Color demosaicing by estimating luminance and opponent chromatic signals in the Fourier domain

David Alleysson\textsuperscript{1}, Sabine Süsstrunk\textsuperscript{1}, and Jeanny Hérault\textsuperscript{2}

\textsuperscript{1}Laboratory for Audiovisual Communications
Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland
\textsuperscript{2}Laboratoire des Images et des Signaux
Institut National Polytechnique de Grenoble, France

Abstract

We propose a new method for color demosaicing based on a mathematical model of spatial multiplexing of color. We demonstrate that a one-color per pixel image can be written as the sum of luminance and chrominance. In case of a regular arrangement of colors, such as with the Bayer color filter array (CFA), luminance and chrominance are well localized in the spatial frequency domain. Our algorithm is based on selecting the luminance and chrominance signal in the Fourier domain. This simple and efficient algorithm gives good results, comparable with the Bayesian approach to demosaicing. Additionally, this model allows us to demonstrate that the Bayer CFA is the most optimal arrangement of three colors on a square grid. Visual artifacts of the reconstruction can be clearly explained as aliasing between luminance and chrominance. Finally, this framework also allows us to control the trade-off between algorithm efficiency and quality in an explicit manner.

1. Introduction

Today, most consumer digital cameras capture an image with a single CCD chip to minimize cost and size of the camera. To retain color information, a color filter array (CFA) is placed before the CCD. As a result, there is only one color sensitivity (red, green or blue) available at each spatial location. But for viewing, editing and printing, three colors per pixel (red, green, and blue) are necessary, and therefore reconstructed and available as the output of the camera. Color demosaicing refers to the algorithms that allow recreating a three-color per pixel image from a one-color per pixel image.

The most common spatial arrangement of color filters is called the Bayer CFA, named after its inventor [1] (see Figure 1). This CFA has two times more green than red and blue filters. It is composed of alternate lines of two colors: red/green on odd lines, and blue/green on even lines.

The demosaicing method we propose is not restricted to this arrangement of color filters. However, the Bayer CFA pattern is optimal in terms of spatial frequency representation when arranging three colors on a square grid, as will be discussed in section 2.3.

Creating a three-color per pixel image from a one-color per pixel image can be seen as an interpolation problem [2, 12, 13]. Demosaicing algorithms should reconstruct missing colors. The simplest and most efficient way for demosaicing is the bilinear interpolation. The following convolution filters can easily be implemented in a camera-internal processor, such as a DSP, to compute the interpolation efficiently.

$$F_{RB} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} / 4, F_{G} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 0 \end{bmatrix} / 4$$ (1)

$F_{RB}$ is the bilinear interpolation filter for the red and blue grids, and $F_{G}$ is the interpolation filter for the green grid.

Figure 2 shows an example of such a bilinear interpolation. On the left is the original three-color per pixel image, in the middle the sub-sampled color pixels according to the Bayer CFA, and on the right the image reconstructed with the bilinear interpolation filters of equation (1). The interpolated image shows two artifacts inherent to demosaicing: blurring and the generation of false color, also called color aliasing.
Several solutions have been published that improve this demosaicing result [2-18]. Some use the concept of computing hue as the ratio of two colors [3-5]. Assuming that hue does not change over the surface of an object, it is more reliable to interpolate color ratios instead of R, G, and B in terms of avoiding false color. This approach gives better results than bilinear interpolation, also in terms of reducing blurring. However, artifacts are still visible around the border of an object or in textured regions where the assumption of constant hue does not hold.

To avoid this problem, methods called template matching have been put forth [5,6] that estimate edges in the Bayer CFA image and change the interpolation procedure according to edge behavior. This approach works quite well, despite the problem of having to choose a threshold for the estimation. However, the algorithm varies with the content of a particular image and can become computationally heavy. A variant of this approach uses median filtering instead of convolution filtering [11,14].

Another technique, proposed by Kimmel [10], uses gradient-based interpolation. This algorithm exploits the fact that at each missing pixel, the gradient is known from the existing neighboring pixels. The bilinear interpolation is then weighted by the estimated gradient. This differential method achieves good results except where the gradient is close or equal to zero. The algorithm requires three iterations and is computationally heavy.

Methods using regularization theory [7] or a Bayesian approach [16,17] have also been published. Authors report good results, not easily reproducible, to the detriment of computational complexity. Taubman [23] proposes a preconditioned efficient approach of Bayesian demosaicing that is used in some digital cameras today.

Our approach is similar to the one found by Ingling and Martinez-Urelias [19,20]. They have developed a model of luminance and chrominance, and spatial and temporal multiplexing in the human retina. From this model, Crane et al. derived a color demosaicing algorithm [9]. In this paper, we extend their work to include the relationship between spatial multiplexing and the spatial cone arrangement in the human retina. In particular, a mathematical framework was developed, which demonstrates the properties of spatial multiplexing of colors in the spatial frequency domain. This model, applied to regular arrangements of colors, such as a CFA, shows that luminance and opponent chromatic signals are well localized in the Fourier spectrum [21].

We use this spatial Fourier-transform information to develop a color demosaicing algorithm by selecting appropriate spatial frequencies corresponding to luminance and opponent chromatic signals. When considering spatial frequency localization of luminance and chrominance, we can also prove that the Bayer CFA is the optimal spatial arrangement of three colors on a square grid. Therefore, this CFA pattern is well designed for our algorithm.

Recently, there have been efforts to understand how visual artifacts such as blurring and false colors are created in the reconstruction process, and methods have been proposed on how to minimize them [22]. We will see that our approach facilitates the understanding of artifact generation, as they can be explained in terms of aliasing due to the interaction of luminance and chrominance signals.

2. Spatial-chromatic sampling

For one-chip color CCD cameras, only one color sensitivity per spatial location is available. With this kind of sensor, spatial and chromatic information are therefore intrinsically mixed. In section 2.2, we model this effect as a spatial multiplexing of color. In section 2.3, we demonstrate that this model has the property of localizing luminance and chrominance frequencies in the Fourier domain. In the first section, 2.1, we will define what we understand by luminance and chrominance signals.

2.1 Luminance and chrominance definition

Color images are defined by three-color components at each spatial location. This can be expressed as a vector of three dimensions for each pixel:

\[
\{ C(x,y) \}, i \in \{R,G,B\}, (x,y) \in \mathbb{N}^2
\]  

(2)

Each color vector \( C_i \) corresponds to the sampling of spatially and spectrally variable input illuminance \( E(x,y,\lambda) \) through a spectral sensitivity function \( \phi_\lambda(\lambda) \) given by the photo-sensor characteristics.

\[
C_i(x,y) = \int E(x,y,\lambda) \phi_i(\lambda) d\lambda.
\]  

(3)

It is generally accepted that the color triplets \( \{ C_i(x,y) \} \) form a linear vector space of three dimensions. From this space, spatial information could be computed as a projection onto a one-dimensional axis. Any vector could be chosen as a support for this projection. Since all three-wavelength domains are represented, the probability of detecting spatial information is independent of the wavelength domain it is contained in. Given the projection for spatial information, i.e. luminance, any color image could then be represented as the luminance value plus a vector of three-dimensions called chrominance. Figure 3 shows an example of such decomposition on a three-color per pixel image.

There are several luminance and chrominance definitions in the literature, each applicable in a different
context. In case of the human visual system, the CIE has normalized the photopic function \( V(\lambda) \) found to be representative of the human luminosity sensation. This function corresponds to a positive weighted sum of the sensitivity functions \( \varphi_L, \varphi_V \) and \( \varphi_S \) of the human cones L, M and S.

\[
L = \frac{(R+G+B)}{3} \quad \text{and} \quad C = \{(R-L, G-L, B-L)\}.
\]

In imaging, luminance is often defined to correspond as closely as possible to \( V(\lambda) \), and is calculated as a positive weighted sum of R, G and B. The weights are calculated to minimize the error between the luminance function and the photopic function \( V(\lambda) \). In the demosaicing community, luminance is usually assumed to correspond to the information carried by the green pixels, as the green sensitivity of the camera is closer to \( V(\lambda) \) than either red or blue. The fact that in the Bayer CFA, there are twice more green filters than red and blue is thought to improve the spatial acuity. But this is not necessarily correct because there is no real reason to consider that spatial information is only related to \( V(\lambda) \). For example, we will show that the spatial acuity is identical if we exchange green and red in the Bayer CFA.

Let us start with the premise that luminance can be any weighted sum of R, G and B. We can then try to find the optimal weights such that the luminance channel carries the maximum spatial acuity or frequency resolution. For example, assume that we want to estimate luminance by a 3x3 unitary symmetric convolution kernel with the constraint that the luminance signal would have exactly the same amount of R, G and B. There are three different 3x3 RGB patterns in the Bayer CFA, which result in two conditions. The remaining pattern, exchanging the position of blue and red, does not result in an additional condition. The third condition is that the filter is unitary:

\[
L = \begin{bmatrix} R & G & R \\ G & B & G \\ R & G & R \end{bmatrix} \begin{bmatrix} a & b & a \\ b & c & b \\ a & b & a \end{bmatrix} = \begin{bmatrix} G & R & G \\ B & G & B \\ G & R & G \end{bmatrix} \begin{bmatrix} a & b & a \\ b & c & b \\ a & b & a \end{bmatrix}
\]

\[
\begin{align*}
4a &= 2b \\
2b &= c \\
4a + 4b + c &= 1
\end{align*}
\]

Thus, if the luminance signal is defined as \( L = (R + 2G + B)/4 \), the convolution kernel of equation (4) allows extracting the signal directly from the multiplexed image. Note that the weights of R, G and B for the calculation of the luminance are equal to the probability of the presence of each color on the CFA. Intuitively, this means that luminance is equal to the mean of R, G and B, estimated at each position from the neighborhood.

It is evident that if we exchange, for example, the green and red pixels in the Bayer CFA, the luminance will be calculated by \( L = (G + 2R + B)/4 \). Intuitively, such a pattern layout will work as well as the arrangement in the original Bayer configuration for estimating luminance.

Figure 4 shows the reconstruction of an image with the luminance estimated using the above convolution kernel. For a complete description of the algorithm, see section 3.

We see immediately that the luminance signal is better reconstructed compared to a bilinear interpolation: there are less blurring effects (compare figures 2 and 4). In the next section, we describe a model of multiplexing that formalizes this approach.

### 2.2 Model of spatial multiplexing of colors

We can find a mathematical formulation for the spatial multiplexing of color. First, we have to sub-sample the three-color image according to the CFA array, and then we project the three sub-sampled images into one image. If we call \( R(x,y) \) the spatial multiplexed image, we have:

\[
R(x,y) = \sum_i C_i(x,y)m_i(x,y) \quad (5)
\]

where \( m_i(x,y) \) are three modulation functions having a value of 1 if the color \( i \) is present at position \((x,y)\in\mathbb{N}^2\) and 0 otherwise. For the Bayer CFA shown in Figure 1, the modulation functions can be written as:

\[
\begin{align*}
m_r(x,y) &= (1 + \cos(\pi x))(1 + \cos(\pi y))/4 \\
m_g(x,y) &= (1 - \cos(\pi x)\cos(\pi y))/2 \\
m_b(x,y) &= (1 - \cos(\pi x))(1 - \cos(\pi y))/4
\end{align*}
\]

The modulation functions have a constant part equal to the probability of color presence in the CFA, \( p_i \), plus a fluctuation part with null mean value \( \tilde{m}_i(x,y) \), thus \( m_i(x,y) = p_i + \tilde{m}_i(x,y) \). Therefore, we can rewrite \( R(x,y) \) as follows:
written as:  

\[ R(x, y) = \sum_{\lambda(b, c)} p\lambda(x, y) + \sum_{\lambda(b, c)} \lambda(x, y) \tilde{m}(x, y) \]  

(7)

We decide to call the first part luminance, because it does not depend on the modulation functions at each spatial location. The second term is called chrominance, containing color information that is spatially sub-sampled.

Equation (7) shows that it is possible to split spatial multiplexing of color into two signals. One is defined with the highest spatial frequency since it does not depend on the spatial location. However, it has no spectral behavior. The second signal changes with spatial location and defines a spectral difference.

This “splitting” depends on the CFA under consideration. In the case of Bayer, where there is twice more green than blue or red, luminance should be defined as \((R + 2G + B)/4\) to allow for maximum resolution.

2.3 Fourier representation

Given the formulation (equation 5) of the spatially multiplexed color image, we can compute its Fourier transform as:

\[ R(f, f) = \sum_{(f_1, f_2)} C(f_1, f_2) * \tilde{m}(f_1, f_2) \]  

(8)

Where \(\tilde{m}\) represent the Fourier transform and \(*\) the convolution operator. The modulation functions defined in equation (6) are based on cosine functions and have their Fourier transforms expressed as Diracs:

\[ \tilde{m}_{\lambda}(f_1, f_2) = \pi^{-2} \left( \sum_{1}^{1+1} \frac{1}{1+1} \delta(f_1) \right) \left( \sum_{1}^{1+1} \frac{1}{1+1} \delta(f_2) \right) \]

\[ \tilde{m}_{\lambda}(f_1, f_2) = 2\pi^{-2} \delta(f_1) \delta(f_2) \left( \sum_{1}^{1+1} \frac{1}{1+1} \delta(f_1) \right) \]

\[ \tilde{m}_{\lambda}(f_1, f_2) = \pi^{-2} \left( \sum_{1}^{1+1} \frac{(-1)^{1}}{1+1} \delta(f_1) \right) \left( \sum_{1}^{1+1} \frac{(-1)^{1}}{1+1} \delta(f_2) \right) \]  

with \(f_1 = f_1 - r/2, f_2 = f_2 - s/2\)  

(9)

These modulation functions localize luminance and chrominance in the frequency domain. Thus, the Fourier spectrum of a spatially multiplexed color image can be estimated by simply selecting corresponding frequency domains. Luminance is estimated by low-pass filtering, while chrominance is estimated by high-pass filtering.

For natural images, luminance and chrominance are usually not well separated in the Fourier domain. This results in aliasing between luminance and chrominance, i.e. their spectra overlap. Consequently, our frequency selection procedure will give errors in the estimation. Actually, this aliasing causes the artifacts seen in demosaicing. Artifacts could be classified in four classes.

1. One artifact is the apparent blurring due to the under-estimation of the luminance spectrum, (i.e. choosing a too narrow low-pass filter).
2. Choosing a larger filter could result in a grid effect in flat (homogeneous) regions of luminance due to the high frequency content in the chrominance signal.
3. The same applies for chrominance. If the high pass filter is too large, false color appears, due to high frequencies of luminance in the chrominance signal.
4. If the filter is too small, a “watercolor” effect could appear as colors go beyond the edges of an object.

Generally, for demosaicing algorithms, the two most visible effects are blurring and false color.

Note that this method also allows us to show that the Bayer CFA is the most optimal arrangement in terms of spatial frequency representation when placing three colors on a square grid. The replication periodicity of each color in the CFA is 2 in the horizontal and vertical direction. This replication periodicity in the spatial domain carries the chrominance in the Fourier domain at frequency \(1/2\) in the horizontal and vertical direction. Otherwise said, the chrominance and luminance spectra are maximally

Figure 5 shows an example of the modulus of the fast Fourier transform (FFT) of a one-color per pixel image. We clearly see nine regions where energy is concentrated. These regions correspond to luminance and opponent chromatic signals, arranged as illustrated in Figure 5.

![Figure 5: Fourier representation of a one-color per pixel image.](image)
2.4 Chrominance de-multiplexing

Assume that we are able to estimate completely and without error the luminance part of the signal given by equation (7). We now have to deal with the chrominance part, which is defined by:

\[ \text{Chr}(x,y) = \sum C_i(x,y) \tilde{m}_i(x,y) \quad (11) \]

To better understand what the chrominance signal is, we multiply it by the modulation functions defined in equation (6). This is similar to selecting the chrominance signal in front of a particular photo-sensor type. Note that \( \tilde{m}_i(k) = (1 - p_i) m_i(k) - \sum p_j m_j(k) \) with \( k = (x,y) \), \( m_i, m_j = 0 \) and \( m_i, m_j = m_i \).

\[ \text{Chr}(k)m_i(k) = m_i(k) \left[ (1 - p_i) m_i(k) - \sum p_j C_i(k) \right] \]

\[ = m_i(k) \left[ (1 - p_i) C_i(k) - \sum p_j C_i(k) \right] \quad (12) \]

This leads to:

\[ \text{Chr}(k) = \sum \text{Chr}(k)m_i(k) \]

\[ = \left[ \frac{3}{4} C_x(k) - \frac{1}{2} C_y(k) - \frac{1}{4} C_z(k) \right] m_x(k) \]

\[ + \left[ \frac{1}{2} C_x(k) - \frac{1}{4} C_y(k) - \frac{1}{4} C_z(k) \right] m_y(k) \]

\[ + \left[ \frac{3}{4} C_x(k) - \frac{1}{2} C_y(k) - \frac{1}{4} C_z(k) \right] m_z(k) \quad (13) \]

Equation (13) shows that the chrominance signal is in fact composed of three opponent chromatic signals equivalent to red minus cyan, green minus magenta and blue minus yellow. The high-pass filter on the color spatially multiplexed image results in a multiplexed version of the chrominance. As shown in equation (12), de-multiplexing is obtained by multiplying with the modulation functions.

The de-multiplexed chrominance is still sub-sampled; we can interpolate it using bilinear filters mentioned above. At this point, the interpolation is not critical since human vision is not very sensitive to high spatial frequencies in opponent chromatic channels.

3. The new demosaicing algorithm

In summary, the new algorithm we propose is composed of five steps. First, estimating the luminance signal from the spatially multiplexed image. Second, estimating chrominance by high-pass filtering. This high-pass filter could be designed to be orthogonal to the luminance filter. In that case, the chrominance signal is obtained by subtracting luminance from the multiplexed signal. Third, de-multiplexing the chrominance by multiplying with the modulation functions results in the three opponent chromatic and sub-sampled signals. Fourth, interpolating the opponent chromatic signals and fifth, reconstructing the original image from the luminance and interpolated opponent chromatic signals. Figure 6 shows the synopsis of the algorithm and an example on an image.

As the two main visible artifacts are blurring and false color, the luminance and chrominance filters can be designed in such a way as to reduce them. But it should be noted that high frequencies in chrominance can improve the high frequencies in the reconstructed luminance. This is visible, for example, in Figure 6. Actually, the luminance we have estimated (central figure) appears more blurred than the luminance part of the reconstructed image, because high frequencies in chrominance improve high frequencies in luminance.

For our simulation in Figure 6, we have used the following filter (equ. 14) for estimating the luminance. Its orthogonal filter has been used for chrominance.

\[ F_L = \begin{bmatrix} -2 & 3 & -6 & 3 & -2 \\ 3 & 4 & 2 & 4 & 3 \end{bmatrix} \]

\[ F_C = \begin{bmatrix} -6 & 2 & 48 & 2 & -6 \end{bmatrix} / 64 \quad (14) \]

\[ \begin{bmatrix} 3 & 4 & 2 & 4 & 3 \\ -2 & 3 & -6 & 3 & -2 \end{bmatrix} \]
4. Conclusion

We have developed a mathematical framework for spatial multiplexing of color. This model shows that a one-color per pixel image is equivalent to a sum of luminance and opponent chromatic signals. In the case of a regular CFA arrangement, luminance and chrominance are well localized in the spatial frequency domain, providing a way to estimate them by appropriate frequency selections.

This method also allows us to show that the Bayer CFA is the most optimal arrangement of three colors on a square grid. Additionally, we can clearly explain the artifacts generated by the reconstruction of the image, which can be helpful in the design of selection filters for a particular application or device.

Compared to other demosaicing algorithms, our proposed method has the advantage that only the chrominance is subject to interpolation. The luminance is directly extracted from all known pixel values, i.e. green and red and blue. Interpolating only chrominance information matches well with the color/luminance contrast sensitivity properties of the human visual system. For good quality image reproduction, spatial acuity in color is less critical than luminance acuity.

What remains to be investigated is an appropriate estimator for the luminance. This poses the question of what exactly composes the spatial information in a three dimensional image, where each dimension corresponds to integrated signals of a wideband wavelength domain.

Finally, with this framework, we can define the conditions for which mosaicing can be reconstructed without perceived errors. This can lead to new insights into visually lossless compression of color images.

References


Biography

David Alleysson has a PhD in cognitive sciences from University Joseph Fourier at Grenoble France. He studied the processing of color signal in the retina as a basic model for human color perception. He is now a Post-Doc at the Audiovisual Communications Laboratory continuing his research on color-coding and color non-linearity.