An Empirical Camera Model for Internet Color Vision

Ayan Chakrabarti[1]
http://www.eecs.harvard.edu/~ayanc/

Daniel Scharstein[2]
http://www.cs.middlebury.edu/~schar/

Todd Zickler[1]
http://www.eecs.harvard.edu/~zickler/

[1] Harvard School of Engineering and Applied Sciences
Cambridge, MA, USA 02139

[2] Department of Computer Science
Middlebury College,
Middlebury, VT, USA 05753

Images harvested from the Web are proving to be useful for many visual tasks, including recognition, geo-location, and three-dimensional reconstruction. These images are captured under a variety of lighting conditions by consumer-level digital cameras, and these cameras have color processing pipelines that are diverse, complex, and scene-dependent. As a result, the color information contained in Internet imagery has often been ignored.

The goal of this paper is to determine an efficient representation for the color processing pipelines of consumer cameras. We seek a parameterized function that maps input spectral radiance distributions (SPDs) to output sRGB color vectors that is "efficient" in the sense of being complex enough to accurately model real consumer cameras but simple enough to be used for vision-related inverse problems (stereo reconstruction, object recognition, shape from shading, color constancy, and so on).

Since camera processing pipelines are trade secrets, discovering an efficient parametrized map requires a phenomenological approach. Accordingly, we have acquired a new database of registered images from varying camera models, varying lighting conditions, and varying camera settings. Our database exploits the increasing availability of consumer-level cameras that output both RAW data and JPEG-encoded data, and it currently includes over 1000 images taken with 35 different camera models, ranging from simple point-and-shoot cameras to professional DSLRs. The database is available at

http://vision.middlebury.edu/color/

We use our database to explore models for the mapping from a camera's RAW sensor measurement κ to its JPEG output y. Motivated by an analysis of the components of a typical camera processing pipeline, we consider models of the form

$$y = g(C \cdot \kappa), \tag{1}$$

with $C \in GL(3)$ and $g\colon \mathbb{R}^3 \to \mathbb{R}^3$ a nonlinear function. Unlike most existing work, we do not expect C or g to be fixed properties of a camera, and instead, we assume them to be scene-dependent. The matrix C accomodates a camera's scene-dependent white balance operation as well as its transformation to a standard linear color space. The function g is a composite of the camera's scene-dependent color rendering processes and the standard compressive nonlinearity that is part of the sRGB specification.

We evaluate a series of models that increase in complexity. The simplest is a 12-parameter model comprised of an arbitrary matrix C and an independent per-channel exponentiation. The most complex is a an 8th degree polynomial from $\mathbb{R}^3$ to $\mathbb{R}^3$. Additionally, we test the linearity of the camera sensors and verify that the RAW measurement κ is a linear transform away from the CIE standard observer's reading of the corresponding SPD.

Our results suggest that for a large number of consumer cameras, the sensors are indeed a linear combination of standard CIE color matching functions. Furthermore, we find that a 24-parameter model, consisting of an arbitrary matrix C and an independent per-channel 5th-degree polynomial, is able to represent the nonlinear color processing pipelines (see Fig. 2), and we recommend this representation as providing a good balance between accuracy and model complexity. In particular, as shown in Fig. 1, it offers a significant improvement relative to the common assumption of a fixed per-channel exponentiation (or "gamma") of 2.2.

In order to fully exploit the Internet as a data source for computer vision, we must compensate for the scene-dependent nonlinear color processing performed by consumer cameras, and deriving and evaluating models like those proposed here is an important first step in this direction.

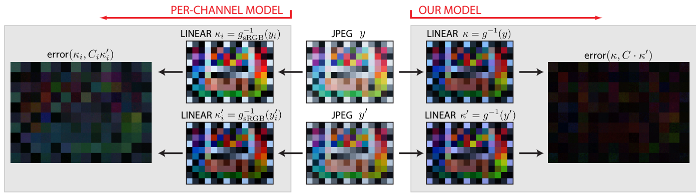

Figure 1: Modeling camera color processing. Pixel intensities are commonly assumed to be standard sRGB maps of spectral image irradiance, i.e., white-balanced linear RGB with standard per-channel nonlinearity. Alternatively, we seek a model that accounts for different cameras having different spectral sensors and nonlinear maps that may vary with scene content. *Middle*: Two JPEG images of the same scene under the same illuminant captured by different cameras; we seek to match their colors. *Left*: Images matched using the standard sRGB model with per-channel gain leads to high residual error (RMSE: 19.5 gray levels). *Right*: The proposed 24-parameter model properly accounts for variations across cameras, and achieves higher accuracy (RMSE: 7.5 gray levels).

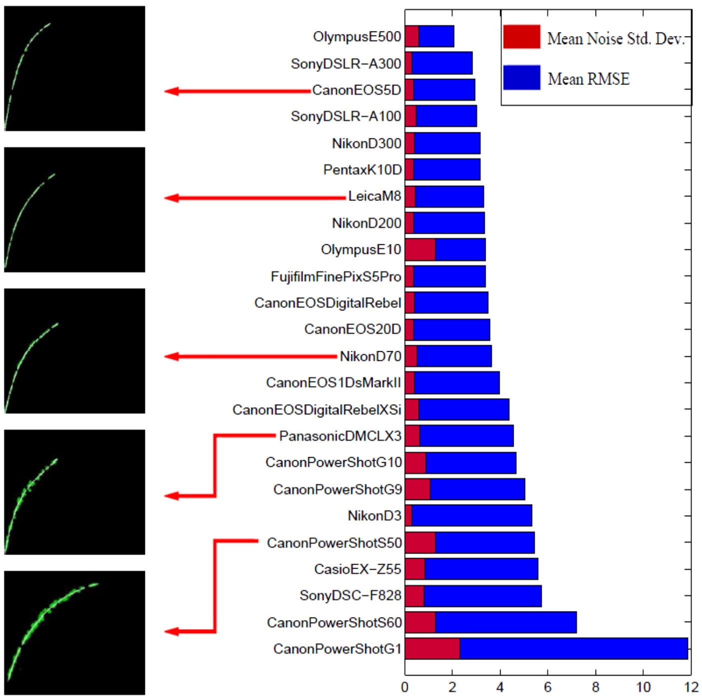

Figure 2: Quality of fit using a 24-parameter model for the RAW→JPEG maps of the 24 RAW-capable cameras in our database. *Right*: Bar graph showing mean RMSE values for each camera, with inset red bar showing mean noise standard deviation. *Left*: Joint histograms of the green channel of the output color vectors y and the green channel of the linearly-transformed RAW vectors $C \cdot \kappa$ for images from five different cameras. The estimated green-channel portion of the nonlinearity g is super-imposed in white. This demonstrates the differences that exist between the color processing pipelines of different cameras.