Experiments in Sensor Sharpening for Color Constancy

Kobus Barnard and Brian Funt
Simon Fraser University
Burnaby, British Columbia

Abstract

Sensor sharpening has been proposed as a method for improving color constancy algorithms but it has not been tested in the context of real color constancy algorithms. In this paper we test sensor sharpening as a method for improving color constancy algorithms in the case of three different cameras, the human cone sensitivity estimates, and the XYZ response curves. We find that when the sensors are already relatively sharp, sensor sharpening does not offer much improvement and can have a detrimental effect. However, when the sensors are less sharp, sharpening can have a substantive positive effect. The degree of improvement is heavily dependent on the particular color constancy algorithm. Thus we conclude that using sensor sharpening for improving color constancy can offer a significant benefit, but its use needs to be evaluated with respect to both the sensors and the algorithm.

Introduction

Sensor sharpening has been proposed as a method for improving color constancy algorithms [1], but it has not been tested in the context of real color constancy algorithms. Rather, the experimental results available are limited to the minimum error possible with and without sharpening. Since the error in current color constancy methods is often substantially larger than the minimum error possible, we felt it necessary to investigate further the utility of sensor sharpening for color constancy. In this paper we provide results of color constancy with and without sharpening for a Sony DXC-930 CCD video camera, a Kodak DCS-200 digital camera [2], a Kodak DCS-420 digital camera [2], the XYZ response curves [3], and the Vos and Walraven human cone sensitivity estimates [3]. The general conclusion is that when the sensors are already relatively sharp (e.g. the Sony camera), further sensor sharpening is not worth the trouble, and often has a small detrimental effect. However, when the sensors are not sharp (e.g. the DCS-200 and DCS-420), sensor sharpening can have a substantial positive effect, depending on the algorithm, thus validating the original work.

Sensor Sharpening

We begin with an explanation of sensor sharpening [1]. The motivation for sensor sharpening is the observation that most color constancy algorithms make use of a diagonal model of illumination change. To understand this model, consider a white patch under two different illuminants. Suppose that under the first illuminant the color is [r,g,b] and under the second illuminant the color is [r', g', b']. It is possible to map the color of white under the first illuminant to the color under the second by post-multiplication by a diagonal matrix: [r', g', b'] = [r, g, b] diag(r'/r, g'/g, b'/b). If the same diagonal matrix transforms the RGB of all surfaces (not just the white ones) to a good approximation, then we say that we have a diagonal model of illumination change. It turns out that the accuracy of the approximation is a function of the vision system's sensors.

The idea of sensor sharpening is to map the data by a linear transform T into a new space where the diagonal model holds more faithfully. Colour constancy algorithms which rely on the diagonal model can then proceed more effectively. The final result is then mapped back to the original RGB space with the inverse transformation. Working in the transform space is like having new sensors which are a linear transformations of the old ones. Further, the sensitivity functions of sensors which support the diagonal model tend to look sharper with narrower peaks than ones that do not—in the extreme case, if the sensors are delta-functions, the diagonal model holds exactly. From these two observations, we get the name: sensor sharpening.

The main technical result in sensor sharpening is finding the transformation T. In [1], Finlayson et al propose three methods for finding T: "sensor based sharpening", "database sharpening", and "perfect sharpening". For this work we chose database sharpening over sensor based sharpening due to the clean correspondence between the sharpening method and a color constancy error metric. Perfect sharpening did not work well for us because our test illuminant set did not meet the key requirement of being two-dimensional, partly due to the inclusion of fluorescent lights.

In database sharpening, RGB are generated using a database of reflectance spectra, together with an illuminant spectrum and the sensors. This is done for two separate illuminants. Let A be the matrix of RGB for the first illuminant and B be the matrix for the second, with the RGB's placed row-wise. In the sharpening paradigm we map from B to A with a sharpening transform, followed by a diagonal map, followed by the inverse transform. If we express each transform by post multiplication by a matrix we get: $A \approx BTDT^{-1}$ In database sharpening the matrix T (and implicitly D) is found that minimizes the RMS error, $|A \approx BTDT^{-1}|_2$. The sharpening transform gives exactly the same error as the best linear transform M. In fact, T is found by diagonalizing M, where M minimizes $||A - BM||_2$.

One implementation issue should be noted. The result of the diagonalization is ambiguous up to scaling and swapping of the columns of T. As is standard, we use columns of norm 1. Furthermore, we put the element of T of largest absolute value on the diagonal by swapping columns, and ensure that it is positive by multiplying the column by -1 if necessary. Then in a similar way we attempt to make the other diagonal elements as large as possible. This procedure is used to reduce the number of negative components of sharpened data.

In this work we view color constancy as finding a transformation from the image of a scene taken under an unknown illuminant, to the image of the same scene as though it were taken under a known, “canonical”, illuminant [4]. A priori, the nature of the transformation is open, but most algorithms find a diagonal transform, and it is these algorithms which interest us here. Of course, the best linear transformation will give at most the same error as any diagonal transformation, but it should be clear from the above that the generalized diagonal transform TDT^{-1} gives us a chance of having this lower error with a diagonal model [5]. Thus we should be able to improve diagonal color constancy if the right sharpening transform is available.

For the first part of our work we estimate upper bounds for improvement with sharpening by giving the algorithms the appropriate T. Here T is computed assuming that we can correctly “guess” the illuminant spectrum. It should be noted that such an upper bound depends on the assumption that sharpening works with, or at least not against, the specific algorithm. When the sensors are already sharp, it is more likely that the breakdown of this assumption becomes noticeable because a smaller portion of the error is due to the lack of sharpness. Thus with sharp sensors, using this “optimal” sharpening often gives worse results than not using sharpening at all.

For the real color constancy problem, we do not know the illuminant spectra used above to compute T. We have experimented with first running the algorithms in non-sharp space to estimate the illuminant, and then using this estimate to help choose T. However, we have not yet found much benefit of this strategy over simply using the average of all the illuminants in the database as the sharpening illuminant. Specifically, our work suggests that when sharpening is beneficial for real color constancy algorithms, all reasonably chosen T work similarly. However, in the case of the comparison “algorithm” which uses the actual illuminant (unknown to the real algorithms) the only source of error is the diagonal model approximation, and the optimal sharpening is significantly better.

Colour Constancy Algorithms

We will now discuss briefly the color constancy algorithms investigated here. The first algorithm is based on a gray world assumption. It assumes that the average RGB of a scene under a given illuminant is that of “gray”, where gray is defined as the average of the entire reflectance database. The implied diagonal map is the ratio of the known RGB of gray under the canonical illuminant, to the RGB of gray under the test illuminant, estimated by the average of the scene RGB. This algorithm performs very well in our experimental context, as the assumption is statistically valid here, but it should be noted that this algorithm does not do nearly as well on real images because in this context the average reflectance is not generally known [6].

The second algorithm is based on the Retinex model of human vision [7, 8, 9]. The result is computed using the maximum in each channel as an estimate of the color of white under the test illuminant. Similar to the gray world algorithm, the implied diagonal map is the ratio of the known color of white under the canonical illuminant to the estimate.

The gamut mapping approaches introduced by Forsyth [4] directly estimate the diagonal map from a set of possible maps. The possible maps can be constrained by considering scene RGB [4], and insisting on plausible illuminants [10]. Given a set of possible RGB, a solution needs to be chosen from this set. One method for doing so is maximizing the volume of the convex hull of the mapped RGB [4]. A second method is to take the centroid of the convex hull of the maps [11, 12].

Each color constancy algorithm relates to sharpening transforms differently. For example, Forsyth’s CRULE algorithm [4] and Finlayson’s extended version for chromaticity [10] rely heavily on the diagonal model, and are likely candidates for improvement by the use of sharpening. At the other extreme, gray world algorithms are not much affected by sharpening. Finally, for the variant of Retinex used here, the relationship of the algorithm design to sharpening is unclear, as the maximum in each channel intuitively estimates white in RGB space, but its choice as an estimator in sharp space is less clear. Similar considerations are also relevant in analyzing methods for choosing a solution from the constraints sets found with gamut mapping algorithms.

Additional problems can occur with the gamut mapping algorithms due to negative RGB which can be introduced by sharpening. Depending on the variant of the algorithm, it may or may not be problematic to have negative components in the sharpened input data, canonical data, or illuminant database data. We found that the chromaticity version of Forsyth’s method [10] is very sensitive to these problems, and therefore we do not include results for it. In the case of the RGB variants used below, a few of the illuminants cause trouble; in the context of an application, one could default back to non-sharp computation. However, for the purposes of testing we simply exclude that generated scene from the test, thus ensuring that all algorithms are run on the same data.

Results

We have investigated sensor sharpening in the case of a Sony DXC-930 CCD video camera, a Kodak DCS-200 digital camera [2], a Kodak DCS-420 digital camera [2], the

XYZ response curves [3], and the Vos-Walraven human cone sensitivity estimates [3]. As our main interest here is machine vision and image reproduction, we are most interested in the results for the cameras. However, sharpening is also of interest in relation to human vision[1].

A sampling of our results is shown in Table 1. The data was used was generated from a test set of 100 measured illuminants (normalized to be the same magnitude), a database of roughly 2000 reflectances, and the 5 different sensors. The 100 measured illuminants include a variety of indoor and sources and outdoor illumination, as well as a complex combinations thereof obtained at random locations in and around our university campus. We note that the results of sharpening are a function of the spectra databases; our data sets were chosen to be as general as possible.

As discussed above, the result of each algorithm can be interpreted as supplying a mapping from unknown illuminant images to canonical ones. To obtain the errors presented here, we applied that mapping to a large set of RGB computed using the entire reflectance database, together with the test illuminant and the sensors. We then computed a similar set using the canonical illuminant, which was chosen to be a Tungsten illuminant, and tabulated the RMS difference of the two sets. This error metric was chosen to coincide with the error which database sharpening strives to reduce. Lack of space prevents us from providing errors using other metrics.

Each entry in the table is the average of 600 results. The first five rows in the two tables are target results. The first row is the best linear fit (which is also the best diagonal fit with optimal database sharpening), the second is the best diagonal fit, the third is the result of “knowing” the RGB of the illuminant, the forth is the same with optimal database sharpening, and the fifth is the same with sharpening using the global average illuminant. Following the target results are the results of several algorithms estimating the best diagonal map on the basis of 8 randomly chosen data points.

The “sharpness” of the sensors can be taken as the relative magnitudes of the best linear fit and the best diagonal fit. The results below indicate that sharpening the already “sharp” DXC-930 sensors is troublesome at best. However, in the case of the less sharp DCS-200 and DCS-420 sensors, sharpening can give a substantial improvement, as is the case with the E-CRULE-HA algorithm. For example, with the DCS-200, average illuminant based sharpening reduces the error of E-CRULE-HA from 81.3 to 33.0. E-CRULE-HA is better tuned than E-CRULE-MV for both the data set and the error metric, and thus the difference between the two algorithms in the case of sharp sensors is not surprising. However, it is interesting to note that in the case of the DCS-200 and DCS-420 sensors, sharpening was required to obtain the advantages of this algorithm in this situation. In summary, the results indicate that the benefit of sharpening is quite dependent on the algorithm, which is understandable based on the discussion above, but the magnitude of the effect is still surprisingly large.

In Table 1 we exclude results for the cone sensitivity estimates. Most algorithms did not work well in the sharp space for these sensors. Only the comparison algorithms and the gray world algorithm gave reasonable results. With these sensors, the sharpened RGB typically had negative R and G. This almost always made the gamut mapping approach untenable, and also made the Retinex algorithm perform very poorly.

Finally, we note that when sharpening is beneficial for the real color constancy algorithms, using the global average illumination for sharpening is not much worse than the optimal. On the other hand, in the case of the known illuminant “algorithm”, optimal sharpening gives substantially better results (relative to the already small error) than average illuminant sharpening. This suggests that as color constancy improves, the method of choosing the sharpening transform becomes more relevant. However, when we ran some of the experiments again with the number of surfaces increased to 32 in order to improve color constancy, the advantage of using the optimal sharpening was still slight, even though the RMS error for some of the algorithms was less than 20.

Table 1. Mapping Error vs. Algorithm

		DXC 930	DCS 200	DCS 420	XYZ
Linear Fit		2.88	1.32	1.67	2.04
Diagonal Fit		4.42	8.48	10.7	9.79
Known Illum.		4.53	8.91	11.2	10.6
Known Illum.	(opt)	2.99	1.33	1.68	2.07
Known Illum.	(ave)	3.90	1.63	2.02	3.00
Gray World		44.5	39.8	39.0	40.3
Gray World	(opt)	45.8	41.4	40.5	42.9
Gray World	(ave)	45.5	41.3	40.4	43.4
Retinex		146	112	120	123
Retinex	(op)	144	102	111	114
Retinex	(ave)	143	102	111	113
E-CRULE-MV		74.3	69.8	66.0	52.4
E-CRULE-MV	(opt)	73.2	58.1	60.9	57.7
E-CRULE-MV	(ave)	72.0	57.3	60.4	55.6
E-CRULE-HA		35.8	81.3	80.5	39.2
E-CRULE-HA	(opt)	36.4	33.0	33.7	33.2
E-CRULE-HA	(ave)	37.6	33.4	34.2	35.5

Table 1: RMS mapping error between the full data set under the unknown illuminant and the canonical illuminant. The (opt) results are for database sharpening using the unknown illuminant. The (ave) results are for database sharpening using the global average illuminant spectra. No designation in the second column is used for no sharpening. The E-CRULE algorithms are based on Forsyth’s CRULE algorithm [4], but are extended with Finlayson’s illumination constraint [10]. The (MV) variant uses the original “maximum volume” method to chose the final answer from the constrained set, whereas the (HA) variant uses the average. Note that the experimental conditions are optimal for the gray world, since the average of the test data base is known and used. Similarly, the E-CRULE-HA results are somewhat optimistic, although not to the same degree as the gray world results. The maximum variability of the numbers between successive runs is estimated to be 5%.

Conclusion

Sensor sharpening can have a substantive positive effect on color constancy processing, but this is highly dependent on the both sensors and the algorithm used. Furthermore, some sensor/algorithm combinations are either unusable, or yield very poor results, due to the introduction of negative data values into the data from sharpening. Thus using sharpening for improving color constancy algorithms requires care.

On a more positive note, our results suggest that in the context of real color constancy algorithms, the method of sharpening is not very critical. We expected that using the somewhat ad hoc sharpening method based on the average illuminant would be distinctly worse than the optimal method based on the actual illuminant (not normally available). However, we did not find much difference between the two. Nonetheless, the potential advantage of the optimal method is clear from the results of the comparison algorithm based on the actual illuminant. Thus we feel that further research into the choice of sharpening method is warranted. Finally, in some cases, the dependence on the algorithm greatly exceeded our expectations, and additional work towards understanding the relationship between sensor sharpening and color constancy algorithms is also needed.

Acknowledgments

We are grateful for the support of Hewlett-Packard Corporation and the Natural Sciences and Engineering Council of Canada.

References

1. G. D. Finlayson and M. S. Drew and B. V. Funt, "Spectral Sharpening: Sensor Transformations for Improved Color Constancy," *Journal of the Optical Society of America A*, **11**(5), 1553-1563 (1994).
2. P. L. Vora, J. E. Farrell, J. D. Tietz, D. H. Brainard, "Digital color cameras—2—Spectral response," from http://color.psych.ucsb.edu/hyperspectral/
3. G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulas*, 2nd edition, (Wiley, New York, 1982).
4. D. Forsyth, "A novel algorithm for color constancy," *International Journal of Computer Vision*, **5**, 5-36, (1990).
5. G. D. Finlayson and M. S. Drew and B. V. Funt, "Color Constancy: Generalized Diagonal Transforms Suffice," *Journal of the Optical Society of America A*, **11**(11), 3011-3020 (1994).
6. Brian V. Funt, and Kobus Barnard, "Is Colour Constancy Good Enough," 5th European Conference on Computer Vision, 1998.
7. John J. McCann, Suzanne P. McKee, and Thomas H. Taylor, "Quantitative Studies in Retinex Theory," *Vision Research*, **16**, 445–458 (1976).
8. E. H. Land, "The Retinex theory of Color Vision," *Scientific American*, **237**, 108-129 (1977).
9. B. K. P. Horn, "Determining lightness from an image," *Computer Vision, Graphics, and Image Processing*, **3**, 277-299 (1974).
10. G. D. Finlayson, "Color in perspective," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **18**(10), 1034-1038 (1996).
11. Kobus Barnard, "Computational color constancy: taking theory into practice," MSc thesis, Simon Fraser University, School of Computing (1995).
12. Graham Finlayson and Steven Hordley, "A theory of selection for gamut mapping color constancy," Proceedings IEEE Conference on Computer Vision and Pattern Recognition, 1998.