Linear Regression Enhancements to eXcalibrator
by
Bruce Waddington

Bruce Waddington

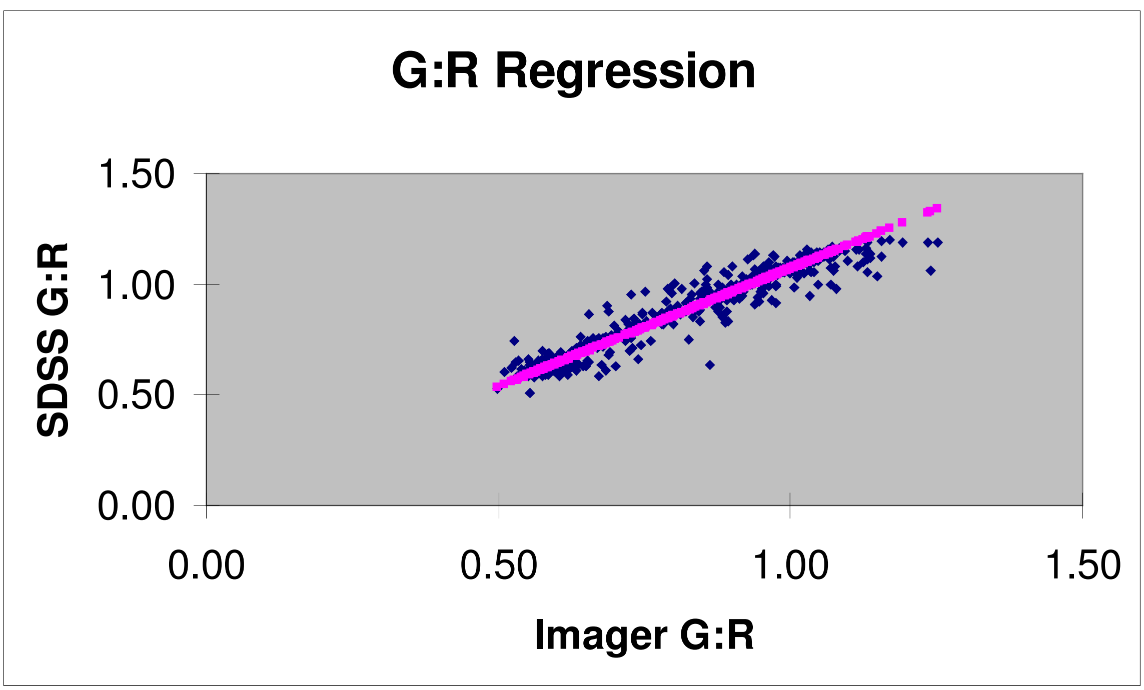
Background

The use of SDSS photometry data in eXcalibrator (eXcal) generally results in excellent color balance. With versions 2.0 and earlier, eXcalibrator confines the SDSS star sample to be “nearly white” and then computes averages for the green:red (G:R) and blue:red (B:R) ratios to “force whiteness.” In situations where amateur images are captured with small fields of view, this approach can result in very few or even zero stars being included in the “nearly white” sample, and the available stars may be skewed somewhat to the red or blue. When this happens, the color balance ratios computed by eXcal may be compromised. To overcome this limitation, we pursued the option of relaxing the “nearly white” constraint and using a linear regression (linear least-squares) model to fit the G:R and B:R ratios.

Linear Regression Algorithm

The steps required to do the linear regression are straightforward. Since we want to fit G:R and B:R ratios, we need to have SDSS red, green, and blue *flux values* rather than photometric magnitudes. Further, these values need to be transformed into filter passbands more typical of amateur imaging equipment. We chose the Johnson-Cousins photometric system as a better match for typical amateur imagers and then used published linear transforms from the SDSS ugr magnitudes into BVR magnitudes (Lupton 2005 as referenced at http://www.sdss.org/dr5/algorithms/sdssUBVRITransform.html#Lupton2005). We use the Lupton equations for B and V as functions of the SDSS g-r values. Conversions of the computed BVR magnitudes to their respective flux values use the standard Vega zero point equations (http://bf-astro.com/eXcalibrator/ZeroPoints.pdf), taking into account we are measuring photon fluxes rather than energy fluxes.   Once the magnitude-equivalent flux values and ratios are in place, the regression is done with standard equations where the y-intercept is forced to zero. Experimentation with an unconstrained y-intercept led to poor outcomes, so we chose the constrained approach.

For most targets we have tested so far, the linear fits are quite good, although the B:R ratio invariably has a higher variance than the G:R ratio. A fairly typical example of a G:R fit is shown below:

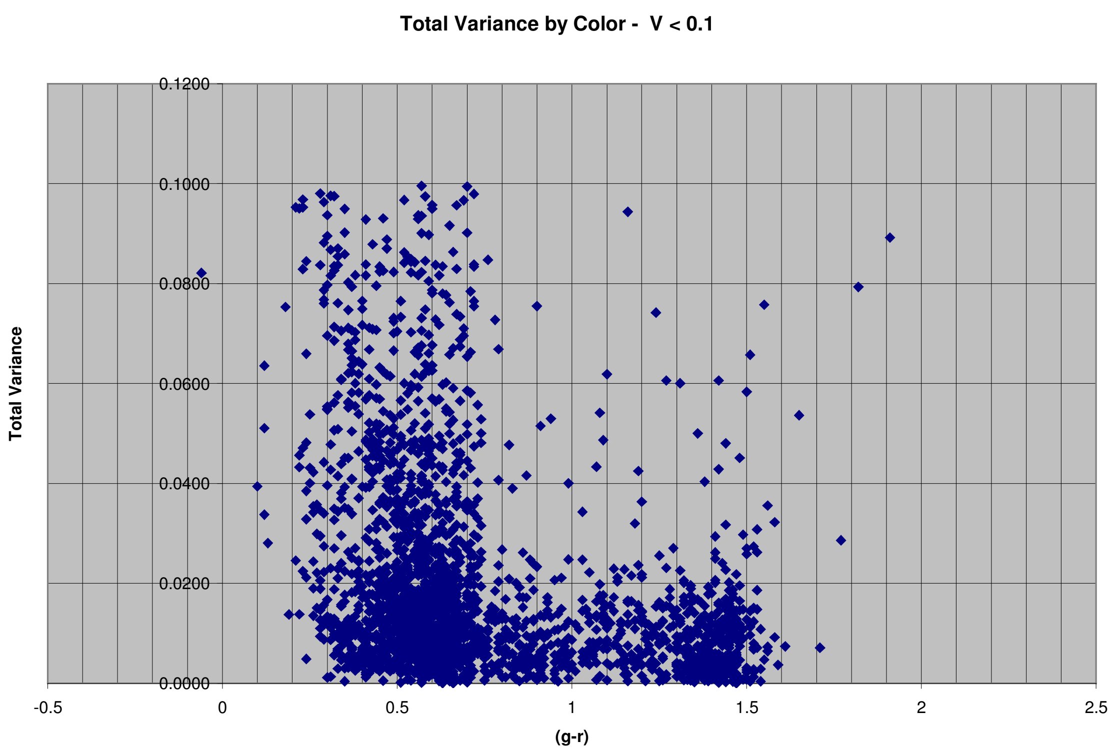


Obviously, the slope of the line in the graph above represents the “best fit” estimate of the G:R ratio for this target.

Filtering and Outlier Rejection

Since we are processing a wide range of SDSS star colors, some level of outlier rejection is mandatory. The raw data almost always contain some grossly mismatched star data, and these can have a significant negative impact on the first pass of the linear regression. Consequently, "gross outliers" are removed before any regression algorithms are even applied. This typically involves a small number of stars for any one target, usually in the range of 0-10. To perform subsequent, more refined, outlier rejection, we adopted the usual metric for "standard error of regression". This is the statistic used internally when the user clicks on 'reject outliers' and it is re-computed after each pass of outlier rejection. One or possibly two passes of outlier rejection typically produces results that quickly converge and subsequently remain stable. Even with a two-pass outlier rejection, we typically still have star samples of over 50 stars and often well over 100. This approach for handling outliers has thus far proven to be fairly robust even with relatively diverse samples of star colors.

Despite taking a broad selection of SDSS stars, some level of up-front filtering is still required. SDSS stars with magnitudes fainter than ~20 do not appear to be well-measured in the amateur images we have available, so those are excluded. We also found that stars at the extreme blue end of the spectrum had systematically higher variances, so we imposed an initial filter condition of (g-r) > 0.36. A similar dependency was noted at the red end of the spectrum although not nearly to the same degree. At the red end, taking stars with a (g-r) value of <= 2.0 avoids any obvious systematic errors. Of course, with the outlier rejection process, these limits need not be very precise. We were somewhat surprised by the breadth of stars that could safely be used in the regression algorithm and the relative unimportance of star color as a discriminator. The chart below shows a color dependency analysis of total variance for the union of all the stars used in all of the targets processed so far. In this case, only a small number of "grossly mismatched" stars were excluded, as described previously:



Conclusion

The linear regression model implemented in V3 of eXcal provides a complementary approach for computing color balance ratios with SDSS stars. It should be particularly helpful for imagers using small fields of view or who have a small number of available SDSS stars for some other reason. As with the other calibration options in eXcal, users are encouraged to try a number of approaches and develop a sense of how well they converge on what is likely to be the "best" set of calibration ratios.

end of document