

# The Ratio Method for Multi-view Color Constancy

Trevor Owens<sup>1</sup>, Kate Saenko<sup>1</sup>, Ayan Chakrabarti<sup>2</sup>, Ying Xiong<sup>2</sup>, Todd Zickler<sup>2</sup> and Trevor Darrell<sup>1</sup>  
<sup>1</sup>UC Berkeley, <sup>2</sup>Harvard University

<sup>1</sup>{trevero, saenko, trevor}@eecs.berkeley.edu <sup>2</sup>{ayanc@eecs, yxiong@seas, zickler@seas}.harvard.edu

## Abstract

*Color constancy is the ability to infer stable material colors despite changes in lighting, and it is typically addressed computationally using a single image as input. In many recognition and retrieval applications, we have access to image sets that contain multiple views of the same object in different environments; we show in this technical report and a related publication [8], that correspondences between these images provide important constraints that can improve color constancy. In this report, we present another method to solve the multi-view color constancy problem, the Ratio Method. This method provides a means to recover estimates of underlying surface reflectance based on joint estimation of these surface properties and the illuminants present in multiple images. In contrast to the multi-view Spatial Correlations method (MVSC), this method can leverage any single image color constancy method as a bootstrap for the multi-view solution. The method exploits image correspondences obtained by various alignment techniques, and we show examples based on matching local region features. Our results show that the Ratio Method performs similarly to the MVSC method, both of which are improvements over a baseline single-view method.*

## 1. Introduction

In addition to its intrinsic reflectance properties, the observed color of a material depends on the spectral and spatial distributions of its surrounding illumination. Thus, in order to use color as a reliable cue for recognition, we must somehow compensate for these extrinsic factors and infer a color descriptor that is stable despite changes in lighting. The ability to make this inference—termed *color constancy*—is exhibited by the human visual system to a certain degree, and there are clear benefits to building it into machines.

One important part of computational color constancy, and the part we consider in this paper, is compensating for the “color cast” that affects a scene as a whole. For this, one

ignores spatial variations in lighting spectra, and makes the assumption that the spectrum of the illuminant is approximately uniform throughout the scene. The problem is then one of inferring the map  $M: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  from the space of observed tristimulus vectors (RGB, human cone, etc.) to the space of *canonical colors*—those that would have been obtained for the same scene by a standard observer under a standard spectral power distribution.

This color constancy problem has traditionally been addressed using single images as input. Accordingly, one begins with a model for the distribution of canonical colors in a typical scene and then chooses a map that takes the input image colors to a distribution that matches this model. Examples of this approach include gamut mapping [5], gray world [2], white-patch [3], and gray edge [10] algorithms in addition to Bayesian formulations [1, 9, 6].

This basic approach ignores the fact that, increasingly, we have simultaneous access to *image sets* that contain multiple views of the same salient objects. For example, Fig. 1 shows images of three different objects, each acquired in a distinct environment with unknown illumination. Techniques for finding correspondences between object instances in multiple images are well known, e.g., based on local invariant features (SIFT, etc.). Such correspondences provide additional constraints that can improve single-image color constancy methods: the per-image maps  $M$  can only be correct if the toy’s resulting canonical colors agree between the images.

Given a set of images with associated correspondences our goal is to simultaneously infer the maps that compensate for the color cast in each image. In a typical scenario, these point correspondences are derived from patches on an object (Fig. 1), and when this is the case, our approach automatically provides the object patch’s canonical color descriptor as a by-product, with this descriptor being jointly optimized with respect to the multiple input views.

## 2. Background

This section parallels the background description in the related published work [8]. Let  $f(\lambda) = (f^1(\lambda), f^2(\lambda), f^3(\lambda))$  be the three spectral filters of a lin-



Figure 1. Each row depicts an object observed in three different environments with distinct and unknown illumination, and with different unknown cameras. For each row, what is the object’s intrinsic color? Can knowledge of object-level correspondences improve estimation of the illuminant in each image? The top row shows images of a abacus, the middle row a baby’s toy, and the bottom row a popular cartoon character. The latter is from the series “One object 365 days”, a popular group on flickr.com

ear sensor, and denote by  $y_p$  the color measurement vector produced by these filters at pixel location  $p$ . Assuming perfectly diffuse (Lambertian) reflection, negligible inter-reflections between surface points, and a constant illuminant spectrum throughout the scene, we can write

$$y_p = (y_p^1, y_p^2, y_p^3) = \int f(\lambda) \ell(\lambda) x(\lambda, p) d\lambda, \quad (1)$$

where  $\ell(\lambda)$  is the spectral power distribution of the illuminant, and  $x(\lambda, p)$  is the spectral reflectance of the surface at the back-projection of pixel  $p$ . Our goal is to infer a canonical color representation of the same scene, or one that would have been recorded if the illuminant spectrum were a known standard  $\ell_s(\lambda)$ , such as the uniform spectrum (Illuminant E). We express this canonical representation as

$$x_p = (x_p^1, x_p^2, x_p^3) = \int f(\lambda) \ell_s(\lambda) x(\lambda, p) d\lambda, \quad (2)$$

and we seek to obtain it by inferring the map  $y_p \mapsto x_p$ .

We follow convention by parameterizing this map using a linear diagonal function, effectively relating input and canonical colors by

$$y_p = M x_p, \quad (3)$$

with  $M = \text{diag}(m^1, m^2, m^3)$ . According to this model, the input color at every pixel is mapped to its canonical counterpart by gain factors that are applied to each channel independently. This process is termed *von Kries adaptation*,

and the conditions for its sufficiency are well understood. It can always succeed when the filters  $(f^1(\lambda), f^2(\lambda), f^3(\lambda))$  do not overlap, and in this case it is common to refer to the parameters  $m \triangleq (m^1, m^2, m^3)$  as the “illuminant color”, and the canonical values  $x_p$  as being the “scene reflectance” (e.g., [9]). For overlapping filters, including those of human cones and a typical RGB camera, the mapping  $y_p \mapsto x_p$  need not even be bijective, but von Kries adaptation can succeed nonetheless provided that the “world” of spectral reflectances  $x(p, \lambda)$  and illuminants  $\ell(\lambda)$  satisfy a tensor rank constraint [4]. Regardless, for the remainder of this document, we will use the terms “illuminant” and “reflectance” to reference  $m$  and  $x_p$ , respectively.

### 3. Multiview Color Constancy Model

The ratio method combines ratio constraints derived from multi-view correspondences with initial illuminant estimates from monocular algorithms.

The key idea in this model is the exploitation of multi-view color correspondence constraints. Here the correspondence is not geometric, but exists in color space. This correspondence occurs whenever the measured color of a patch,  $p_1$  in image 1 and another patch  $p_2$  in image 2 are equivalent, when viewed under the same illuminant. So given  $M_1 = M_2$ , we have:  $y_{1,p_1} = y_{2,p_2}$ , which is equivalent to the statement,  $x_{1,p_1} = x_{2,p_2}$ . We denote this reflectance as  $x_p$ , since it is independent of the image.

Often, when training on specific instances of objects, we have multiple instances of the same object in different scenes and hence different illuminants, as in Figure 1. In this case we have corresponding color patches, which gives us a constraint on the illuminants. Since we have patches  $p$  in common among images  $i$ , we have satisfied the above requirement that we have a reflectance  $x_p$  in common among all images and hence, we obtain a constraint on the possible illuminants in different images at each shared patch. In contrast to monocular models, when jointly modeling the illuminants of an object in multiple views we must consider a possible shading term per patch per image,  $\alpha_{i,p}$ .

$$\alpha_{i_1,p} M_{i_1}^{-1} \vec{y}_{i_1,p} = \vec{x}_p = \alpha_{i_2,p} M_{i_2}^{-1} \vec{y}_{i_2,p} \quad (4)$$

#### 3.1. Multi-view Shading Normalization

When considered in the raw space, the shading terms  $\alpha_i$  add an unknown variable for each added constraint, yielding an ill-posed problem. We could assume a smooth model of shading variation and regularize the solution using a variational penalty term, but in real images shading can vary quite abruptly.<sup>1</sup> Instead, we cast observed pixels into a nor-

<sup>1</sup>We could also use a locally planar model, which would be true for our experimental dataset, but not true in general. In our experiments we make no assumption of local surface geometry.

malized color space where the shading terms cancel:

$$\frac{\vec{y}_{i,p}}{\|\vec{y}_{i,p}\|} = \frac{M_i \vec{x}_p \alpha_i}{\|M_i \vec{x}_p \alpha_i\|} = \frac{M_i \vec{x}_p \alpha_i}{\|M_i \vec{x}_p\| |\alpha_i|} = \frac{M_i \vec{x}_p}{\|M_i \vec{x}_p\|} \quad (5)$$

Therefore  $\vec{y}_{i,p}' = (M_i \vec{x}_p)'$ , using the following notation for a normalized vector:  $v' = v/|v|_\rho$ . For  $\rho = 1$  (the  $L_1$  norm) this normalization is equivalent to using  $rg$  chromaticity space: for each pixel we project the  $rgb$  space to the  $rg$  space, e.g.  $R' = \frac{R}{R+G+B}$ . For the sake of convenience in the following derivations we drop the normalized notation, and simply use  $x$  and  $y$  to refer to their normalized quantities,  $x = x', y = y'$ , so that we can write our color model as:

$$\vec{y}_{i,p} = M_i \vec{x}_p \quad (6)$$

### 3.2. Ratio Method

We take the color model (6) and Von Kries adaptation as assumptions. Under these assumptions, we can form a constraint for each color channel, on the ratio of the illuminants  $m_1^c/m_2^c$  between images 1 and 2, when the images share a patch  $p$  with the same reflectance  $x_p$  in common. Component-wise, letting  $c \in [1, 2, 3]$  denote the color channel,  $i \in [1, \dots, n]$  the image number and  $p \in [1, \dots, m]$  the patch number, this constraint looks like:

$$\forall i_1 \neq i_2 \quad \frac{y_{i_1,p}^c}{m_{i_1}^c} = \frac{y_{i_2,p}^c}{m_{i_2}^c} \quad (7)$$

Solving for  $m_i^c$ 's we have:

$$\forall i_1 \neq i_2 \quad \frac{m_{i_2}^c}{m_{i_1}^c} = \frac{y_{i_2,p}^c}{y_{i_1,p}^c} \quad (8)$$

If our measurement of the transmittance of light integrated over a period of time in the camera was exact, then this would be a hard constraint on the illuminant in one image, given that in another. Since this is not the case, we model the noise in pixel values as  $\epsilon_y$ . In addition, since (8) only gives us a constraint on the ratio of the illuminants, we must have an initial *bootstrap* estimate of the illuminant in each image, to infer the other illuminant. So let  $\hat{m}_i$  denote the bootstrap estimate of the illuminant in image  $i$ , which has noise  $\epsilon_m$ . Under our noisy model assumption,

$$\hat{y}_{i,p}^c = y_{i,p}^c + \epsilon_y \quad (9)$$

$$\hat{m}_{i_1}^c = m_{i_1}^c + \epsilon_m \quad (10)$$

where  $\epsilon_\beta \sim \mathcal{N}(0, \gamma_\beta)$ , and  $\beta \in \{y, m\}$ . We assume that the standard deviation of the noise is independent of the measurement value. Since we have multiple estimates, we are interested in the expected value of the illuminant. From (8), we have:

$$E[\hat{m}_{i_2}^c] = E\left[\frac{\hat{y}_{i_2,p}^c}{\hat{y}_{i_1,p}^c} \hat{m}_{i_1}^c\right] = E\left[\frac{y_{i_2,p}^c + \epsilon_y}{y_{i_1,p}^c + \epsilon_y} (m_{i_1}^c + \epsilon_m)\right] \quad (11)$$

Expanding the right-hand side into four terms and assuming the noise terms are independent of one another, three terms are 0 and we have:

$$E[\hat{m}_{i_2}^c] = E\left[\frac{y_{i_1,p}^c m_{i_1}^c}{y_{i_1,p}^c} \frac{1}{1 + \frac{\epsilon_y}{y_{i_1,p}^c}}\right] \approx \quad (12)$$

$$E\left[\frac{y_{i_1,p}^c m_{i_1}^c}{y_{i_1,p}^c} \left(1 - \frac{\epsilon_y}{y_{i_1,p}^c}\right)\right] = E\left[\frac{y_{i_1,p}^c m_{i_1}^c}{y_{i_1,p}^c}\right] \quad (13)$$

Where we have used the power series expansion of  $(1 + x)^{-1}$  for small  $x$ . Hence, for our approximation and assumptions to be correct, we must discard patches which have small values or saturate in any channel.

Given an existing color constancy method, we first form the estimate of the illuminant conditioned on the non-object portion of the input images, and then apply the ratio constraint and infer refined illuminant estimates using the average of the ratios or finding the median ratio values, as described below. In practice, in order to have an unbiased estimate of the mean or median illuminant, we first take the log of the ratio and add 1 to insure all values are positive. If we assume all of the illuminants have a uniform probability distribution, then we can simply perform inference by averaging to find the mean illuminant estimate over color channels.

Since each term in (7) is equal to patch reflectances,  $x_p^c$ , per color channel, which are equivalent across the images, these ratios give us multiple samples of the same ratio under different illuminants. Since we are considering a noisy model, we can choose the average as the maximum likelihood estimator of the mean of these ratios. Alternatively, we could take the median value among these estimates. Another possible approach, which we have not experimented with, would be to estimate the distribution of illuminants and then discard outliers before applying either estimate of the center of the distribution. Below we present the averaging method to estimate the most likely ratio that explains the distribution of ratios we see in the images.

We use the initial estimate of the illuminant for color channel  $c$  from a monocular color constancy method (such as gray world) on image  $j$  to estimate  $m_j^c$ . We can then estimate illuminant colors in all images by averaging the ratio constraint over patches,  $p$ , using only the initial estimate for this image  $j$ :

$$\bar{m}_i^c(j) = \frac{1}{P} \sum_{p=1}^P \frac{y_{i,p}^c}{y_{j,p}^c} m_j^c \quad (14)$$

This, however, biases the estimate of all illuminants based on the initial estimate of image  $j$ . So as not to favor any single image illuminant estimate, we average over all possible initial single image illuminant estimates to get our final





Figure 2. In the color checker database color patches on calibrated color checker charts were used for the image patch correspondences.

illuminant estimate:

$$\bar{m}_i^c = \frac{1}{N} \sum_{j=1}^N \bar{m}_i^c(j) \quad (15)$$

Using these illuminant estimates, we can form an estimate of the patch colors  $x_p^c$  by averaging over the images:

$$\bar{x}_p^c = \frac{1}{n} \sum_{j=1}^n \frac{y_{j,p}^c}{\bar{m}_j^c} \quad (16)$$

This algebraic method is very fast and extends symmetrically to any number of patches, images and even extra color channels<sup>2</sup>. Now consider  $M$  such single image color constancy methods, which we index by  $m$ . Index  $m_i^c(j)$  in equation (14) by its single view method:  $m_i^c(j)(m)$ . We can then average over these (by assumption, independent) single image illuminant estimates to avoid any shortcomings of any single method. Hence, the ratio method also works with any, or multiple single image color constancy methods.

#### 4. Establishing Color Correspondences Between Views

Our method relies on finding corresponding homogeneous color regions across several images. This is done using standard alignment techniques given an object in common among several views. A watershed method is used for finding stable regions. We refer the reader to the related publication [8] for details on establishing these correspondences. For the remainder of the document, we assume these correspondences have been found reliably for both datasets.

### 5. Results

The ratio method is evaluated in two ways. To more effectively evaluate how varying number of patch correspondences effect the methods, we use a large database with color checker charts in each image. The corresponding patches in this case are provided by the color checker chart.

<sup>2</sup>We will release a MATLAB version of our implementation for unrestricted research use.

The second evaluation metric is computed using images in the wild. We use the same data set of 26 images, each with a color checker chart, split into sets which share an object in common as described in the related paper [8].

#### 5.1. Color Checker Database

The first database contains 568 images acquired by Gehler et al. [6]. This dataset has been split into indoor scenes (246 images) and outdoor scenes (322 images), and we report results on each of these sets separately, in addition to results on the 568 image-set as an undistinguished whole. (We refer to these as *indoor*, *outdoor*, and *undistinguished*, respectively.) Each scene contains a color checker that provides the ground-truth illuminant  $m_o$  for that image, and the performance of the color constancy algorithms is quantified by computing the root mean squared error (RMSE) between the “chromaticity” of the inferred illuminant,  $(m^1, m^2)/(m^1 + m^2, m^3)$  and that of the ground-truth illuminant. In figures 4 and 3 we report the mean RMSE. In addition, to compare the Ratio Method with the multi-view Spatial Correlations method, as in figures 5 and 6, we report illuminant estimates using a similar consistent measure of error, the angular error,  $\theta = \arccos \frac{\langle m_1, m_2 \rangle}{\|m_1\| \|m_2\|}$ . There is a simple conversion between the two measures. Suffice it to say that doing better in one metric implies doing better in the other. We follow the evaluation protocol of [6] and mask out the color checker charts before using the images as input to any inference algorithm.

As described in the section 3.2, the ratio method approach can be paired with a variety of single-image algorithms. We choose two different approaches for our evaluation. The first is the grey-edge algorithm [7], with parameter-sets optimized for either indoor scenes ( $\sigma = 1$ ,  $n = 0$ ,  $p = 1$ ) or outdoor scenes ( $\sigma = 1$ ,  $n = 2$ ,  $p = 5$ ). (See [6] for optimization details.) The second is the single-image Bayesian color constancy approach [6]. For each of these approaches, we compare the performance of the original single-image algorithm to those obtained by augmenting them with multi-view information, using both the Ratio Method and the multi-view Spatial Correlations method.

Before making comparisons between the single-image and multiview cases, we first assess the performance of the multiview approach as we increase the color diversity of the correspondences between images. As depicted in Fig. 2, we use manually-identified squares from the color checker chart to simulate the matches that are provided by an object detector as described in section 4. That is, each selected square provides a reflectance  $x$  that is “shared” between the multiple images. Figure 3 shows the mean RMSE as the number of corresponding patches increases. For this test, we use two images as input and we average the results over the whole *outdoors* set uniformly at random. We pair the

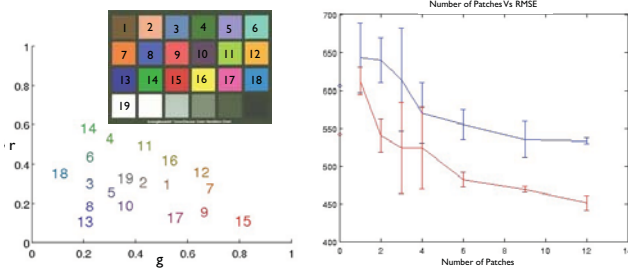


Figure 3. Performance with increasing number of patches on outdoor images in the color checker database. The blue line represents the RMSE of the Ratio Method using the grayedge algorithm as the single image color constancy method. The red line represents the RMSE of the Ratio Method using Gehler’s Bayesian color constancy method. The diamonds on the left axis represent the baseline methods, grayedge (blue) and bayesian (red). Two images are used as input, drawn randomly from the undistinguished set, and we run our algorithm with a growing set of corresponding patches. We select a single random order of these patches, indexed 15, 18, 9, 10, 11, 12, 7, 3, 19 and test this on many pairs of images. In the related publication [8], we found that if we chose several orders of patches and averaged, the corresponding curves decrease more smoothly as we increase the number of patches and on average the performance for only a single patch in common was on par with the single image method. As can be seen here, as the number of patches grows, the confidence in the value increases and the RMSE decreases. The plot on the left is the gamut of the macbeth color checker chart in  $rg$ -chromaticity space. It appears that the closer the patches selected are to the center of the gamut, the fewer patches that are needed for an accuracy equivalent to using the whole set. Patches with colors far from the center tend to skew the estimate. For another interpretation of why certain patches perform more poorly, see the discussion on “angular spread” in section 5.1 of the related publication

images randomly several times to assess the confidence in the RMSE. The baseline error numbers reported (plotted at the zero patches line) are obtained by running the single-image version of each algorithm on the entire set. The ratio method may perform slightly worse than the baseline when only one patch is available in correspondence. In the real world dataset experiments, we found that we always had several patches in common, so this should not be a problem, in practice. A potential solution to this problem would be to avoid using the Ratio Method in situations where 1 or fewer patches are in common. The multi-view spatial correlations method, on the other hand, deals with this issue by forming an optimization problem which weights the multi-view constraint and single image constraint appropriately and when only one or fewer patches are in common, performs no worse than the single view color constancy baseline.

For a more thorough comparison between the single-image and multiview scenarios, we run an additional set of tests with 8 matched patches but with an increasing input set

RMSE $\times 10^4$ 8 patches	Gray Edge	Multiple Images Ratio Method (Gray Edge, Median)			Bayesian	Multiple Images Ratio Method (Bayesian, Median)		
Images		2	5	10		2	5	10
Indoors	91 $\pm$ 9	75 $\pm$ 5	72 $\pm$ 4	69 $\pm$ 4	93 $\pm$ 7	79 $\pm$ 6	69 $\pm$ 7	66 $\pm$ 4
Outdoors	61 $\pm$ 9	54 $\pm$ 5	50 $\pm$ 3	50 $\pm$ 2	46 $\pm$ 2	43 $\pm$ 3	32 $\pm$ 1	31 $\pm$ 1
Average	76 $\pm$ 9	65 $\pm$ 5	63 $\pm$ 4	43 $\pm$ 3	69 $\pm$ 6	61 $\pm$ 5	50 $\pm$ 4	49 $\pm$ 4
Undistinguished	73 $\pm$ 7	64 $\pm$ 4	62 $\pm$ 4	61 $\pm$ 5	69 $\pm$ 5	61 $\pm$ 7	58 $\pm$ 4	53 $\pm$ 2

Figure 4. Average RMSE of the illuminant estimates from the ground truth illuminant as given by averaging the gray patches on the color checker chart. From this, you can see that the Ratio Method improves over two single view color constancy methods, Gehler’s Bayesian Color Constancy Method [6] and the Gray Edge Method [7].

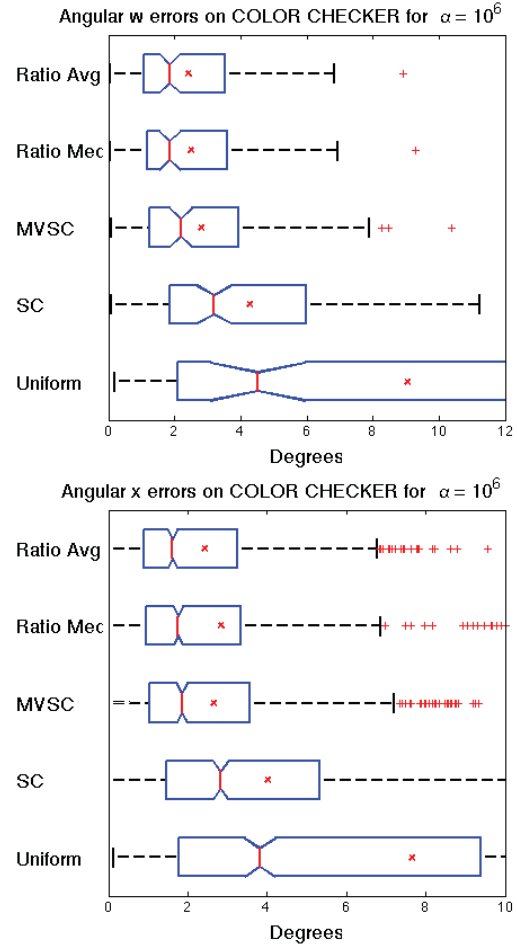


Figure 5. Angular errors for the Ratio Method using the Median and Average of the single view Spatial Correlations method, the multi-view and single view Spatial Correlations methods and the naive, uniform illuminant estimate on the COLOR CHECKER dataset. These are calculated using four images and 6 randomly selected patches (among the 12 used for testing the MVSC method) in correspondence. The  $\alpha$  parameter in the MVSC method is chosen using cross validation on a different subset of patches.

size. The results are shown in Fig. 4. Both the gray edge and Bayesian single-image approaches are trained twice—once

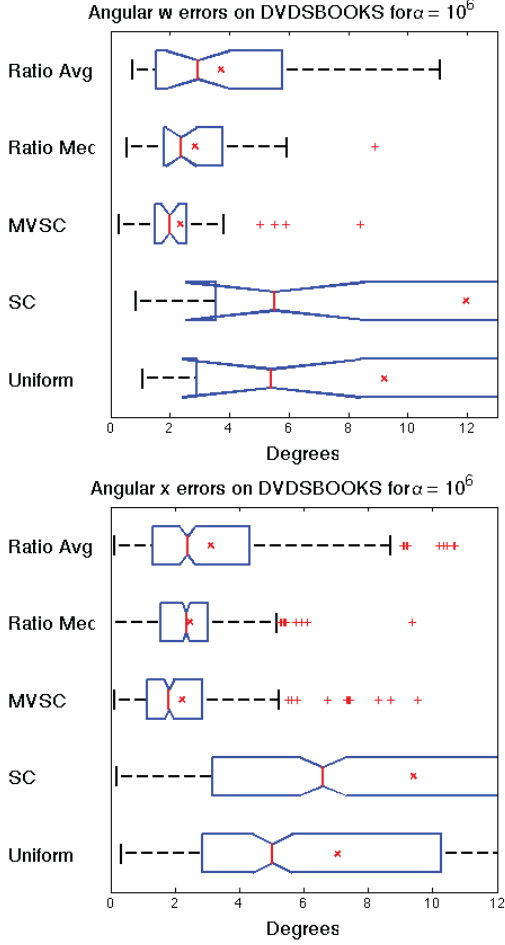


Figure 6. Angular Errors for the same methods as in figure 5 on the real-world dataset of DVDSBOOKS. The  $\alpha$  parameter in the multi-view Spatial Correlations method was chosen using the larger COLOR CHECKER data set. The spatial correlations method performs poorly on this data set as there are a few backgrounds repeated which had very little texture.

for the indoor and once for the outdoor sets—and we report the error for each algorithm trained on the appropriate test set. For the undistinguished set, we report the better of the grayedge indoor tuned algorithm and grayedge outdoor tuned algorithm. The Bayesian algorithm is run using a prior which was learned on the undistinguished set. As before, the multiview results are obtained by incorporating the single-image algorithm into our framework, and averaging the error over the whole set divided up into the appropriate number of images, either 2, 5, or 10. No image is used more than once in the trial. Image-sets are selected from the respective datasets uniformly at random. The confidence in the RMSE is calculated by testing on the whole data set repeatedly each time with different, randomly selected image-sets. For all of the single-image algorithms that we tested, the multi-view framework provides

an improvement. The error decreases with the number of images and patches used in the correspondence.

Another comparison<sup>3</sup> is made between the Ratio Method and Spatial Correlations method in figures 5 (on this COLOR CHECKER data set) and 6 on the DVDSBOOKS data set. Both the Ratio Method and the MVSC method improve over the baseline single image color constancy methods. There appears to be negligible difference between the MVSC method and the Ratio Method on this data set.

## 5.2. Real World Object Database

To evaluate our method on a real-world task, we make use of the DVDSBOOKS “real world data set”, as described in [8]. This data set is composed of 39 images with 5 different objects in different scenes, under natural illuminants, both indoors and outdoors. Some illuminants are heavily colored and thus present a significant challenge for most single image color constancy methods. Each image additionally contains a color checker chart, used only to determine the ground truth illuminant, in the same way as for the colorchecker dataset. The colorcheckers are all in full view and are oriented towards the dominant source of light in the scene. As described in section 4, we automatically find patch correspondences between images with the same object in common. This provides a more realistic setting under which the multiview algorithm can be used for actual objects, compared to the color checker chart patches as described in 5.1. Several of objects have over 100 patches in common between all images; most have on the order of 40 stable regions in common. In our experiments, we use the matches corresponding to the top 10 largest MSER regions. Decreasing the number of patches to 10 puts an upper bound on the error we would reasonably expect from the method.

As in the colorchecker database evaluation, we mask out the color checker and object in each scene to obtain the single image illuminant estimate. The results are shown in Figure 6. For some scenes, as described in [8], this data proved quite difficult for the single-view method, which was trained on the large color checker dataset for these experiments. Indeed, the monocular SC performance is worse than the uniform illuminant estimate. The Ratio Method again solidly outperforms the single view color constancy methods. The difference between the Ratio Method and

<sup>3</sup> Plots are presented using MATLAB boxplot format. The line in the middle of the box is the median error. The red x shows the mean error. The edge of the box shows the edge of the first and third quartiles of errors. The box whiskers extend to the highest and lowest errors not considered outliers. The edge of some whiskers were cropped to focus on the middle 50 percent of errors, contained inside the boxes. Outliers are plotted separately. A Kolmogorov-Smirnov goodness-of-fit test shows the errors to be normally distributed at a 95% confidence level. The notches around the median denote two standard deviations above and below the median error. While the box of the multiview method errors overlaps with that of the single image method errors, the notches in the box do not overlap, suggesting statistical significance.

MVSC method is more pronounced in this test.

## 6. Conclusion

We have presented the Ratio Method, another multi-view color constancy method, similar to the MVSC method, as described in [8]. We presented experiments on two databases, a standard color constancy dataset and a real-world dataset of objects. Our results show that multi-view constraints can significantly improve estimates of both scene illuminants and true object color (reflectance) when compared to baseline methods. Our method performs well even when monocular estimates are worse than a uniform baseline. This method is more flexible than the MVSC method in that it can use any baseline single image color constancy method, and moreover, be used with several single image baseline methods at once. In addition, the Ratio Method using the median of the ratios provides a certain measure of robustness to outliers.

## References

- [1] D. Brainard and W. Freeman. Bayesian color constancy. *Journal of the Optical Society of America A*, Jan 1997. 1
- [2] G. Buchsbaum. A spatial processor model for object colour perception. *J. Franklin Inst.*, 310(1), 1980. 1
- [3] V. Cardei and B. Funt. Committee-based color constancy. In *Proc. IS&T/SID Seventh Color Imaging Conf.: Color Science, Systems, and Applications*, 1999. 1
- [4] H. Chong, S. Gortler, and T. Zickler. The von kries hypothesis and a basis for color constancy. *Proc. ICCV*, 2007. 2
- [5] D. A. Forsyth. A novel algorithm for color constancy. *Int. J. Comput. Vision*, 5(1), 1990. 1
- [6] P. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp. Bayesian color constancy revisited. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–8, 2008. 1, 4, 5
- [7] A. Gijsenij, T. Gevers, and J. van de Weijer. Physics-based edge evaluation for improved color constancy. *IEEE Conference on Computer Vision and Pattern Recognition*, 2009. 4, 5
- [8] T. Owens, K. Saenko, A. Chakrabarti, Y. Xiong, T. Zickler, and T. Darrell. Learning object color models from multi-view constraints. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Color Springs, Colorado, June*, pages 169–176, 2011. 1, 4, 5, 6, 7
- [9] C. Rosenberg, T. Minka, and A. Ladsariya. Bayesian color constancy with non-gaussian models. *Proc. of NIPS*, 2003. 1, 2
- [10] J. Weijer, T. Gevers, and A. Gijsenij. Edge-based colour constancy. *IEEE Trans. on Image Processing*, 16:2207–2214, 2007. 1