Reconstructing Images from Trichromatic Samples: From Basic Research to Practical Applications

David H. Brainard* and Doron Sherman†

*Department of Psychology, UC Santa Barbara, Santa Barbara, California
†HP Laboratories, Imaging Technology Department, Palo Alto, California

Introduction

The human retina encodes information about images through the responses of three classes of photoreceptor, often referred to as the L, M, and S cones. These photoreceptors are arranged in three interleaved mosaics; at any one retinal location only a single cone type samples the retinal image. To create our percept of a continuous colored world, the visual system must reconstruct the responses of the missing two cone types at each retinal location. The algorithm that performs this reconstruction works very well—we rarely perceive artifacts that arise from the interleaved sampling arrangement.

Most CCD based color cameras employ the same interleaved sampling architecture as the human retina. Yet for CCD cameras, color artifacts are quite common near sharp luminance boundaries. CCD cameras are more susceptible to these artifacts because their reconstruction algorithms are not as successful as the one employed by the human visual system.

This talk will begin by reviewing basic research designed to elucidate reconstruction by the human visual system. We will then show how ideas that emerged from the basic research have led to a new algorithm for processing images acquired with CCD cameras.

Sampling and Reconstruction by the Human Visual System

The initial encoding of light by the human visual system is accomplished by specialized nerve cells called photoreceptors. Each photoreceptor produces an electrochemical response when light is imaged on it by the optics of the eye. The response of an individual photoreceptor carries information about the intensity of the image at a single location. Information about the image as a whole is carried by the ensemble of responses from the entire photoreceptor mosaic. We may think of the retina as an image sensing device that discretely samples the image.

Spatial sampling eliminates information about the image. The information loss is illustrated in Figures 1 and 2. Figure 1 shows a regular one-dimensional sampling mosaic. Sampling by such mosaics has been extensively analyzed in the engineering literature. The figure also shows the intensity profiles of two one-dimensional sinusoidal images. Note that the images have the same intensity at each location where there is a photoreceptor; they differ only in the gaps between photoreceptors. For this reason, the images cannot be distinguished by the responses of the photoreceptor mosaic; the crucial information has been lost though the sampling process.

When physically different images generate the same ensemble of responses from a mosaic of photoreceptors, we say that the images are aliases of one another. An ensemble of photoreceptor responses defines an equivalence class of images. Members of the equivalence class have the property described above: they have same intensity at each location where there is a receptor. A visual system cannot determine with certainty which member of an equivalence class of aliases was actually present.

![Figure 1. The figure plots the intensity profile of two one-dimensional sinusoids with spatial frequencies of 2 c/°image (dashed) and 6 c/°image (solid). The gratings are sampled by a one-dimensional mosaic of 8 sensors. The sensor positions are indicated by the cones at the top of the figure. The Nyquist limit for this mosaic is 4 c/°image. The two gratings have the same intensities at each sensor location. They cannot be distinguished by any reconstruction scheme and are said to be aliases of one another. Standard low-pass filtering will reconstruct the 6 c/°image grating as its 2 c/°image alias.](image)

In the classic frequency domain analysis of regular spatial sampling, the ambiguity introduced by sampling is resolved by adding prior information that the physical image is band-limited. The fundamental result for regular one-dimensional monochromatic sampling is that a signal may be reconstructed from samples if it contains no frequency components greater than the Nyquist limit of the sampling mosaic. The Nyquist limit is defined as half the sampling rate. Signal components at frequencies above the Nyquist limit cannot be distinguished from lower frequency components, as illustrated in Figure 1. When standard methods are used, high frequency components are reconstructed at lower frequencies. This phenomenon is referred to as aliasing. The one-dimensional result has been extended to handle two-dimensional sampling both for regular (see e.g. Pratt) and irregular\footnote{Pratt} sampling mosaics.

Although the photoreceptors sample the image discretely, our conscious percept is of a spatially continuous image. Apparently, our visual system uses the information available in the ensemble of photoreceptor responses to reconstruct a continuous representation of the image. It is relatively easy to demonstrate the action of this reconstruction process. There is a substantial area of the retina where there are no photoreceptors at all. This area is referred to as the optic disk; it is where nerve fibers pass through the retina to connect with the primary visual cortex. Figure 2 emphasizes the formal similarity between the situation at the optic disk and the regular spatial sampling illustrated in Figure 1.

In spite of the fact that there are no photoreceptors in the optic disk, we do not perceive a hole in our visual field. Rather, our visual system "fills-in" a percept for us, based on the information provided by photoreceptors at other image locations. To experience this filling-in, the reader may proceed as follows: a) close the left eye, b) with the right eye, fixate on some object straight ahead, c) hold up the right thumb at arms length, d) slowly move the thumb horizontally to the right while carefully maintaining fixation straight ahead. When the thumb has moved about seven inches on its horizontal traverse to the right, the image of the thumbnail will fall on the optic disk. At this point, the thumbnail will disappear and be replaced by a percept similar to the image surrounding the optic disk. The disappearance is not due simply to the fact that visual resolution decreases with eccentricity: if the thumb is moved further to the right it will reappear clearly enough. Apparently, whatever algorithm the visual system uses to reconstruct the image does not create for us the percept of a thumbnail in the absence of any direct information that the thumbnail is there. Generally, however, the reconstruction algorithm works quite well, since we are rarely aware of the blind spot corresponding to the optic disk. The detailed nature of the algorithm that reconstructs the image near the blind-spot is not currently well-understood. See Walls for an interesting discussion.\footnote{Walls}

The filling-in that occurs at the optic disk is easy to demonstrate because the disk is relatively large. Formally, however, there is no difference between reconstructing an image across the large inter-receptor spacings of the optic disk and across the much smaller spacings between neighboring photoreceptors in the rest of the retina. By using very fine patterns, it is possible to demonstrate that reconstruction occurs at all retinal locations. Williams\footnote{Williams} constructed a specialized laser interferometer capable of imaging high spatial frequency gratings on the retina. Even in the fovea, where the inter-receptor spacing is very small, he found that high-spatial frequency gratings were misperceived as much lower spatial frequency patterns.

The patterns perceived in Williams' and other\footnote{Other} experiments are consistent with the notion that the visual system uses low-pass filtering to reconstruct a representation of the image and most formal models of the reconstruction process are based on this or closely related notions.\footnote{This} Low-pass reconstruction is a reasonable strategy for the visual system to employ for monochromatic reconstruction at the fovea. This is because outside the laboratory blurring by the eye's optics constrains retinal spatial frequencies to be at or below the Nyquist limit for the foveal photoreceptor mosaic.

\section*{The Trichromatic Reconstruction Problem}

Our discussion so far has focused on spatial sampling and reconstruction without any discussion of color. As mentioned in the introduction, the cone photoreceptors that support human color vision actually come in three types, arranged in three interleaved mosaics. Color CCD cameras share this same basic design. In this sec-
tion, we discuss the trichromatic reconstruction problem. To start the transition from human vision to engineering application, we introduce this problem in the context of color CCD cameras.

The optics of a CCD camera produce an image which is spatially sampled by the sensing elements of the CCD. Each individual element functions in a manner analogous to a single photoreceptor. We typically specify color images by a triplet of red, green, and blue (RGB) values at each pixel on a rectangular sampling grid. This representation is not available directly from the output of most color CCD cameras. Rather, these cameras provide only single sensor response at each pixel. Color information comes because the camera as a whole contains multiple classes of sensing elements, with each class characterized by a distinct spectral sensitivity. Usually there are three classes, referred to as red (R), green (G), and blue (B). Figure 3 provides a schematic illustration of this design.

To obtain a full color image, we must reconstruct the responses of all three sensor classes. We call this trichromatic reconstruction. Trichromatic reconstruction generalizes the classic problem of reconstructing signals from samples.

![Figure 3. Color CCD camera. Left: The CCD camera contains a rectangular mosaic of sensing elements. Each element has either an R, G, or B spectral sensitivity. Right: We may think of the overall sensor mosaic as three interleaved submosaics. Each submosaic corresponds to one of the three sensor classes. Note that although the overall geometry of the mosaic is rectangular, this is not the case for the individual submosaics.](image1)

![Figure 4. Chromatic aliasing. The upper and lower panels plot the R and G components of two sinusoids. One sinusoid is an intensity grating with spatial frequency 3 c/image (solid). The R and G components for this sinusoid are in phase. The second sinusoid is a red/green grating with spatial frequency 1 c/image (dashed). The R and G components for this sinusoid are out of phase. The sinusoids are sampled by two interleaved submosaics. The R sensor positions are indicated by the shaded cones above the top panel. The G sensor positions are indicated by the open cones above the bottom panel. The two sinusoids produce the same responses in all of the sensors and cannot be distinguished by any reconstruction scheme. Low-pass filtering applied separately to each submosaic will reconstruct the 3 c/image intensity grating as its 1 c/image red/green alias. Although the total number of sensors is the same as it was in Figure 2, distortion from aliasing occurs at lower spatial frequencies.](image2)
The most straightforward approach to reconstructing full color images is to treat each sensor class separately. Standard methods may then be applied to each submosaic in turn. The disadvantage of this approach is that aliasing will occur when the image contains spatial frequencies above the Nyquist limits of the individual submosaics, as illustrated in Figure 4. This sort of chromatic aliasing can produce objectionable artifacts in images recorded with color CCD cameras, particularly near intensity edges (see Figure 5). Such artifacts are readily observed in many commercially available camcorders.

Figure 5. Chromatic aliasing at edges. The upper and lower panels plot the R and G components of two signals. One signal is an intensity step edge (solid). The R and G components of this edge are in phase with one another. The second signal is a low frequency alias of the intensity step edge (dashed). The sinusoids are sampled by two interleaved submosaics. The R sensor positions are indicated by the shaded cones above the top panel. The G sensor positions are indicated by the open cones above the bottom panel. The two signals produce the same responses in all of the sensors. Careful examination of the low frequency alias reveals that its R and G components are not in phase at locations adjacent to the edge. This alias will contain visually salient colored fringes.

Because the cut-off frequency of the human optical system is well-matched to the Nyquist frequency of the overall retinal sampling mosaic, the human visual system should be subject to the same sort of chromatic fringing exhibited by CCD cameras. Yet such fringing is not part of our typical perceptual experience. We can view fine spatial patterns without perceiving objectionable artifacts from the interleaved design of our trichromatic retinal mosaic.

Williams et al. studied in detail the appearance of intensity gratings at a spatial frequency of 20 cycles per degree. This spatial frequencies is well-below the Nyquist limit for the foveal cone mosaic as a whole but is close to the Nyquist frequency for the L and M cone submosaics. What Williams et al. observed is that most observers perceived low contrast red and green splotches superimposed on the perceived intensity grating. They measured the apparent contrast of the splotches by asking observers to match it to that of an adjustable red-green grating of low spatial frequency. Williams et al. developed a simple model that explained these splotches as an example of chromatic aliasing. They assumed sampling by a trichromatic retinal mosaic and a reconstruction algorithm that treated the signals from each of the three submosaics separately. The output of this model was a trichromatic image that contained exactly the sort of splotches reported by observers. The contrast of these splotches, however, was much higher than that reported by observers. Indeed, the splotches dominate the model output (see the color plates in Williams et al.'s paper), whereas for the human observer they are quite subtle.

One interesting possibility that could explain Williams et al.'s result is that the visual system combines information from all three cone types rather than processing each type separately. To understand this idea, consider Figure 6. Each numbered column of the matrix corresponds to a spatial location. Each row corresponds to a cone type. A trichromatic image can be completely specified (at the sampling resolution of the matrix) by specifying an intensity value for each of the matrix cells, that is by specifying the intensity seen by each class of cone at each spatial location. A dot in the figure indicates locations where a cone photoreceptor is present. Note that at most one dot occurs in each row, illustrating that there is at most one cone at any location. The trichromatic reconstruction problem is to use the information available from the cells where cones are present to estimate the image intensity in all the cells.

As noted above, the natural generalization of monochromatic reconstruction methods is to restrict the reconstruction procedure to process each cone class separately. Under this general strategy, the missing intensity values within each row would be estimated using only photoreceptor responses from the same row. Such estimation could be accomplished using any standard monochromatic reconstruction method. Adopting this strategy, however, prevents the use of information carried by, say, the L cones about the intensities that would be seen by, say, the missing M cones.

In general, the information carried by the response in one cell of the matrix about the intensity value of another cell depends on the correlation between the two: the higher the correlation, the more useful the response.
Separate submosaic methods take advantage of correlations between neighboring locations within a single cone class. If there are correlations across cone classes at a single location, this information is ignored.

The few measurements that exist about the correlation of signals in different color bands suggest that these correlations are very high for natural images. This suggests that the performance of separate submosaic methods can be improved by combining information across photoreceptor classes. Brainard and Williams\textsuperscript{14} conducted experiments on how the human visual system reconstructs signals from S cones. They were able to demonstrate conditions where the spatial structure of the percept corresponding to S cones is influenced by signals from the L and M cones. Their result suggests strongly that human vision employs a reconstruction strategy more sophisticated than separate submosaic reconstruction. Rather, it seems to take advantage of correlations that exist across the responses of different cone classes.

Based on their result and the earlier work of Williams et al., we became interested in developing a normative computational model for reconstruction that could integrate information across cone classes. We outline our approach in the next section.

### Statistical Decision Theory Approach

Our approach to trichromatic reconstruction is to apply Bayesian decision theory. The Bayesian approach provides a general prescription for how to use all of the information in a data set (e.g. the ensemble of photoreceptor responses) to estimate the values of a set of unknown parameters (e.g. the image intensities in each color band at all locations). Prior information about the parameters is expressed as a probability distribution. If we are trying to estimate parameters described by the vector $\mathbf{x}$, then the prior information is given by the probability density $p(\mathbf{x})$. The relation between the parameters $\mathbf{x}$ and the data $\mathbf{y}$ is also expressed as a probability density $p(\mathbf{y} | \mathbf{x})$, often referred to as the likelihood. The likelihood is essentially a forward model of the data acquisition device and is generally readily specified. Given the prior $p(\mathbf{x})$ and the likelihood $p(\mathbf{y} | \mathbf{x})$, the probability of any set of the parameter values, given the data, is computed using Bayes rule

$$p(\mathbf{x} | \mathbf{y}) = \frac{C}{p(\mathbf{y} | \mathbf{x})} p(\mathbf{x})$$

(1)

where $C$ indicates a normalization constant that depends on the data $\mathbf{y}$ but not on the parameters $\mathbf{x}$. The distribution $p(\mathbf{x} | \mathbf{y})$ is referred to as the posterior. The posterior gives the probability that the parameters $\mathbf{x}$ generated the data $\mathbf{y}$. The posterior expresses what is known about the parameters given the prior and the data.

To go from the posterior to an actual parameter estimate we need to specify a loss function $L(\tilde{\mathbf{x}}; \mathbf{x})$. This function specifies the penalty for choosing $\tilde{\mathbf{x}}$ when the actual parameters are $\mathbf{x}$. Given the posterior and a loss function, we may compute the expected loss corresponding to any estimate $\tilde{\mathbf{x}}$. This is called the Bayes risk and is given as

$$R(\tilde{\mathbf{x}} | \mathbf{y}) = \int_{\mathbf{x}} L(\tilde{\mathbf{x}}; \mathbf{x}) p(\mathbf{y} | \mathbf{x}) d\mathbf{x}$$

(2)

The estimate $\tilde{\mathbf{x}}$ is chosen to minimize the Bayes risk. See Berger\textsuperscript{15} for a general discussion of Bayesian methods.

Bayesian estimation provides a principled way to choose an optimal estimate. As a practical matter, a number of difficulties can obstruct its use. First, it may be difficult to specify a prior distribution that adequately captures what is known about the structure of the parameters. In the case of trichromatic reconstruction, the prior must specify how likely it is that any given image will occur. Second, it may be difficult to specify a loss function that captures how costly errors of various types are. Freeman and Brainard\textsuperscript{16,17} consider loss functions in detail. Finally, it may be computationally difficult to minimize the Bayes risk once the prior, the likelihood, and the loss function are specified. For the case of trichromatic reconstruction, however, it has been possible to make progress using the Bayesian approach.

Brainard\textsuperscript{18} provides a formal description our trichromatic reconstruction algorithm. Space limitations require that here we only provide a qualitative description. To develop a prior distribution for images, we wanted to incorporate two simple facts about natural images. First, the average power spectrum of natural images falls off rapidly as a function of spatial frequency.\textsuperscript{19,20} Second, the signals in different color bands are positively correlated.\textsuperscript{20} These two facts can be described probabilistically by assuming that images are drawn at random from a multivariate Normal distribution with known mean and covariance. This class of prior does not capture all of the regularity in natural images. Random draws from such priors appear like military camouflage: they contain no edge-like structures. Compared to the simple a-
assumption that the image is band-limited, however, Normal priors are quite expressive.

It is straightforward to specify the likelihood for the sampling problem. Once the spectral sensitivities and locations of the photoreceptors are known, it is possible to compute the expected response of each photoreceptor to any given image. To convert this deterministic computation into a probability distribution, we assume that the photoreceptor responses are subject to independent Normally distributed random noise. Thus the likelihood is also characterized by a multivariate Normal distribution with known mean and covariance.

Given Normal prior and likelihood, it may be shown that the posterior is also Normal. Normal distributions have the attractive feature that they are unimodal. Thus the Bayes risk for most reasonable loss functions is minimized by the posterior mean. For our problem, an analytic expression for this mean exists and it is a linear function of the photoreceptor responses. Thus the Bayesian approach leads to a linear reconstruction algorithm that combines information from all three photoreceptor classes. The reconstruction algorithm is not, however, shift invariant and is not equivalent to filtering in the frequency domain.

We have performed a number of calculations using the Bayesian approach to reconstruction from samples. Brainard and Williams simulated the performance of the method for the human photoreceptor mosaic. In contrast to separate submosaic reconstruction methods, the Bayesian algorithm is able to predict the low apparent contrast of the red and green splotches observed by Williams et al. Interestingly, it also predicts that luminance splotches should be perceived when high contrast chromatic gratings are viewed at spatial frequencies of about 30 cycles per degree. Sekiguchi, Williams, and Brainard have observed hints of such splotches in an experiment whose primary purpose was to measure sensitivity to chromatic gratings.

Second, Brainard simulated the performance of the algorithm for a regular one dimensional dichromatic mosaic. These simulations indicate that the type of chromatic fringing illustrated by Figure 5 may be substantially reduced.

Finally, we have implemented the algorithm for a Kodak DCS-420 digital camera and made qualitative comparisons of its performance to that of commercially used algorithms. In our judgment, the Bayesian method reduces chromatic fringing more effectively than these other algorithms. We will show examples of these reconstructed images in our talk.

Summary and Discussion

Our interest in trichromatic reconstruction began with the observation that the human visual system is effective at reconstructing images from the responses of interleaved submosaics of retinal cones. In trying to understand how this is possible, we cast the trichromatic reconstruction problem in the form of a statistical decision problem: given the responses of interleaved submosaics of sensors and a model of the statistical distribution of images, what image estimate minimizes the expected reconstruction error? This line of thinking led to a method that may be applied to color CCD camera data. In this paper, we summarized this analysis and its performance.

Trichromatic reconstruction techniques have also been developed in the engineering community. These share with our work the feature that responses from the entire sensor mosaic are used jointly in the reconstruction process. Our work differs in that it starts with explicit description of the structure of natural images (expressed as the prior) and of the sampling device (expressed as the likelihood). Although we have emphasized the trichromatic aspect of the reconstruction problem, it is worth noting that the method we have developed is very general. It provides a recipe for designing optimal reconstruction algorithms that handle irregular polychromatic sampling in the presence of optical blur and sensor noise. Because the method tailors reconstruction algorithms to the properties of the camera and the image population, it can be used to compare how well different camera designs will perform. For example, the number of sensor classes, their spectral sensitivities, the relative numbers and placement of sensors in each class, and the amount of optical blur are all design parameters of CCD cameras. Our method could be used to compare the performance of different cameras when signals from each are reconstructed optimally.

Other trichromatic reconstruction algorithms employ heuristics that attempt to identify the location of edges as part of the reconstruction process. It may be that the good performance exhibited by human vision is driven in part by taking advantage of similar information. Our algorithm is linear and does not perform any explicit segmentation of the image. This is a consequence of the fact that our priors do not incorporate information about the spatial structure of natural objects, many of which are delineated by edges. Our results show that it is worthwhile to exploit the simple facts about natural images that are currently at our disposal. We close by noting that our approach can be extended as we learn to describe more completely the structure of such images.

Acknowledgment

Much of the work discussed here was done while Brainard was a post-doctoral fellow in David Williams' laboratory. We thank him for many useful discussions on the general topic of sampling and reconstruction. Figures 1, 3, 4, and 5 are reprinted from Brainard. Figure 2 is similar to a figure in Brainard and Williams. Brainard is supported by NEI EY 10016.

References


