Learning Object Color Models from Multi-view Constraints

Trevor Owens[1], Kate Saenko[1], Ayan Chakrabarti[2], Ying Xiong[2], Todd Zickler[2] and Trevor Darrell[1]
[1]UC Berkeley, [2]Harvard University
[1]{trevoro,saenko,trevor}@eecs.berkeley.edu[2]{ayanc@eecs,yxiong@seas,zickler@seas}.harvard.edu

Abstract

Color is known to be highly discriminative for many object recognition tasks, but is difficult to infer from uncontrolled images in which the illuminant is not known. Traditional methods for color constancy can improve surface reflectance estimates from such uncalibrated images, but their output depends significantly on the background scene. 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 paper that correspondences between these images provide important constraints that can improve color constancy. We introduce the multi-view color constancy problem, and present a method to recover estimates of underlying surface reflectance based on joint estimation of these surface properties and the illuminants present in multiple images. The method can exploit image correspondences obtained by various alignment techniques, and we show examples based on matching local region features. Our results show that multi-view constraints can significantly improve estimates of both scene illuminants and object color (surface reflectance) when compared to a baseline single-view method.

1. Introduction

While it is well known that color is very discriminative for many object recognition and detection tasks, contemporary methods rarely attempt to learn color models from images collected in uncontrolled conditions where the illuminant is not known. This is because illuminant color can have a huge effect on the reported image color (see Fig. 1), and when the illuminant is not known, the learned color model ends up having limited discriminative utility in new images.

One approach to dealing with this difficulty is to compensate for illumination effects by applying a traditional color constancy technique to each training image. This approach has limited benefit, however, because methods for computational color constancy—including various forms of Bayesian estimation [3, 18, 11], gamut mapping [9], and "gray world" [4, 20, 5]—rely on prior models for the distribution of surface reflectances in a scene. This means that a color model that is learned for a particular object can be heavily influenced by the background scenes that happen to appear during training, and again, its utility is limited.

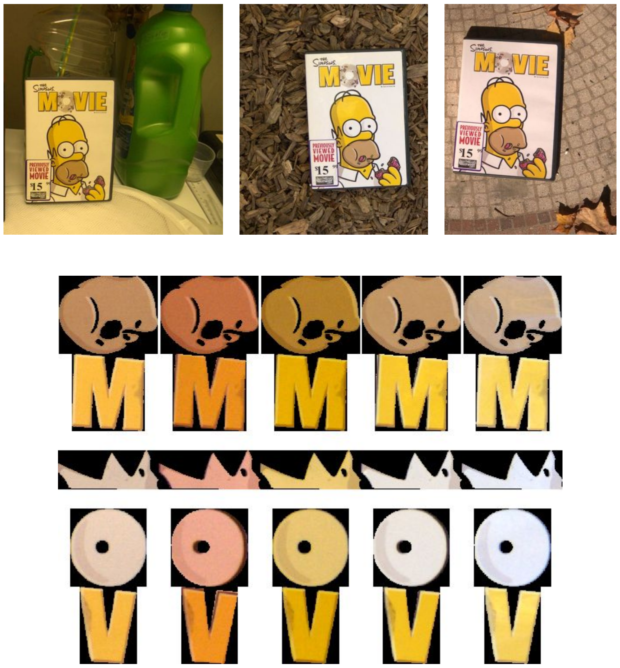


Figure 1. *Top row*: An object in three different environments with distinct and unknown illuminants. *Bottom rows*: Five local regions from the same object, extracted from five uncontrolled images like those above, demonstrating the extent of the variation in observed color. Our goal is to jointly infer the object's true colors and the unknown illuminant in each image.

This paper exploits the simple observation that when learning object appearance models from uncontrolled imagery, one often has more than one training view available. Figure 1 shows an example where images of an object are acquired in distinct environments with unknown lighting. When images like these are available, we can establish region-level correspondences between images and use *multi-view color constraints* to simultaneously improve our object color model and our estimates of the illuminant color in each image. This can be useful, for example, for obtaining better object color estimates for image-based modeling from uncontrolled images, and for object recognition in domains where the training images are uncontrolled but the sensor is known or color-calibrated at run-time (*e.g.*, on a robot or mobile phone).

Our task can be interpreted as one of *multi-view color constancy*, where we leverage well-known techniques for establishing correspondence between images (*e.g.*, spatial interest points and invariant descriptors) and re-formulate the color constancy problem in terms of joint inference across multiple images. When correspondences are derived from patches on an object, this process naturally provides a more useful object color model, one that is jointly optimized with respect to the multiple input views.

We evaluate our multi-view approach using the database of Gehler et al. [11] as well as a new real-world database containing books and DVD covers. Our results suggest that multi-view constraints are effective in improving color constancy and learning object color models from uncontrolled data, particularly in cases where the training set includes images that are challenging for traditional single-image color constancy.

2. Background and Related Work

In order to use color as a reliable cue for recognition, we seek to compensate for illumination effects and learn an object color model that is stable despite changes in lighting (and variations in the background scene). An alternative to this approach would be to use color invariants for recognition. When using an invariant, one avoids the explicit estimation of object colors and instead computes a feature, such as a color ratio [10, 13] or something more sophisticated [8, 12], that does not change with illuminant spectrum. Invariant-based approaches have been shown to improve recognition performance relative to monochromatic object appearance models, but in exchange for their invariance, they necessarily discard information that could otherwise be used for discrimination. This is perhaps why the human visual system does not solely rely on invariants, as evidenced by the existence of color names [2] and the role of object memory in the perception of color [16].

Our work is related to that of Barnard et al. [1], who exploit correspondences between points in a single image when inferring the spatially-varying illumination of the underlying scene. It is also related to image-based modeling for outdoor scenes, where one seeks to infer color-consistent texture maps from images taken under varying lighting conditions (*e.g.*, [14, 19]).

2.1. Computational Color Constancy

To define notation, let $f(\lambda) = (f^1(\lambda), f^2(\lambda), f^3(\lambda))$ be the three spectral filters of a linear sensor, and denote by y_p the color measurement vector produced by these filters at pixel location p in a single image. Assuming perfectly diffuse (Lambertian) reflection, negligible mutual illumination between surface points, and a constant illuminant spectrum throughout the scene, we can write

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

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. Given only a single image, we use the phrase *color constancy* to mean the ability 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 or Illuminant E. We express this canonical representation as

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

and achieving (single image) color constancy then requires 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

$$y_p = Mx_p, \qquad (3)$$

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 "reflectance" (e.g., [18]). For overlapping filters, including those of human cones and a typical RGB camera, the mapping $y_p \mapsto x_p$ need not be bijective, and von Kries adaptation can only succeed if the "world" of possible spectral reflectances $x_p(\lambda)$ and illuminants $\ell(\lambda)$ satisfies a tensor rank constraint [6] and an optimized ("sharpening") color transform is applied to all measurements [6, 7]. While our cameras' filters overlap, we do not employ a sharpening transform in this paper, and our results therefore give a conservative estimate of what is achievable with our approach.

3. Multi-view Color Constancy

The key idea in our approach is to exploit correspondence constraints between multiple views when attempting color constancy. When a set of images contains one or more common objects, the fact that these objects share the same underlying reflectance provides us with additional information about the illuminant in each scene. In this section, we describe the constraints implied by these shared reflectances, and then we describe how to combine them with an existing single-view method for color constancy to achieve our end goal.

3.1. Joint Constraints from Shared Reflectances

Let us assume that we are given a set of N images $\{I_i\}$ and have identified P uniformly-colored corresponding patches between these images. We let $y_{ip} \in \mathbb{R}^3_+$ be the observed RGB color of the pth patch in image I_i. Let $M_i \in \text{diag}(\mathbb{R}^3)$, referred to as the "illuminant," represent a diagonal (and invertible) linear transform that describes the mapping to the colors in I_i from their canonical counterparts, as in (3).

The observed colors $\{y_{ip}\}_i$ have the same reflectance and therefore the same canonical color, *i.e.*,

$$\forall i,\ y_{ip} = \rho_{ip} M_i \hat{x}_p,\ \|\hat{x}_p\|^2 = 1, \tag{4}$$

where $\hat{x}_p$ is a unit vector representing the canonical color of the patch, as in (2), (also referred to as the "reflectance" or "patch color"), and ρ_{ip} is a scalar brightness term that depends on both the albedo and the per-image shading effects. Equation (4) implies that, ideally, the canonical patch colors $\{M_i^{-1} y_{ip}\}_i$ are scaled versions of each other. We allow the scale factors ρ_{ip} to be arbitrary and express (4) in normalized form as

$$\frac{M_1^{-1} y_{1p}}{\|M_1^{-1} y_{1p}\|} = \frac{M_2^{-1} y_{2p}}{\|M_2^{-1} y_{2p}\|} \ldots = \frac{M_N^{-1} y_{Np}}{\|M_N^{-1} y_{Np}\|} = \hat{x}_p, \tag{5}$$

which is independent of $\{\rho_{ip}\}$. This relates the canonical patch colors in terms of their chromaticities, or the ratios of their different components.

When the illuminants are unknown, (5) provides a constraint on the possible illuminants among the images sharing corresponding color patches. Given approximate estimates of the illuminants $\{M_i\}$, statistics regarding the corresponding patch colors $\{M_i^{-1} y_{ip}\}$ can be used to infer $\hat{x}_p$; and similarly, given estimates of the patch color, $\hat{x}_p$, one can update the illuminants $\{M_i\}$ to values that better satisfy (5). Such an iterative algorithm could take many forms, and we have experimented with two. The method described here uses the constraints (5) to enhance the single-view optimization framework of Chakrabarti et al. [5]. The second, which we refer to as the "ratio method" makes similar use of this constraint. Viewed per color channel, c, working in the normalized space, this constraint for 2 images is: $\forall i_1 \neq i_2\ \frac{y^c_{i_1,p}}{m^c_{i_1}} = x_p = \frac{y^c_{i_2,p}}{m^c_{i_2}}$, which provides information on the ratio of illuminants between images. This method perform similarly and works with any initial single image color constancy method. It is described further in an associated technical report [17].

3.2. Estimation

We configure the estimation problem as one of optimizing a combined cost function that incorporates both $\{M_i\}$ and the patch colors $\hat{x}_p$. This cost function merges the joint reflectance constraints in (5) with single-view illuminant information extracted from each image. Since the constraints are expressed in terms of unit vectors (patch chromaticities), we express them by penalizing the angle between the illuminant-corrected image data $M_i^{-1} y_{ip}$ and the estimated patch color $\hat{x}_p$. For patch p and image i, we denote this angle as θ_{ip}.

Single-view illuminant information could be extracted using almost any existing method for single-view color constancy. We use the spatial correlations method of Chakrabarti et al. [5], which models the distribution of canonical colors in a typical scene using a collection of zero-mean Gaussian distributions of RGB colors in distinct spatial frequency sub-bands. The distribution parameters are fixed as in [5], and given these parameters and a single input image i, the illuminant is estimated through the eigen-decomposition of a 3×3 matrix that measures the difference between the observed RGB covariances in each sub-band and those expected under the model.

Formally, let $w_i \in \mathbb{R}^3$ be the diagonal elements of the corrective transform M_i^{-1} to be applied to I_i, with the additional constraint $\|w_i\|^2 = 1$. (We refer to w_i as the "illuminant" since it is just the inverse of m_i.) Following [5], we write the single-view cost of illuminant parameters w_i for image i as $w_i^T A_i w_i$, where the matrix $A_i \in \mathbb{R}^{3\times 3}$ is computed from the RGB distributions of the spatial frequency sub-bands of image i.

Combining this single-view information with our correspondence constraints, the joint cost is written

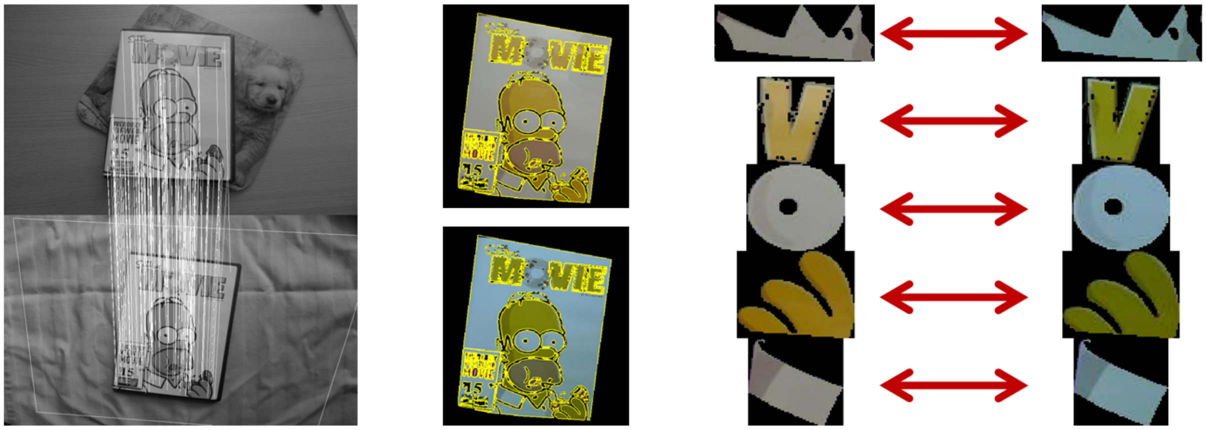
$$C(\{w_i\}, \{\hat{x}_p\}) = \underbrace{\frac{1}{N} \sum_i w_i^T A_i w_i +}_{\text{Single View Cost}} \quad \underbrace{\alpha \sum_{i,p} \sin^2(\theta_{ip})}_{\text{Reflectance Constraints}}\ , \tag{6}$$

with

$$\sin^2(\theta_{ip}) = 1 - \left(\hat{x}_p^T \frac{\text{diag}(w_i) y_{ip}}{\|\text{diag}(w_i) y_{ip}\|} \right)^2, \tag{7}$$

and a parameter α that controls the relative importance of the patch constraints and the single-view information. Since the matrices A_i are all positive semi-definite [5], the overall cost (6) is always positive.

The cost $C(\cdot)$ is hard to optimize due to the appearance of $\{w_i\}$ in the denominators of the second term. To circumvent this, we use the approximation $\|\text{diag}(w_i) y_{ip}\| \approx \|y_{ip}\|/\sqrt{3}$ (corresponding to $w_i \propto [1, 1, 1]$) in the denomi-



(a) Align images, remove background

(b) Detect stable regions

(c) Output dominant color of matched regions

Figure 2. Corresponding color regions are extracted as follows: 2(a) images are aligned and segmented using grayscale SIFT features; 2(b) Maximally Stable Extremal Regions (MSER) are detected; 2(c) the dominant color in each region is used to produce the final observed color correspondences.

nator of (7) to obtain

$$\sin^2(\theta_{ip}) \approx \frac{\|\mathrm{diag}(w_i)y_{ip}\|^2 - (\hat{x}_p^T\mathrm{diag}(w_i)y_{ip})^2}{\|y_{ip}\|^2/3}$$

$$= w_i^T(Y_{ip}Y_{ip}^T)w_i - w_i^T(Y_{ip}\hat{x}_p\hat{x}_p^TY_{ip}^T)w_i \triangleq z^2(\theta_{ip}), \quad (8)$$

with $Y_{ip} = \sqrt{3}\,\mathrm{diag}(y_{ip})/\|y_{ip}\|$. We find this to be a useful approximation in practice, and minimizing $z^2(\theta_{ip})$ typically also decreases the value of $\sin^2(\theta_{ip})$.

We now describe an iterative scheme to minimize this approximate cost with respect to both the illuminants $\{w_i\}$ and the true patch colors $\{\hat{x}_p\}$. We begin by setting w_i to the single-view estimates, *i.e.* the smallest eigenvectors of each A_i, respectively. Then, at each iteration, given the current estimates of w_i, we set $\hat{x}_p$ as

$$\begin{aligned}\hat{x}_p &= \arg\min_{\|\hat{x}_p\|^2=1} \sum_i z^2(\theta_{ip}) \\ &= \arg\max\ \hat{x}_p^T\left(\sum_i Y_{ip}^T w_i w_i^T Y_{ip}\right)\hat{x}_p, \quad (9)\end{aligned}$$

which is the *largest* eigen-vector of $\sum_i Y_{ip}^T w_i w_i^T Y_{ip}$. These estimates of $\hat{x}_p$ are in turn used to update each w_i as

$$w_i = \arg\min_{\|w_i\|^2=1} \frac{1}{N} w_i^T A_i w_i + \alpha \sum_p z^2(\theta_{ip})$$

$$= \arg\min w_i^T \underbrace{\left(A_i + N\alpha \sum_p Y_{ip}(I - \hat{x}_p\hat{x}_p^T)Y_{ip}^T\right)}_{A_i^+} w_i, \quad (10)$$

which is given by the *smallest* eigen-vector of A_i^+.

Since both (9) and (10) reduce the value of the approximate cost function which is bounded below by zero, the iterations are guaranteed to converge to a local minimum. Moreover, we evaluate the true cost in (6) after each iteration, and we terminate the iterations if the decrease of this true cost is sufficiently small.

As a final improvement, we use a weighted version of the cost function to diminish the effects of outliers and avoid converging to undesirable local minima when starting from poor initial estimates of w_i. We use $z'^2(\theta_{ip}) = k_{ip}z^2(\theta_{ip})$ with the weights k_{ip} given by

$$k_{ip} \propto \frac{\sum_{j\neq i}(M_i^{-1}y_{ip})^T(M_j^{-1}y_{jp})}{(M_i^{-1}y_{ip})^T(M_i^{-1}y_{ip})}, \qquad \sum_{i,p} k_{ip} = 1. \quad (11)$$

Note that the stopping criteria (based on the true cost, (6)) remains unweighted.

4. Establishing Color Correspondences

Our method relies on having some number of regions matched across several views of the same object, such that each set of image regions corresponds to the same physical surface patch on an object. Such regions can be found using any number of multi-view stereo techniques, and since the images used in this paper contain mostly planar objects like book covers, we use a straight-forward matching algorithm based on local features and homography-based geometric consistency. Extending the method to more general surfaces and textures is left to future work.

The process is illustrated in Fig. 2. First, we detect SIFT features in the grayscale images of the object, use RANSAC with a homography constraint to extract a set of geometrically-consistent matches (Fig. 2(a)), and use these to align the images. A conservative object mask is computed as the convex hull of the feature matches, and this is used to remove the background (and more) in each image. At this point, we could use the matched SIFT patches to provide us with color correspondences, however, we would have to account for the fact these patches typically fall on corners and blobs and thus have inhomogeneous color. A better choice for extracting uniformly-colored patches is the Maximally Stable Extremal Region detector [15] (Fig. 2(b)). Experimentally, we find that after removing very small regions this provides good coverage of the colors contained in the object. Nevertheless, we find that MSER patches can contain more than one distinct color, so in the final step (Fig. 2(c)), we select the dominant color by computing a histogram of the color values in each detected region, selecting the histogram bin with the largest value, and using the average of the color samples in that bin as the observed patch color $y_{i,p}$ in our method.

5. Results

In this section, the proposed method is evaluated using two datasets of real-world images, with patch correspondences being manually provided and automatically detected respectively. We gauge the benefits of joint multi-view estimation relative to single view estimation obtained by the original spatial correlations method [5], which serves as our baseline. We do this using different numbers of images and patches in correspondence, and in all cases, we use the angular error metric to measure the accuracy of estimated colors (of illuminants or patches):

$$\text{Angular Error}(x, x_g) = \cos^{-1}\left(\frac{x^T x_g}{\sqrt{(x^T x)(x_g^T x_g)}}\right), \tag{12}$$

where x and x_g are the estimated and "true" RGB vectors.

5.1. Color Checker Database

We first use 560 images from the database introduced in [11]. The provided RAW camera data was used to generate gamma-corrected RGB images using the program DCRAW[1], leaving out the camera's auto-white balance correction. The images were scaled down by a factor of five to remove potential demosaicking artifacts. Each image in this database contains a twenty-four patch color-checker chart at a known location. As illustrated in Fig. 3, we use the bottom row of gray patches to estimate the ground truth illuminant in each image. Of the remaining eighteen patches, six are used to train α, and the remaining are used for evaluation. The "true colors" of these patches are computed as per (9), using all 560 images and ground-truth values for w_i.

To evaluate the proposed method on this database, we use three-fold cross validation to train both the baseline spatial correlations method as well as to learn α. We use a simple grid-search to determine the optimal α using six patches (see Fig. 3) and sets of four images in correspondence. When reporting results for P patches in correspondence, we divide the remaining twelve patches into sets of P and average the computed error quantiles for each of these sets.

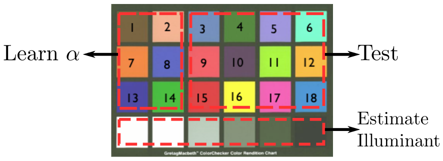


Figure 3. Every image in the Color Checker database contains this 24-patch color chart at a known location. We use the grey patches to estimate the scene illuminant, and the rest as known matched colors between images.

Figure 4 shows various error quantiles in the estimated image illuminants for different numbers of images and patches in correspondence. We note that joint illuminant estimation leads to significant gains in performance over single-view independent estimation (denoted by 0 corresponding patches in the figure). With just one corresponding patch and using pairs of images, the mean error drops from 3.9° to 3.5°, and to as low as 1.7° when using sets of sixteen images with twelve corresponding patches.

We next examine the variation in performance over the choice of patch that is assumed to be in correspondence. As noted in Sec. 2.1, using a diagonal transform to correct all colors is only an approximation and is likely to be less accurate for some patches in comparison to others. Accordingly, we define "angular spread" for each patch in the chart to be the mean angular error between the "true color" of the patch estimated across all images, and the corrected color in every individual image obtained using the diagonal transform corresponding to the ground truth illuminant. Figure 5 shows the relationship between the observed angular spread of a patch and estimation accuracy when using it during joint estimation (with sets of four images). Note that lower values of angular spread typically lead to better performance. However, even though some patches have angular spread greater than 6°, the mean illuminant error is lower than that of baseline single-view estimation in all cases.

Finally, we look at the effect that the number of images has on the accuracy of estimated patch colors (*i.e.* $\hat{x}_p$). Figure 6 shows the mean error in estimated patch colors when different numbers images are available as input for joint estimation (with only one patch in correspondence at a time). Performance improves rapidly with the number of images.

5.2. Real World Object Database

To evaluate our method on real-world objects, we collected a database of 39 images with five objects in differ-

[1] We used DCRAW version 8.77 with command dcraw -C 0.99967 0.99967 -o 0 -4 -q 3 -c RAWfilename > ppmfilename

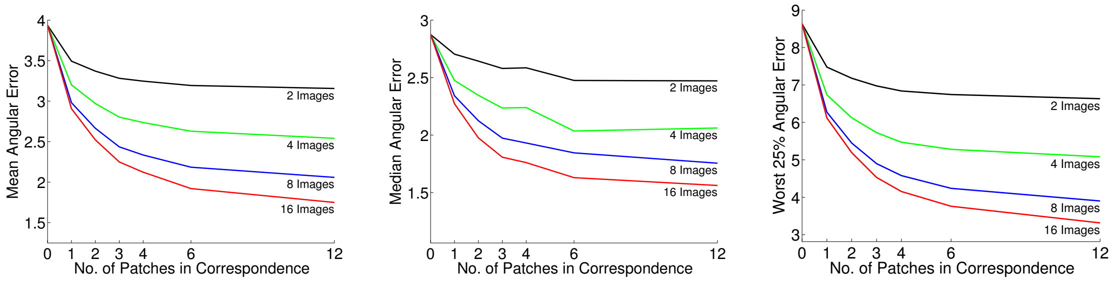


Figure 4. Performance, in terms of angular errors of estimated illuminants, for joint estimation with different numbers of matched patches and images. We report mean and median errors for each case, as well as the mean of the 25% largest error values to gauge robustness (note that axes have different scales). Values for zero matched patches correspond to the "single-view" (baseline) independent estimates.

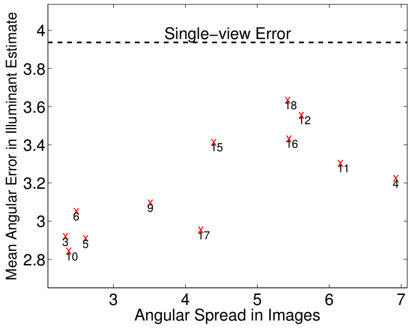


Figure 5. Mean illuminant errors when using different patches. The variation in estimation accuracy across patches is loosely correlated with the "angular spread" in the patch color. A high angular spread indicates that the diagonal approximation is poor for that patch. Performance is always better than the baseline single-view estimation, even for "poor patches".

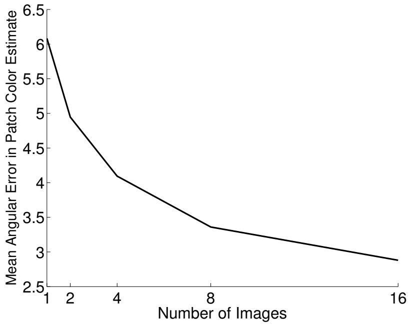


Figure 6. Accuracy of estimated patch colors $\hat{x}_p$ with different numbers of images used for joint estimation.

ent scenes (see Figure 9) under natural illuminants, both indoors and outdoors. Some illuminants are heavily colored and thus present a significant challenge for most single-view 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 color-checker dataset. The color-checkers are all in full view and are oriented towards the dominant source of light in the scene. We refer to this data set as DVDSBOOKS. In the following, we will refer to each set of images with an object in common as an *object set*. We automatically find patch correspondences between images with the same object in common as described in Section 4. This provides a more realistic setting under which the multi-view algorithm can be used for actual objects, compared to the color-checker chart patches as described in Sect. 5.1. Several of the 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 ten largest MSER regions. Decreasing the number of patches to ten puts an upper bound on the error we would reasonably expect from the method. The parameters of the single-view method and the α parameter of the multi-view method were trained on the color-checker database. We mask out the color-checker and object in each scene to obtain the single image illuminant estimate using local image evidence in the spatial correlations method, both the single-view and multi-view versions.

The errors in the estimated illuminants are summarized in Fig. 7. As before, we show the mean, median and mean of the worst 25% of the angular errors between the ground truth and the estimated illuminant. The mean error in true patch color estimate is shown in Fig. 8. As is the case for many single-view color constancy techniques, the single-view baseline method can perform poorly when the color statistics of the scene deviate significantly from its model [5]. One such case is shown in the top row of Fig. 9. However, even for these poor monocular illuminant estimates, we find that the multi-view method is able to leverage the additional information from multiple views to produce strikingly good results. Note that both the illuminant and canonical object color estimates have improved.

6. Conclusion

We have presented a method to learn canonical models of object color from multiple images of the object in different scenes and/or different illuminants. Traditional methods for color constancy can improve surface reflectance

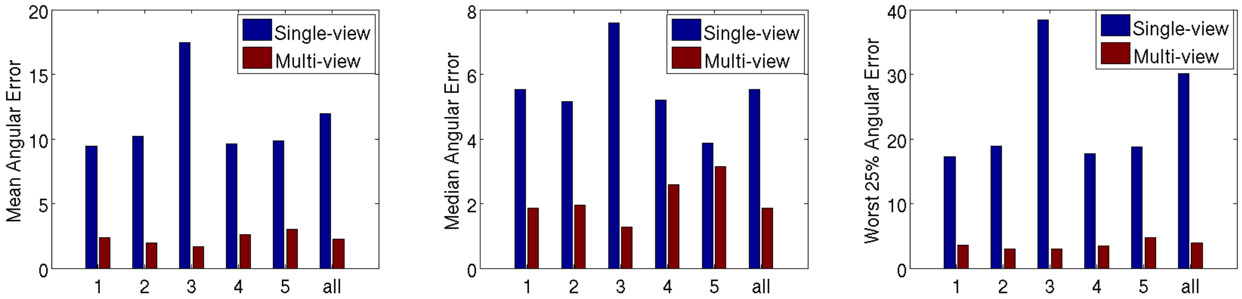


Figure 7. Results of illuminant estimation using the single-view and multi-view methods on the DVDSBOOKS dataset. We report mean and median errors for each case, as well as the mean of the 25% largest error values to gauge robustness (note that axis scales differ). The single-view method performs poorly on images that have color statistics that deviate significantly from the model that it assumes. However, using only ten automatically-identified patches and in seven images, the proposed multi-view approach is still able to leverage the correspondences to provide good illuminant estimates.

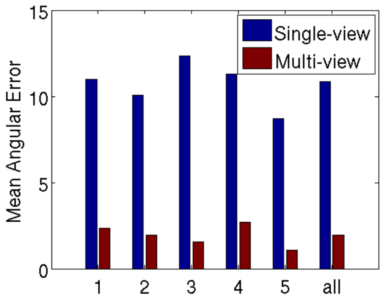


Figure 8. Results of estimating true patch colors $\hat{x}_p$ on the DVDSBOOKS dataset. We report mean angular errors for the single-view and multi-view methods.

estimates in uncalibrated images, but use only a single view and depend significantly on individual backgrounds. This greatly complicates the use of color in many recognition and retrieval applications. We defined a multi-view color constancy task, and presented techniques for aggregating color information across the several views. We develop an efficient multi-view method extending the monocular spatial correlations method; this method jointly optimizes estimates of corresponding patch colors and the illuminants present in multiple images. Correspondences can be formed using a number of alignment methods; we performed matching using local region features. 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 poor.

Acknowledgments

This project was funded by the NSF through grant IIS-0905243. Additional support for T.Z. and A.C. was provided by NSF grant IIS-0546408 and ARO contract/grant number 54262-CI. Various co-authors of this work have been supported in part by awards from the US DOD and DARPA, including contract W911NF-10-2-0059, by NSF awards IIS-0905647 and IIS-0819984, and by Toyota and Google.

References

[1] K. Barnard, G. Finlayson, and B. Funt. Color constancy for scenes with varying illumination. *Comp. Vision and Image Understanding*, 65(2):311–321, 1997.

[2] R. Boynton and C. Olson. Locating basic colors in the OSA space. *Color Research & Application*, 12(2):94–105, 1987.

[3] D. Brainard and W. Freeman. Bayesian color constancy. *JOSA A*, 14(7):1393–1411, 1997.

[4] G. Buchsbaum. A spatial processor model for object colour perception. *J. Franklin Inst.*, 310(1):1–26, 1980.

[5] A. Chakrabarti, K. Hirakawa, and T. Zickler. Color constancy beyond bags of pixels. In *Proc. CVPR*, 2008.

[6] H. Chong, S. Gortler, and T. Zickler. The von kries hypothesis and a basis for color constancy. In *Proc. ICCV*, 2007.

[7] G. Finlayson, M. Drew, and B. Funt. Diagonal transforms suffice for color constancy. In *Proc. CVPR*, 1993.

[8] G. Finlayson and S. Hordley. Color constancy at a pixel. *JOSA A*, 18(2):253–264, 2001.

[9] D. A. Forsyth. A novel algorithm for color constancy. *Int. J. Comp. Vision*, 5(1):5–35, 1990.

[10] B. Funt and G. Finlayson. Color constant color indexing. *IEEE Trans. PAMI*, 17(5):522–529, 1995.

[11] P. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp. Bayesian color constancy revisited. In *Proc. CVPR*, 2008.

[12] J. Geusebroek, R. van den Boomgaard, A. Smeulders, and H. Geerts. Color invariance. *IEEE Trans. PAMI*, pages 1338–1350, 2001.

[13] T. Gevers and A. Smeulders. Color based object recognition. In *Image Anal. and Proc.*, 1997.

[14] S. Hirose, K. Takemura, J. Takamatsu, T. Suenaga, R. Kawakami, and T. Ogasawara. Surface color estima-

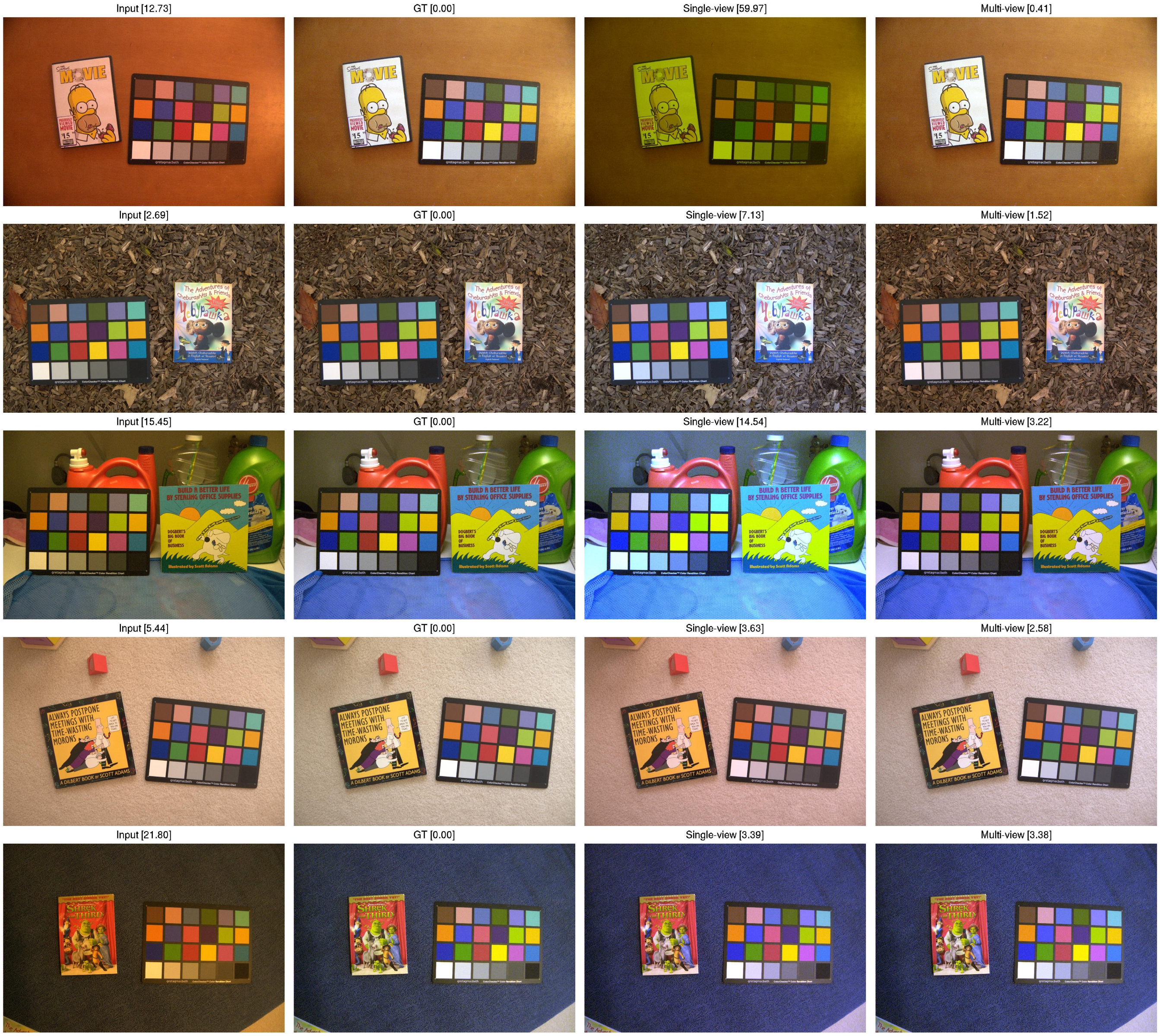


Figure 9. Example results from the DVDSBOOKS dataset. Each row shows the original image (Input), and the illuminant-corrected images $x = M^{-1}y$, using the ground truth (GT), the single-view spatial correlations method and the multiview spatial correlations method (using all available images of that object in the dataset). The number in brackets is the angular error of the illuminant estimate w.r.t. the ground truth. The top row shows an example where the single image estimate of the illuminant is very poor, due to the uniform background, yet the multiple image estimate performs remarkably well.

tion based on inter-and intra-pixel relationships in outdoor scenes. In *Proc. CVPR*, 2010.

[15] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, 2004.

[16] M. Olkkonen, T. Hansen, and K. Gegenfurtner. Color appearance of familiar objects: Effects of object shape, texture, and illumination changes. *J. of Vision*, 8(5), 2008.

[17] T. Owens, K. Saenko, A. Chakrabarti, Y. Xiong, T. Zickler, and T. Darrell. The ratio method for multi-view color constancy. Technical Report UCB/EECS-2011-23, EECS Department, University of California, Berkeley, Apr 2011.

[18] C. Rosenberg, T. Minka, and A. Ladsariya. Bayesian color constancy with non-gaussian models. *Proc. NIPS*, 2003.

[19] A. Troccoli and P. Allen. Building illumination coherent 3D models of large-scale outdoor scenes. *Int. J. Comp. Vision*, 78(2-3):261–280, 2008.

[20] J. Weijer, T. Gevers, and A. Gijsenij. Edge-based colour constancy. *IEEE Trans. Image Proc.*, 16:2207–2214, 2007.