$  image processing pipelines.** (a) Cross-illuminant (XI) illuminant-dependent  $L^3$  processing. (b) Same-illuminant (SI) illuminant-independent  $L^3$  processing followed by a linear color correction transform (T).

shirt) across conditions (daylight to tungsten light). Because the human visual system adapts to the illuminant, to preserve color appearance the linear transforms in the  $L^3$  method used for rendering must adapt, too.

In the original algorithm, we anticipated learning a table of linear transforms for each illumination condition (cross-illuminant, Figure 1(a)). These cross-illuminant tables would be stored and applied for rendering. This approach pays a significant cost in computation and storage. Here, we explore an approach that separates illuminant correction from the rest of the pipeline. We learn a same-illuminant table, that renders data acquired under one illuminant into XYZ values under the same illuminant. This accomplishes demosaicking, denoising and sensor correction, but it does not perform illuminant correction (Figure 1(b)). We then apply an illuminant correction transform to obtain the final rendering. This architecture requires training and storing only one table of  $L^3$  transforms to be used for all illuminants and one illuminant correction matrix for each illuminant of interest.

### 3. METHOD

#### 3.1 General simulation conditions

We used ISET<sup>5,6</sup> to simulate digital camera processing. The ISET camera model includes the simulation of scene radiance, optics, sensor electronics and image processing pipeline (here the  $L^3$  pipeline). We model the camera as an f/4 lens with diffraction limited optics and focal length of 3mm. The sensor uses a RGB/W color filter array (CFA) layout described in,<sup>1</sup> with parameter values as in Figure 2.

The  $L^3$  algorithm uses simulated training data to learn a table of linear transforms.<sup>1</sup> Each pixel in the simulated sensor data is assigned to a distinct class based on (a) its color type (R, G, B, W), (b) the response voltage level (low, middle, high), and (c) the neighborhood spatial pattern (uniform or textured). The transform for a given class is derived by optimizing a Wiener filter between the data in the pixel neighborhood in the sensor, and the corresponding ideal value in the target color space (XYZ).

Optics(diffraction-limited)	
Focal length(mm)	3
F-number	4
Aperture diameter (mm)	0.75
Sensor	
Rows/Columns	204 × 240
Size (mm)	0.4488 × 0.528
Dark noise nonuniformity (mV)	1.41
Photoreceptor nonuniformity (%)	$2.218 \times 10^{-3}$
Analog gain	1
Analog offset (V)	0
Sensor pixel	
Width/height ( $\mu\text{m}$ )	$2.2 \times 2.2$
Fill factor	0.45
Dark voltage (V/sec)	$1 \times 10^{-5}$
Read noise (mV)	1.34
Conversion gain ( $\mu\text{V}/e$ )	200
Voltage swing (V)	1.8
Well capacity (electrons)	9000



Figure 2. Sensor parameter values (left) and Natural-100 color test chart (right)

The rendering process identifies the class membership for each pixel in the sensor data and then applies the appropriate stored linear transform. The image processing pipeline, which performs demosaicking, denoising, sensor conversion comprises the application of these precomputed local, linear transforms.

### 3.2 Learning the $L^3$ tables

**Learning the cross-illuminant tables** Cross-illuminant (XI) tables of  $L^3$  transforms are learned by simulating the sensor data under tungsten ( $XI_{\text{Tun}}$ ) and fluorescent ( $XI_{\text{Flu}}$ ) illumination and rendering to a target representation (XYZ) under a D65 illuminant.

**Learning the same-illuminant table** Same-illuminant (SI) tables of  $L^3$  transforms are learned by simulating sensor data and the target display XYZ values using the same illuminant. We learned same-illuminant tables for D65 ( $SI_{\text{D65}}$ ), tungsten ( $SI_{\text{Tun}}$ ) and fluorescent ( $SI_{\text{Flu}}$ ). To complete the rendering, we apply an illuminant correction transform from the acquisition illuminant to the final D65 rendering. The matrix T is derived by a linear regression between the XYZ values in the acquired and target (D65) representations. T is derived from first principles and is not dependent on the camera structure or  $L^3$ .

In principle, any of the SI tables can be used. In fact, it is possible to learn a same-illuminant table using sensor data from an ensemble of illuminants. (e.g., a mixture of D65 and tungsten scenes). The linear transforms reflect the properties of the multiple illuminants used in the ensemble, rather than the requirements of a specific illuminant. For example, the response level of a blue pixel under tungsten illuminant is typically quite low as reflected in the images obtained from tungsten-based table. The  $L^3$  compensates by using linear transforms with more blurring on those pixels compared to the D65 same-illuminant transforms. We learned tables for an ensemble comprising data from a mixture of D65 and Tungsten illuminants ( $SI_{\text{Ens}}$ ) which results in a more robust table for both illuminants.

### 3.3 Color evaluation

To compare the XI and SI methods, we used data from hyperspectral scenes and simulated test charts.

**Multispectral scenes** We used a small set of multispectral scenes including the Macbeth (Gretag) color chart, fruits, and people.<sup>8</sup> These images compare the algorithms visually.

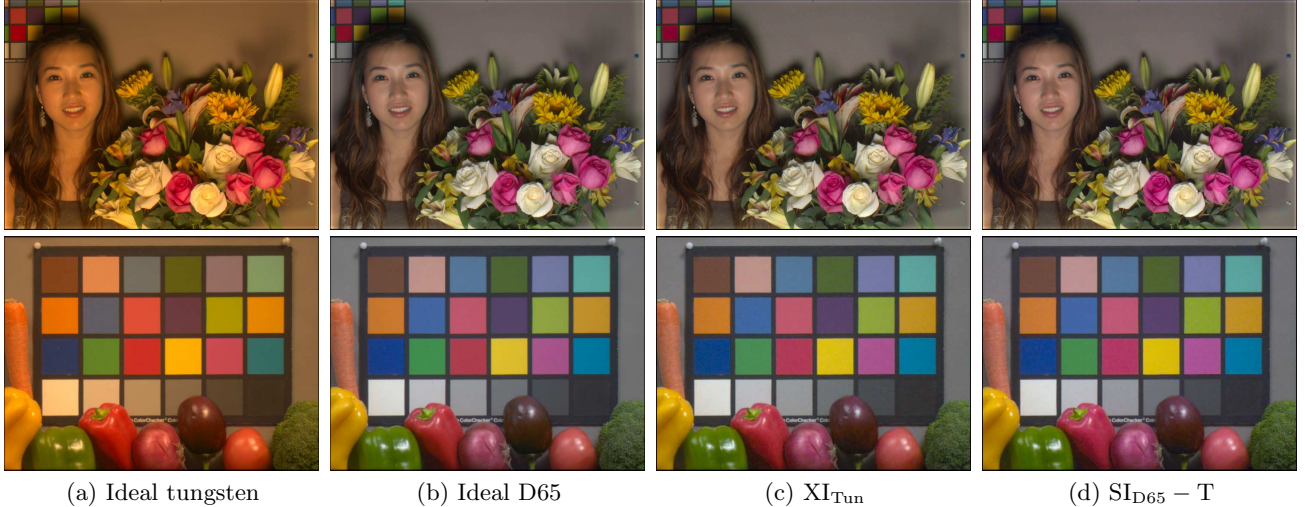


Figure 3. **Cross-illuminant and same-illuminant algorithms perform equally well in color rendering for a tungsten illuminant.** (a) The sensor image under tungsten, uncorrected. (b) The ideal image under a D65 illumination. (c) A cross-illuminant analysis for the same image but acquired under a tungsten light. The  $L^3$  table was built from the tungsten-acquisition data and XYZ target values for a D65-rendered image. (d) A same-illuminant rendering. In this case the L3 table was built using the same-illuminant D65 table, and the final rendering combined this table with an illuminant correction matrix T. The cross-illuminant and same-illuminant architectures produce similar images.

**Natural-100 reflectance chart** To generate a systematic and quantitative assessment of the color reproduction error, we created a Natural-100 test chart (Figure 2). The chart includes 100 samples with natural object reflectances from five categories,<sup>9</sup> and a 10-sample gray color strip. The reflectances are grouped by category (20 food, 20 clothes, 20 nature, 5 hair, 15 skin and 20 paint color samples). To select the category samples we calculated 7 principal components for the entire set of natural reflectances, represented each sample by its component weights, and then formed clusters of the samples. If N samples are needed from a category, we formed N clusters for that category and chose one sample from each cluster. By selecting one sample per cluster per category we ensure that the final chart has a variety of reflectances that span the sample pool.

We quantified color reproduction error using CIE2000  $\Delta E$  units.<sup>10</sup> The error was the difference between the chart XYZ under a D65 illuminant and the data rendered using either the SI or XI method.

#### 4. EXPERIMENTAL RESULTS

We compared the quality of the color rendering using two natural hyperspectral scenes (Figures 3 and 4). Each column shows a different representation of the scene. The leftmost columns show the scene as measured under a tungsten illuminant (Figure 3(a)) and a fluorescent illuminant (Figure 4(a)). The scene rendered under a D65 illuminant, calculated from the multispectral data, is shown in panel (b). The original data rendered using XI table is shown in panel (c), and rendered using the  $SI_{D65}$  table and linear transformation T ( $SI_{D65} - T$ ) is shown in (d). The XI and  $SI_{D65} - T$  renderings (panels c and d) are good approximations to the calculated D65 rendering (panel b).

We quantified the color rendering quality accuracy using the Natural-100 test chart (Table 1). The XI color reproduction errors are very similar to the SI - T color errors. Hence, the color accuracy of the efficient single-illuminant approach matches the cross-illuminant table for the tungsten illuminant.

Interestingly, we found that the  $SI_{D65}$  table trained with tungsten illumination can replace the  $SI_{Tun}$  table with only a small reduction in color reproduction accuracy ( $1.5\Delta E$ ). Using a SI table derived from training on an ensemble (tungsten and D65) of illuminants reduces the color accuracy by even less ( $0.4\Delta E$ ). As a result, we believe that training a SI table from data with an ensemble of illuminants may provide the best SI approach.

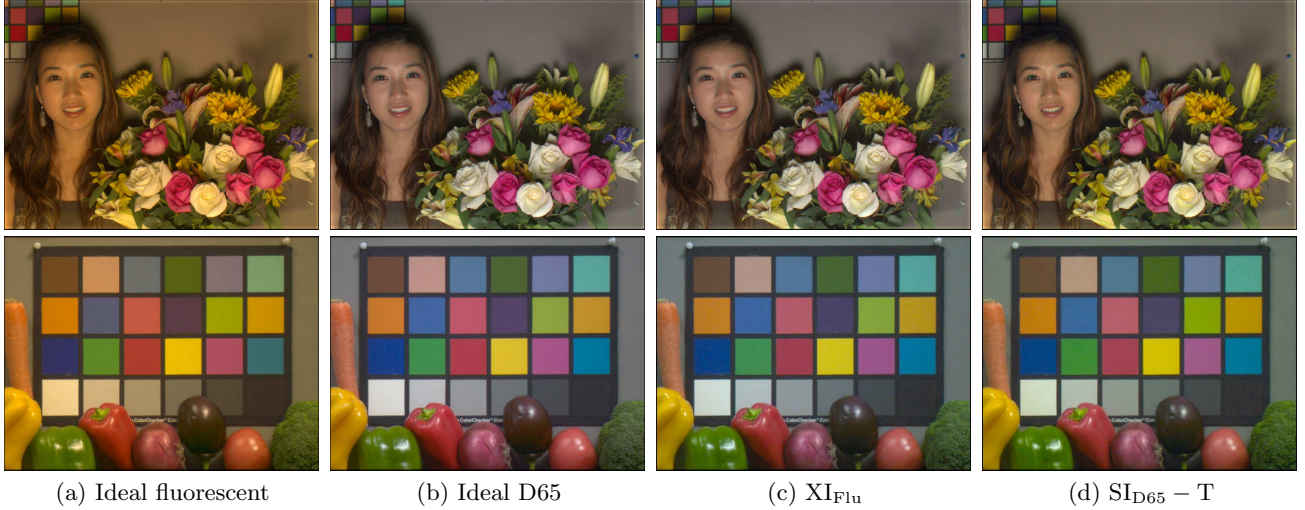


Figure 4. **Cross-illuminant and same-illuminant algorithms perform equally well in color rendering for a fluorescent illuminant.** (a) The sensor image under fluorescent, uncorrected. (b) The ideal image under a D65 illumination. (c) A cross-illuminant analysis for the same image but acquired under a fluorescent light. The  $\mathbf{L}^3$  table was built from the fluorescent-acquisition data and XYZ target values for a D65-rendered image. (d) A same-illuminant rendering. In this case the  $\mathbf{L}^3$  table was built using the same-illuminant D65 table, and the final rendering combined this table with an illuminant correction matrix  $\mathbf{T}$ . The cross-illuminant and same-illuminant architectures produce similar images.

A more detailed account of the color reproduction error for the cross-illuminant tables ( $\mathbf{XI}_{\text{Tun}}$ ) and two different same-illuminant tables followed by the illuminant transform ( $\mathbf{SI}_{\text{Tun}} - \mathbf{T}$  and  $\mathbf{SI}_{\text{D65}} - \mathbf{T}$ ) is presented in Figure 5 with the distribution of the  $\Delta E$  values recorded for each reflectance on the Natural-100 chart and the scatter plots comparing the CIE XYZ as well as the  $L^*a^*b^*$  color values of the true and reproduced colors.

## 5. CONCLUSIONS

We compared color reproduction accuracy following illuminant correction for two versions of the  $\mathbf{L}^3$  algorithm. In the SI version we store a single, same-illuminant table and a set of  $3 \times 3$  illuminant correction transformations. In the XI version, we store a set of tables, one for each illumination. The SI method requires far less storage than the XI method and costs very little in terms of color reproduction accuracy. We present different options regarding the choice of the same-illuminant table. We suggest that the best option for a same-illuminant table might be derived by training the  $\mathbf{L}^3$  algorithm with data simulated using an illuminant ensemble.

Color reproduction error ( $\Delta E$ )	Average (Std. dev.)	90th percentile
$\mathbf{XI}_{\text{Tun}} \mathbf{L}^3$	2.5 (1.8)	4.9
$\mathbf{SI}_{\text{Tun}} \mathbf{L}^3$ with linear color transform $\mathbf{T}$ ( $\mathbf{SI}_{\text{Tun}} - \mathbf{T}$ )	2.5 (2.0)	5.4
$\mathbf{SI}_{\text{D65}} \mathbf{L}^3$ with linear color transform $\mathbf{T}$ ( $\mathbf{SI}_{\text{D65}} - \mathbf{T}$ )	4.0 (3.0)	8.3
$\mathbf{SI}_{\text{Ens}} \mathbf{L}^3$ with linear color transform $\mathbf{T}$ ( $\mathbf{SI}_{\text{Ens}} - \mathbf{T}$ )	2.9 (2.2)	6.0

Table 1. **Color reproduction errors ( $\Delta E$ ) for the Natural-100 color chart.** The original images were simulated using a tungsten illuminant. The left column shows the training method used to derive the  $\mathbf{L}^3$  table. The mean reproduction error (middle) and the relatively large reproduction errors (right) are shown. The smallest reproduction error is from the cross-illuminant table. But the errors from several same-illuminant tables followed by a linear transform are only slightly larger.



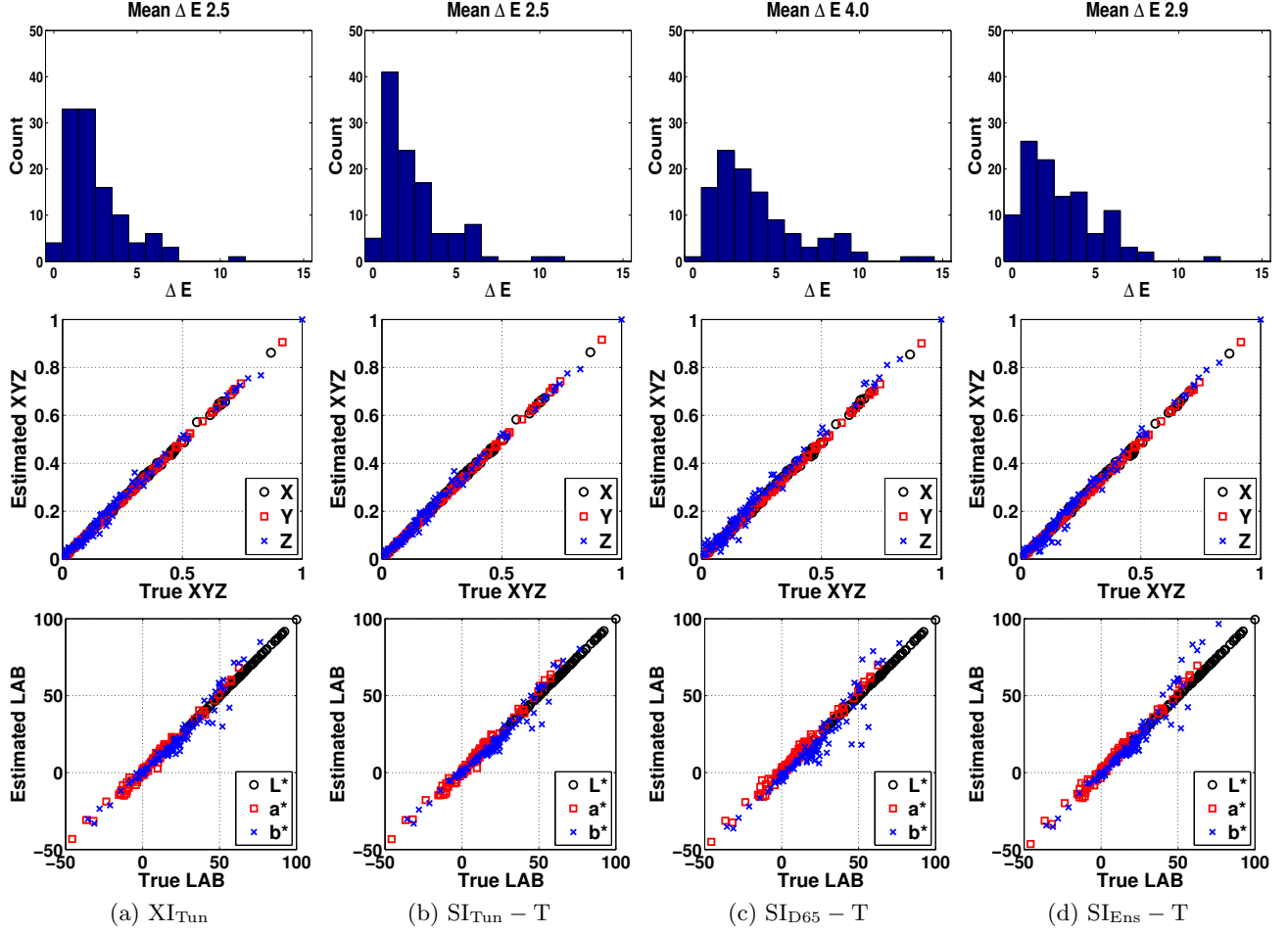


Figure 5. Color error distributions for a Natural-100 chart acquired under a tungsten illuminant and rendered using different  $L^3$  tables. We compare the distribution of color reproduction error in  $\Delta E$  units (top) and the comparison between true and estimated XYZ (middle) and  $L^*a^*b^*$  coordinates (bottom) for the 110 chart reflectances for (a) the cross-illuminant table ( $XI_{Tun}$ ), (b) the same-illuminant table trained under tungsten followed by the illuminant transform T ( $SI_{Tun} - T$ ), (c) the same-illuminant table trained under D65 followed by the illuminant transform T ( $SI_{D65} - T$ ) and (d) the same-illuminant table trained under an ensemble of illuminants followed by the illuminant transform T ( $SI_{Ens} - T$ ).

## REFERENCES

- [1] Tian, Q., Lansel, S. P., Farrell, J. E., and Wandell, B. A., “Automating the design of image processing pipelines for novel color filter arrays: local, linear, learned (l3) method,” *Proceedings SPIE* **9023**, 90230K–90230K–8 (2014).
- [2] Lansel, S. P. and Wandell, B. A., “Local linear learned image processing pipeline,” in *[Imaging Systems and Applications]*, Optical Society of America (2011).
- [3] Lansel, S. P., *Local Linear Learned Method for Image and Reflectance Estimation*, PhD thesis, Stanford University (2011).
- [4] Lansel, S. P. and Wandell, B. A., “Learning of image processing pipeline for digital imaging devices,” (Dec. 7 2012).
- [5] Farrell, J. E., Xiao, F., Catrysse, P. B., and Wandell, B. A., “A simulation tool for evaluating digital camera image quality,” in *[Electronic Imaging 2004]*, 124–131, International Society for Optics and Photonics (2003).

- [6] Farrell, J. E., Catrysse, P. B., and Wandell, B. A., “Digital camera simulation,” *Applied Optics* **51**(4), A80–A90 (2012).
- [7] Wandell, B. A., [*Foundations of Vision*], Sinauer Associates (1995).
- [8] <http://imageval.com/scene-database/>.
- [9] Vrhel, M. J., Gershon, R., and Iwan, L. S., “Measurement and analysis of object reflectance spectra,” *Color Research & Application* **19**(1), 4–9 (1994).
- [10] Sharma, G., Wu, W., and Dalal, E. N., “The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations,” *Color Research & Application* **30**(1), 21–30 (2005).