Recovering Color from Black and White Photographs

Sven Olsen
University of Victoria

Rachel Gold
University of Victoria

Amy Gooch
University of Victoria

Bruce Gooch
University of Victoria

Abstract

This paper presents a mathematical framework for recovering color information from multiple photographic sources. Such sources could include either black and white negatives or photographic plates. This paper's main technical contribution is the use of Bayesian analysis to calculate the most likely color at any sample point, along with an expected error value. We explore the limits of our approach using hyperspectral datasets, and show that in some cases, it may be possible to recover the bulk of the color information in an image from as few as two black and white sources.

1. Introduction

The modern tradition of photography started in the 1820s. Before Kodak's introduction of Kodachrome color reversal film in 1935, color image reproduction was a laboratory curiosity, explored by a few pioneers such as Prokudin-Gorskii [1]. Tens of thousands of black and white plates, prints, and negatives exist from the hundred year time period between the birth of modern photography and the widespread introduction of color film [15]. Currently, it is only possible to create color images from early black and white data sources in those rare occasions when the original photographer contrived to create a color image by taking matched exposures using red, green, and blue filters.

In this paper, we explore how variations in spectral sensitivity between different black and white sources may be used to infer color information to within a known margin of error. We assemble a toolkit that scholars can use to recover color information from black and white photographs in historical archives.

The paper's main technical contribution is the use of Bayesian analysis to calculate the most likely color at any sample point, along with an expected error value. This analysis requires some knowledge of the types of light spectra likely to occur in the scene.

We believe that, for many antique photographs, the introduction of accurate color information will greatly enhance their visual quality. Color recovery may prove invaluable in understanding archived materials of historical, cultural, artistic, or scientific significance.

2. Related Work

Both the computer graphics and photography communities have long investigated techniques for colorizing black and white photographs. Previous research in the computer graphics community proposed adding color to black and white images by either transferring color from a similar color image to a black and white image [13, 17] or used color-by-example methods that required the user to assign color to regions [7, 14].

Colorizing old black and white photographs can be as simple as painting colors over the original grey values. While difficult to do by hand, with the help of the right software tool, a convincingly colorized version of a black and white photograph can be created provided that a user selects reasonable colors for each object. Software capable of quickly colorizing a greyscale image given minimal user interaction has been a topic of recent research interest, pioneered by Levin et al. [7]. Nie et al. [9] presented a more computationally efficient formulation. Qu et al. [12] presented a colorization system specialized for the case of colorizing manga drawings, while Lischinski et al. [8] generalized the idea to a broader category of interactive image editing tools. Such colorization techniques naturally complement the process of color recovery. Provided an appropriately chosen sparse set of recovered colors, colorization can be used to create complete color images from any of the black and white originals (see also Figure 2).

In the field of remote sensing, it is sometimes necessary to construct color images from multispectral image data. For example, color images of the Martian landscape were generated from the 6 channel multispectral image data returned by the Viking lander [10, 11]. The basis projection algo-
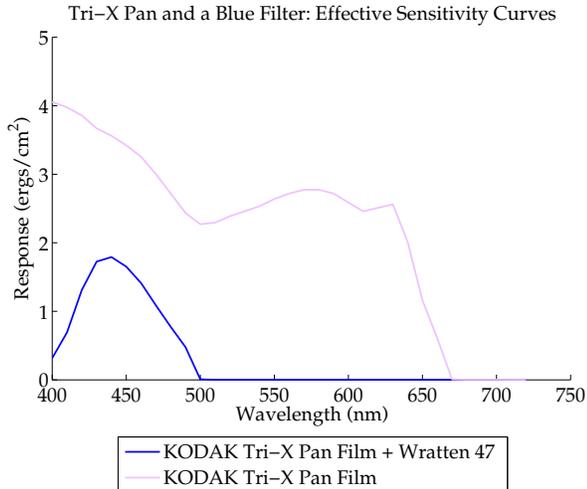
Figure 1: *Our mathematical framework allows the recovery of color with known error thresholds from as few as two images. This example uses a hyperspectral dataset to create two simulated photographs using KODAK Tri-X Pan film with and without a blue filter (Wratten 47). The first image shows the effective sensitivity curves used for creating the two simulated images used as source data. The remaining three images show the default recovered colors, the recovered colors with statistical error correction, and ground truth color values for the dataset.*

rithm used to create those color images is nearly identical to the projection technique that appears in our own methods section (see Equation (4)). However, in our own framework, this initial result can be significantly improved using natural light statistics.

A technique developed by the photographic community combines several registered black and white images taken in rapid succession with red, green, and blue filters to create a color image. This method dates back to at least the early 1900s when Sergey Prokudin-Gorskii set out to document the Russian Empire using a pioneering camera and red, blue, and green filters to capture three different images of the same scene. The slides were then projected through red, green and blue filters of a device know as a "magic lantern" which superimposes the images onto a screen, producing a color projection [1]. While the projection equipment that Prokudin-Gorskii used to display his photographs has been lost, it is possible to extract color information from the surviving glass plates. Recovering color from Prokudin-Gorskii's photographs is relatively straightforward, because we can assume that the red, blue, and green channels in an output color image are just scaled versions of the red, blue, and green filtered photographs. In order to create color images, scholars working with the Library of Congress adjusted the weightings of each color channel until the resulting image looked sensible. In this work we provide a solution to the general problem of recovering color from an arbitrary collection of black and white photographs.

3. Methods

3.1. Terminology and Background

3.1.1   Terminology

In this paper, we use the convention of referencing per-photograph variables with a i subscript. To distinguish between the different dimensions of a color space, we use a k subscript.

Given a spectral sensitivity curve, $G(\lambda)$, we define its *ideal response* to incoming light having radiance distribution $V(\lambda)$ to be,

$$\int G(\lambda)V(\lambda)d\lambda. \tag{1}$$

Under certain assumptions, equation (1) can be used to model photoreceptors with spectral sensitivity $G(\lambda)$. In order for equation (1) to predict the response of a photoreceptor, it must be the case that the incoming light does not change as a function of space or time. The activation level of a photoreceptor is influenced by the amount of time that it is exposed to light, thus, for Equation (1) to predict a photoreceptor's measurements, we must assume that the receptor will always be exposed for a fixed length of time, and that the sensitivity of the photoreceptor does not change over time. These assumptions are reasonable in the case of
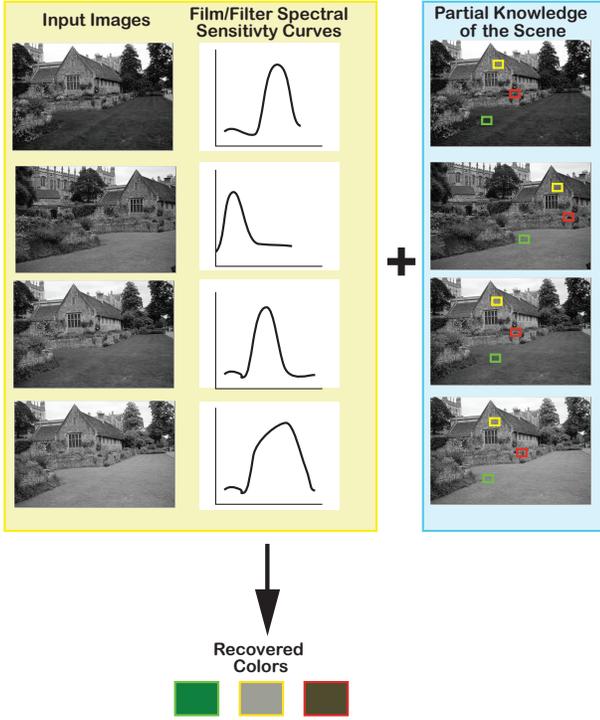
Figure 2: *Application to non-aligned images. Our methods can operate on data samples taken from non-aligned images. Color recovery at point samples is possible, provided that a user selects pixel sets that share a common material and lighting properties, and the spectral sensitivity curves of the film and transmission functions of any lens filters are known. Such point color samples are a natural input for standard colorization techniques.*

a CCD camera with fixed exposure time, but they are overly restrictive as a model for film photography.

The CIE Colorimetric system defines the RGB and XYZ colorspaces in terms of ideal response equations [18]. Specifically, if we assume that the light $V(\lambda)$ passing through a small patch of our image is constant over time, and $C_k(\lambda)$ is the sensitivity curve associated with the k-th dimension of the CIE RGB color space, then the k-th color value of that patch should be,

$$\int C_k(\lambda)V(\lambda)d\lambda. \qquad (2)$$

In a black and white photograph, an integral similar to equation (1) plays a role in determining brightness at each point. The film's spectral sensitivity and the transmittance functions of any applied filters will combine to form an *effective spectral sensitivity*, which we denote as F_i. The relation of the ideal response of the film to the density of the developed

negative is determined by the parameters of the developing process; see Section 3.3 for further discussion.

3.1.2 The Linear Algebra of Response Functions

Putting aside, for a moment, the question of whether or not it is possible to find our films' ideal response values, we now consider whether given such responses, it is possible to approximate the color values at any point in the image.

First, note that the set of all response functions forms an *inner product space*, with the inner product defined as,

$$F \cdot G := \int F(\lambda)G(\lambda)d\lambda.$$

Define $\mathbf{R}(V)$ to be a linear transformation that maps radiance distributions V to the ideal responses of a set of film sensitivity functions $\{F_i\}$,

$$\mathbf{R}(V)_i := F_i \cdot V.$$

It follows that any color response function C_k can be uniquely decomposed into two components,

$$C_k = S_k + D_k. \qquad (3)$$

Here S_k is an element of the span of $\{F_i\}$, and D_k is an element of the null space of $\mathbf{R}$ [2]. As $S_k \in span(\{F_i\})$, we can find a set of scalar weights w_i such that,

$$S_k = \sum_i w_i F_i.$$

Therefore, for any radiance distribution V, the scalar product $S_k \cdot V$ can be recovered from the set of measured film response values $\mathbf{R}(V)$, as,

$$S_k = \sum_i w_i F_i \quad \rightarrow \quad S_k \cdot V = \sum_i \mathbf{R}(V)_i w_i. \qquad (4)$$

Equation (4) can be used to infer colors from a set of response measurements $\mathbf{R}(V)$, if we assume that $C_k \cdot V \approx S_k \cdot V$, i.e., if we assume that little color error will be introduced if C_k is replaced by its projection into the spanning space of $\{F_i\}$. Color interference using projection into a spanning space is not a novel technique, and non-trivial examples of the method can be found dating back to the Viking lander project of the 1970's [10, 11].

3.2. Bayesian Error Prediction

Color recovery using a spanning space projection has the potential to be very accurate, provided that a sufficiently

wide range of response functions is available [10, 11]. However, given a limited set of effective film responses, the approach will frequently fail. In the following, we introduce a technique for improving on the results of the spanning space projection method. We use a large collection of example light sources to predict the color information likely to have been lost as a result of the spanning space projection. That expected value can then be used to adjust the initial recovered colors. Additionally, we define a variance term that can be used to predict the accuracy of the adjusted color values.

It follows from Equation (3) that the correct color value will be the sum of the responses of the spanning space and null space components of the color sensitivity curve,

$$C_k \cdot V = S_k \cdot V + D_k \cdot V. \tag{5}$$

Given a limited set of film response curves, it is likely that $D_k \cdot V \not\approx 0$. We suggest that this error can be corrected by finding correspondences between the response of the null space component and the set of known film response values, $\mathbf{R}(V)$. Formally, we propose adopting the approximation,

$$C_k \cdot V \approx S_k \cdot V + \mathrm{E}(D_k \cdot V \mid \mathbf{R}(V)\,).$$

In order to calculate *a posteriori* probabilities for $D_k \cdot V$ given the measured film response values $\mathbf{R}(V)$, we must assume some knowledge of the light spectra likely to be present in the scene. In our approach, this prior knowledge takes the form of a set of known light sources $\{V_q(\lambda)\}$. We make the assumption that that the true radiance distribution, $V(\lambda)$, will be equal to a scaled version of one of the known light sources. Thus, good error predictions require a large database of light sources.

Given a list of film responses $\mathbf{R}(V)$, we begin by finding the scale factor s_q that minimizes the apparent difference between any example light V_q and the correct light V. We accomplish this by solving the following simple optimization problem,

$$\delta_q^2(s) := \sum_i (R_i \cdot sV_q - \mathbf{R}(V)_i)^2,$$

$$s_q := \min_{s \in [s_0, s_1]} \delta_q^2(s).$$

We define an error value associated with the q-th example light source as $\delta_q^2(s_q)$. This error definition is quite similar to the affinity function definition used in many image segmentation algorithms [16], and the pixel weighting function used in Levin et al.'s colorization algorithm [7]. The best approach to modeling error value probabilities would be to develop a mixture model based on known data; however, as is often the case in similar image segmentation tasks, the

computational costs associated with such models are substantial. A more practical approach is to assume a Gaussian distribution of error. Thus, we define a probability function, $P(V_q \mid \mathbf{R}(V)\,)$, as follows,

$$P(V_q \mid \mathbf{R}(V)\,) \propto u_q := \exp\left(\frac{-\delta_q^2(s_q)}{2\sigma^2}\right).$$

Here, we define σ^2 as the m-th smallest $\delta_q^2(s_q)$ calculated for our set of known lights.

Given the above, we can now form complete definitions for the expected error in our recovered colors, as well as the variance in those error terms. Defining $d_{qk} := s_q V_q \cdot D_k$, we have,

$$\mathrm{E}(D_k \cdot V \mid \mathbf{R}(V)\,) = \sum_q P(V_q \mid \mathbf{R}(V)\,) d_{qk},$$

$$= \frac{\sum_q u_q d_{qk}}{\sum_q u_q}.$$

$$\mathrm{Var}(D_k \cdot V \mid \mathbf{R}(V)\,) = \frac{\sum_q u_q d_{qk}^2}{\sum_q u_q} - \left(\frac{\sum_q u_q d_{qk}}{\sum_q u_q}\right)^2.$$

In Section 4 we use hyperspectral datasets to test the accuracy of these error predictions, given a range of different possible film response curves and example light sets of varying quality. Along with the set of known lights, V_q, the accuracy of our predictions is also influenced by the choice of values for the minimum and maximum allowed scale factors, s_0 and s_1, as well as the index m used to determine σ^2. In our experiments, we found that $[s_0, s_1] = [.8, 1.2]$ and $m = 5$ typically produced good results.

3.3. Calculating Ideal Film Response Values

Unfortunately, the realities of film photography deviate sufficiently from equation (1) that we cannot consider the grey values in our black and white photographs to be the ideal response values. However, we can hope to find a function $g_i : \mathbb{R} \to \mathbb{R}$ that will map the grey values in our image, b, to $\mathbf{R}(V)$.

In order to find such a function we must propose a model of black and white photography that makes it possible to approximate ideal response values given the grey values present in the image, along with some other additional information that we can hope to uncover. One such model is,

$$b = f_i(\Delta t_i F_i \cdot V). \tag{6}$$

Here Δt_i is the exposure time of the photograph, and $f_i : \mathbb{R} \to \mathbb{R}$ is a nonlinear scaling function, determined by the
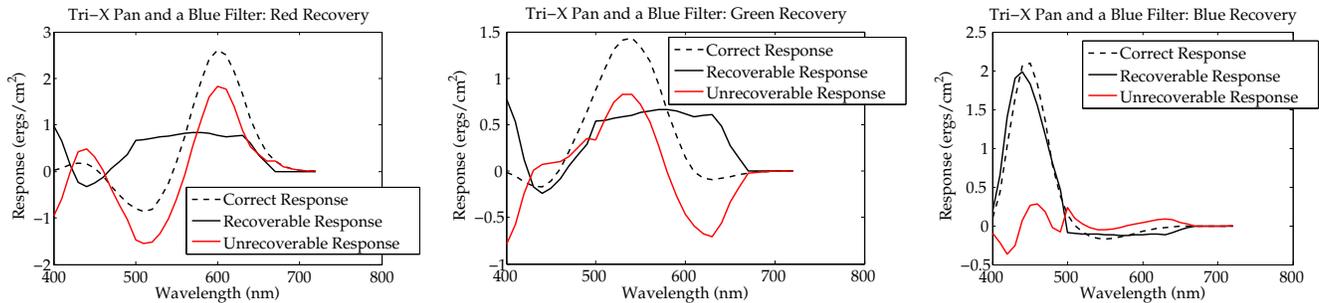
Figure 3: *Here we show the null space and spanning space projections defined in Equation* (3), *for the case of the the filter set used in Figure* 1. *The spanning space projection,* S_k, *represents the portion of the response curve that can be easily recovered from the information present in the photographs, while the the null space projection,* D_k, *represents parts of the spectrum that are impossible recover without the use of Bayesian error prediction.*
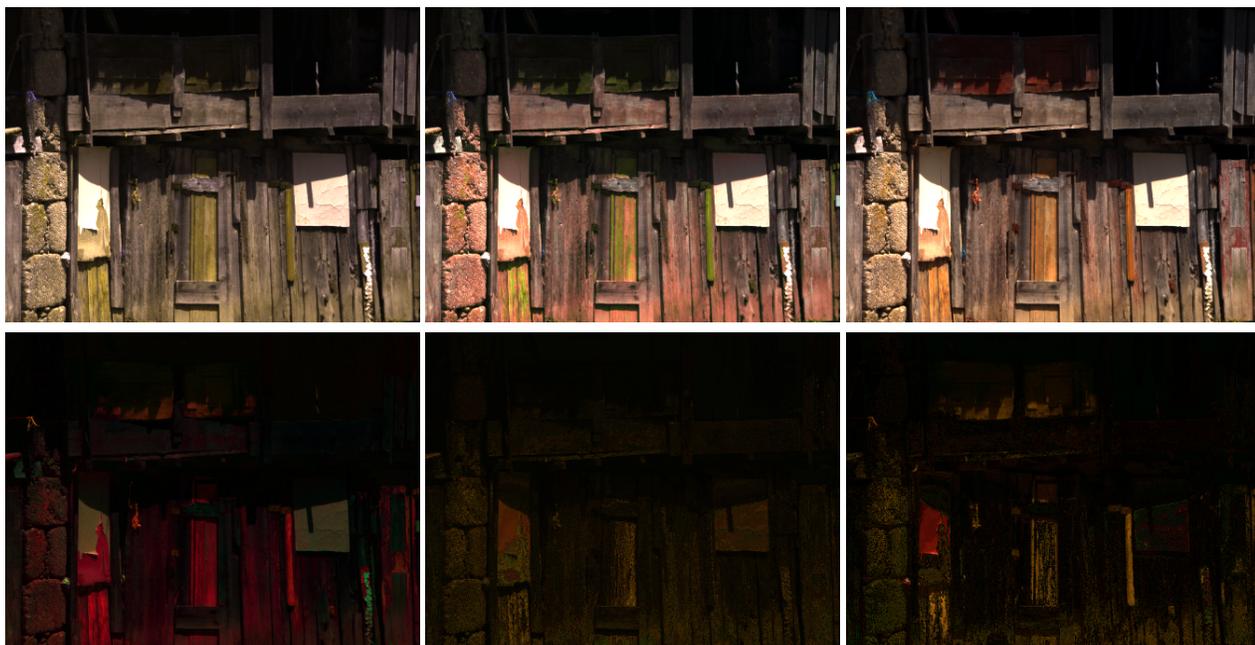

Figure 4: *Hyperspectral Test Results: Here we show the color recovered images from the two simulated photographs of the Tri-X Pan and Blue Filter test case. Row one: initial recovered colors, recovered colors with Bayesian error correction, correct color values. Row two: Initial color error, predicted error, actual error in the final recovered colors. Twenty additional results are shown in the supplemental materials.*

saturation function of the film, and the details of the development process used to create the photographic print [5].

To find f_i, we can either assume that the film saturation function given by the manufacturer will be a good match for f_i, or we can take several photographs of the same scene using known exposure times, develop the film in a way that conforms to our assumptions about the film development process used in our input black and white image, and fit a curve to the resulting data points.

Provided that f_i is known, it is possible to calculate Δt_i,

given at least one point in the image with known incoming light $V(\lambda)$. Such an exposure time inference is similar to techniques used by Land [6]. For example, given a number of photographs of a building taken under daylight, if we can identify a point in the image that contains a diffuse object painted with a white paint having known reflectance function $H(\lambda)$, and assume a standard daylight illuminant $L(\lambda)$, then we can predict that $V(\lambda) = H(\lambda)L(\lambda)$.

4. Evaluation using Hyperspectral Images

The mathematics of color recovery impose few bounds on the amount of error that could occur due to the unrecoverable portion of color sensitivity curve, $D_k(\lambda)$. While we can attempt to model this error statistically, it is ultimately the light present in the scene, taken together with the span of the films' effective response curves, that determines the extent to which color recovery is possible.

In order to access the practical limits of our color recovery process, we have investigated the effectiveness of our algorithm in a series of controlled test cases. We use Foster et al.'s hyperspectral datasets [4] to simulate the result of taking several photographs of the same scene using different film/filter combinations. Each hyperspectral dataset is effectively an "image" that stores thirty or more values per pixel. Each value in the dataset records the narrow band response from a 10nm section of the visible spectrum. By applying a discretized spectral sensitivity curve to this data, we are able to calculate both ideal color values and ideal film response values.

These test cases allow us to analyze the color error that results from unavoidable sources of color information loss, i.e., the unrecoverable $D_k(\lambda)$ curve, independently of any measurement errors that occur in the course of constructing ideal response values from an input photograph set. Thus, the hyperspectral test results can be considered an "upper bound" on the quality of the color information that can be recovered using real photographs. However, it must be noted that Foster et al.'s hyperspectral datasets do not record information in the IR and UV spectrums of light, though some black and white films are sensitive in those wavelengths. Thus, our simulations miss errors that might be introduced due to electromagnetic radiation in those spectrums.

For each hyperspectral scene, we construct a set of ideal film response values $\mathbf{R}(V)$, assuming a collection of photographs taken with varying combinations of films and filters. The effective response curves used are defined using the published spectral sensitivity curves of Kodak and Fujifilm black and white films, along with the published transmittance values of Kodak Wratten filters. The film/filter combinations which we test range from the virtually impossible case of recovering color from a single photograph, to the relatively easy case of recovering color given 5 different photographs, taken with a range of photographic filters.

Foster et al.'s hyperspectral datasets include a range of scenes featuring natural and man-made objects, viewed under daylight illuminant conditions. In order to perform the statistical error correction of Section 3.2, we use radiance values taken from two hyperspectral datasets, Scenes 2 (green foliage) and 7 (sunlit building) from Foster et al. [4]. As the high resolution data can greatly increase the computational cost of finding the expected error at each point, we downsample in image space by a factor of 9. To avoid the case of impossibly good example lights, neither of the datasets used to provide example light sets are used to generate test problems. Our tests are thus run using responses derived from Scenes 3 (English garden), 4 (flower), 6 (city scape) and 8 (barn door). Example recovered images from these trials are shown in Figures 1 and 4, and the complete set of recovered images and associated error values are included in our supplementary materials.

Overall, the results of the hyperspectral tests are quite promising. They suggest that roughly correct color recovery may be possible given as few as two photographs. Predictably, a collection of photographs representing a wider range of spectral sensitivity curves greatly improves the quality of the recovered colors.

4.1. Error Prediction Accuracy

The error prediction step can only be expected to provide accurate results insomuch as the set of example radiance distributions includes rough matches for the radiance distributions that exist inside a given scene. We would expect few similarities between the reflectance functions of metals and those of plants, and thus, using a set of example light sources drawn entirely from images of plants is likely to lead to incorrect results in the case of photographs containing metal objects. A similar failure case is shown in Figure 5.

Interestingly, even rough matches between the content of the example light set and the source photograph often appear sufficient to allow the Bayesian error correction to significantly improve the accuracy of the recovered colors. For example, the materials present in the barn door dataset are only roughly similar to those in the sunlit building dataset. Yet, as show in Figure 4, the error correction is accurate enough to frequently introduce the reds and oranges missing from the initial recovered colors. (That said, there are also several visible failure cases, which attest to the limits of the approach.)

For many of the test cases, the variance term, $\mathrm{Var}(V \cdot D_k \mid \mathbf{R}(V))$, is demonstrated to be a fairly accurate prediction of the true error in the recovered colors. The accuracy of the error predicted by the variance term appears to be a function of the dataset that the tests are run on, rather than the particular filters used. This suggests that the quality of the error predictions is determined primarily by the degree to which the example dataset can provide good matches for the light in the scene.

Figure 5: *Error Correction Failure Case: Here we show another set of hyperspectral test results, once again using the same Tri-X Pan and Blue Filter test case, and the same simulated ideal response functions as Figure 4. However, for this example, we have reduced the set of example light sources, using only hyperspectral data from Foster et al.'s "green foliage" dataset. From left to right: initial recovered colors, recovered colors with Bayesian error correction, correct color values. As these results demonstrate, given sufficiently poor example light sources, the error correction step can increase color errors.*

5. Discussion and Conclusions

Viewed at a high level, the problem of finding ideal response values provides us with an opportunity to incorporate any information that we have on the content of the image into the color recovery process. This information does not need to be as specific as the known response functions and lighting conditions proposed in the previous subsection. For example, if we know that some of the pixels in the images are skin illuminated by daylight, that restricts the possible observed radiance, $V(\lambda)$, and thus the possible response values $\mathbf{R}(V)$ for those points in the image. Developing algorithms that can effectively make use of such partial information has the potential to greatly improve the effectiveness of color recovery algorithms, and such techniques could also be useful in other computational photography applications. There is also considerable potential to incorporate existing computational photography techniques into the color recovery process, for example, it may be possible to combine existing color constancy algorithms with our present work in order to recover colors from photographs of the same scene taken under differing illumination conditions [3].

Prior to the 1920's, most photography made use of orthochromatic films, which had negligible red responses. This would seem to put a hard limit on the time period before which color recovery could be possible – as response information for the red and orange portions of the spectrum will always be missing in photographs taken before the introduction of panchromatic films. One of the most interesting results of our research is that, as the hyperspectral test cases suggest, it may still be possible to recover accurate color information from such photographs, by making use of Bayesian error prediction.

For the vast majority of antique black and white photographs, neither exposure times, nor any filter information have been recorded. Moreover, professional black and white photographers regularly use a range of darkroom techniques that would render the inference of ideal response values for points in their prints exceedingly difficult. Film negatives or photographic plates are more attractive data sources, as all the unknowns associated with printmaking are eliminated. Furthermore, photographers make use of numerous conventions that recommend exposure times, filter use, and development processes in particular situations. Thus, there is potential to develop tools for guessing the likely photographic variables associated with a given collection of antique negatives.

References

[1] The Empire that Was Russia: The Prokudin-Gorskii Photographic Record Recreated. http://www.loc.gov/exhibits/empire/making.html. 1, 2

[2] S. J. Axler. *Linear algebra done right*. Springer-Verlag, second edition, 1997. 3

[3] D. A. Forsyth. A novel algorithm for color constancy. *Int. J. Comput. Vision*, 5(1):5–36, 1990. 7

[4] D. Foster, S. Nascimento, and K. Amano. Information limits on neural identification of coloured surfaces in natural scenes. *Visual Neuroscience*, 21:331–336, 2004. 6

[5] J. Geigel and F. Kenton Musgrave. A model for simulating the photographic development process on digital images. *Computer Graphics*, 31(Annual Conference Series):135–142, 1997. 5

[6] E. Land. The retinex theory of color constancy. *Scientific Am.*, pages 108–129, 1977. 5

[7] A. Levin, D. Lischinski, and Y. Weiss. Colorization using optimization. In *Proceedings of ACM SIGGRAPH 2004*, Computer Graphics Proceedings, Annual Conference Series, Aug. 2004. 1, 4

[8] D. Lischinski, Z. Farbman, M. Uyttendaele, and R. Szeliski. Interactive local adjustment of tonal values. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Papers*, pages 646–653, New York, NY, USA, 2006. ACM. 1

[9] D. Nie, Q. Ma, L. Ma, and S. Xiao. Optimization based grayscale image colorization. *Pattern Recogn. Lett.*, 28(12):1445–1451, 2007. 1

[10] S. K. Park and F. O. Huck. A spectral reflectance estiamtion technique using multispectral data from the viking lander camera. Technical Report D-8292, NASA Tech. Note, 1976. 1, 3, 4

[11] S. K. Park and F. O. Huck. Estimation of spectral reflectance curves from multispectral image data. *Appl. Opt.*, 16:3107–3114, dec 1977. 1, 3, 4

[12] Y. Qu, T.-T. Wong, and P.-A. Heng. Manga colorization. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Papers*, pages 1214–1220, New York, NY, USA, 2006. ACM. 1

[13] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley. Color transfer between images. *IEEE Computer Graphics & Applications*, 21(5):34–41, Sept. 2001. 1

[14] D. Sýkora, J. Buriánek, and J. Žára. Unsupervised colorization of black-and-white cartoons. In *NPAR '04: Proceedings of the 3rd international symposium on Non-photorealistic animation and rendering*, pages 121–127, New York, NY, USA, 2004. ACM. 1

[15] P. Vanderbilt. Guide to the special collections of prints & photographs in the library of congress. U.S. Library of Congress, 1955. 1

[16] Y. Weiss. Segmentation using eigenvectors: A unifying view. *iccv*, 02:975, 1999. 4

[17] T. Welsh, M. Ashikhmin, and K. Mueller. Transferring color to greyscale images. *ACM Transactions on Graphics*, 21(3):277–280, July 2002. 1

[18] G. Wyszecki and W. S. Stiles, editors. *Color Science : Concepts and Methods, Quantitative Data and Formulae*. Wiley-Interscience, 2000. 3