Characterization and measurement of color fringing

Frédéric Cao, Frédéric Guichard, Hervé Hornung and Cédric Sibade

[a]DxO Labs, 3 rue Nationale, 92100 Boulogne-Billancourt FRANCE

ABSTRACT

This article explains the cause of the color fringing phenomenon that can be noticed in photographs, particularly on the edges of backlit objects. The nature of color fringing is optical, and particularly related to the difference of blur spots at different wavelengths. Therefore color fringing can be observed both in digital and silver halide photography. The hypothesis that lateral chromatic aberration is the only cause of color fringing is discarded. The factors that can influence the intensity of color fringing are carefully studied, some of them being specific to digital photography. A protocol to measure color fringing with a very good repeatability is described, as well as a mean to predict color fringing from optical designs.

Keywords: Photography, chromatic aberration, spectral response, blur spot.

1. INTRODUCTION

Dispersion (i.e. refraction index depends on the wavelength) is a physical phenomenon that leads to what optical engineers call "chromatic aberrations" [1]. In a photograph, chromatic aberrations can usually be seen as color fringes along the boundaries of objects. Optical designers take great care to correct chromatic aberrations up to some point, but some residual chromatic aberration can still lead to colored artifacts around objects boundaries in a photograph. This is well known to photographers, who generically call these artifacts lateral chromatic aberration. However, this is improper and lateral chromatic aberration has a precise meaning that shall be detailed further on. As a matter of fact, even with well-corrected optics that exhibit no visible lateral chromatic aberration, purple or blue fringes can be observed at the boundaries of backlit objects, see Fig. 1. The color of the fringes depends on the camera and the observed scene. These fringes can appear at the image center and are symmetrical, that is to say, the color of the fringes is the same on both sides of a backlit object. On the other hand, lateral chromatic aberrations are usually negligible at the image center, and they also yield asymmetric color fringes. Therefore, experimented photographers have roughly classified color fringing into two classes: lateral chromatic aberration and purple (or blue) fringing. The cause of purple fringing has remained a mystery so far, and more or less fancy explanations can be found on the internet, including that purple fringing was actually lateral chromatic aberration (while we will prove it is not). Also, a popular belief is that purple fringing appeared with digital photography. This is not true. However, it can be amplified in digital photography because of several factors that shall be detailed in this article. Most of all, digital images can be zoomed in, much more than silver halide photographs used to be, making color fringing more noticeable. The aim of this paper is to describe the different optical causes of color fringing (Sect. 2), a protocol to measure color fringing in Sect. 3, with the numerous factors that can influence it. A method to predict at different exposure from a single shot and from an optical design is presented in Sect. 4.

2. OPTICAL CAUSES OF COLOR FRINGING

In all what follows, we consider theoretical optical systems, with circular lenses perfectly aligned, that is to say with negligible manufacturing errors, which could include tilt, decentering of the lens and so on.

Further author information: (Send correspondence to Frédéric Cao)
Frédéric Cao: E-mail: fcao@dxo.com, Telephone: +33 (0)155205587

Figure 1. Digital photograph of a backlit tree. Left: luminance channel. Right: blue channel. The blue channel exhibits obviously too large values inside the branches. Since the other channels are dark (as they should be), the branches show large color fringes. The width of these fringes can be at least 4 or 5 pixels.

2.1 Lateral chromatic aberration

When a lens system forms images of different sizes for different wavelengths (due to light dispersion), or spreads the image of an off-axis point into a rainbow, the difference between the image heights for different colors is called lateral color, or chromatic difference of magnification. For a single lens, blue rays will bend more than red rays and thus form their image closer to the optical axis. The effect in the image plane is that magnification is different for each wavelength, see Fig. 2. Therefore, for a system with perfectly aligned revolution lenses, it is essentially a radial phenomenon, null at the image center and it increases when going further in the image field [2, 3]. A white spot on a black background is imaged as a segment oriented to the optical center. Each wavelength is imaged as a point on this segment. Note that the height of a point is not necessarily a monotone function of the wavelength, and depends on the optical design. In this simple model, even though each wavelength is infinitely sharp, a transition containing multiple wavelengths does not appear sharp. See Fig. 4. In photography, it is very usual to have three types of photocells (either in silver halide or digital photography), and lateral chromatic aberrations creates bluish fringe on one side of an object and reddish or yellowish on the other side. A common belief is that lateral chromatic aberration is the cause of purple fringing. This cannot be true for several reasons:

- purple fringing can appear at the image center, whereas there is usually no lateral chromatic aberration (for an optical system with no tilt or centering problem).
- purple fringes are roughly symmetrical on both sides of a thin object, whereas fringes due to lateral chromatic aberrations have a different hue on both sides.
- purple fringes are essentially isotropic whereas lateral chromatic aberrations are mostly directed to the image center.

These remarks are qualitative but the precise chromatic aberrations of an optical system can only be described by the design itself.

2.2 Longitudinal chromatic aberration

Longitudinal chromatic aberration occurs when the position of the plane in which a point source is imaged varies with the wavelength. In this case, it is not possible to focus the optics for all wavelengths simultaneously. The plane in which the image is observed will be called the focusing plane. In particular, for a scene at infinity, the focusing plane is the focal plane. Because of longitudinal chromatic aberrations, all the wavelengths do not focus in the focusing plane. The focusing position is therefore a compromise so that no wavelength is too much "out of focus". Focusing is usually performed for a green reference wavelength, around 550nm [4]. For a wavelength that

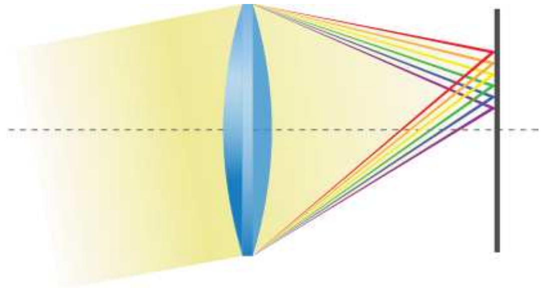

Figure 2. Lateral chromatic aberration. The magnification depends on the wavelength.

does not converge on the image plane, a point light source is imaged as a spot, called the blur spot. When the blur spot is very small, the eye or a sensor cannot discriminate it from a single point. However, it can be large enough so that this approximation no longer holds and the image becomes blurry. A dark/bright transition is more or less blurry, depending on the wavelength. For instance, if the focus in optimized for a wavelength in the middle of the visible spectrum, the wavelengths at the ends of the spectrum may appear blurry. The transition is a mix of long (red) and short (blue) wavelengths and appear as a purple halo. See Fig. 3. That halo can be hardly visible on a low contrasted transition but amplified on an oversaturated high contrasted transition.

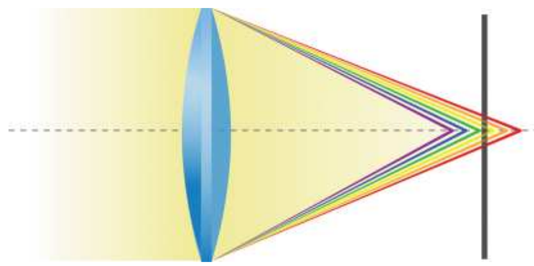

Figure 3. Longitudinal chromatic aberration. The different wavelengths of a light source at infinity on the optical axis converge at different position on the axis. If the image is observed on a plane where the green is optimal, then red and blue appear blurry.

2.3 Third order and higher order aberrations

Up to now, we have only considered first order optics, but real optical systems suffer from higher order aberrations such as the well-known third order Seidel aberrations: spherical aberration, coma, field curvature, astigmatism and distortion. All these aberrations can be wavelength dependent. The consequence is that depending on optical specifications and constraints like small size, small optical track length (TTL), light weight, large aperture, cost and choice of material (which are typical for camera phones optics), an optical designer may not have enough parameters to reduce all aberrations to negligible values. The optical design is therefore a compromise and even though first order (lateral and longitudinal) chromatic aberrations can been made negligible, the blur spot can still depend on the wavelength due to higher order aberrations and to diffraction. Note that the blur spot might be anisotropic, and color fringing is then dependent on the orientation of the edges. Even if not radial, the blur spot usually has an approximate central symmetry, and the colors of the fringes on both sides of an elongated object are usually very close. Moreover, the aberrations described above can also occur close to the optical axis, that is, at the image center. As suggested above, lenses are also optimized to fit the eye sensitivity. It is well known that the sensitivity of the human eye is not constant over the visible spectrum. The relative sensitivity in well-lit conditions is given by the photopic curve, elaborated by the CIE [5]. This curve attains its maximal value at 555nm, so it is sound to make lenses with best accuracy at about this wavelength, corresponding to greenish colors. If the shorter wavelengths are significantly more blurry, a bluish fringe appears; if longer wavelengths are more blurry, a reddish fringe appears. If both are, a purple fringe appears.

2.4 Conclusions

Color fringing can be observed on both digital and silver halide photographs. It has various causes, related to different kinds of chromatic aberrations. To make things simple, we can focus on the consequences of these aberrations on an image that can be observed in a plane. Chromatic aberrations can be essentially summarized by describing the variation of the blur spot as a function of the wavelength. The variation of the position of the blur spot creates lateral chromatic aberration. The difference of shape of the blur spot in the image plane usually creates purple fringing. Although this can be improper since this could create any color in principle, the blur spot is usually optimized for the green (around 550nm). Therefore, different shaped blur spots usually yield purple or blue fringing. Although lateral chromatic aberration can modify purple fringing, it cannot be its main cause.

3. MEASUREMENT OF COLOR FRINGING

3.1 General method

In this section, we describe how color fringing can be measured from an image shot with a digital camera. This measurement can be performed either on JPG or RAW images. The idea is to measure the difference between the color channels across a transition between a neutral dark patch and a neutral bright patch. The method is similar to the one of the ISO 12233 standard for the evaluation of the resolution of camera [6]. The first step is to measure the values of the three channels on lines crossing an edge making a small angle (between $5°$ and $7°$) with the horizontal or the vertical. See Fig. 5. Thanks to this small angle, profiles with subpixellic shifts can be obtained, and agglomerated to obtain a single profile with subpixellic accuracy. This profile is called Edge Spead Function (ESF), since it models the response of the camera to a sharp edge. Let us call $R(x)$, $G(x)$ and $B(x)$ the value of the R, G and B channels at position x, where x belongs to a finite interval, containing all the transition. The process to obtain these channel values influences the quantity of color fringing, and this will be detailed further on. For the sake of simplicity, we just assume that R, G and B is representative of the number of photons coming in bandwidths respectively centered at large, medium and small wavelengths of the visible spectrum.

In what follows, the G channel is always taken as the reference channel. The definitions are given for the blue channel, but the same apply for the red one. Blue fringing is due to an excess of blue with respect to green. To make the measurement invariant with respect to the exposure, the excess must be relative. Hence, the relative difference of blue and green is calculated. It is defined as

$$d_{BG}(x) = \frac{B(x) - G(x)}{\max G - \min G}. \tag{1}$$

The maximal and minimal value of G are obtained far enough from the transition.

Definition 3.1. *Let α be a real value in $(0,1)$. We call* blue fringing area*, the set of points*

$$A_{BF} = \{x,\ d_{BG}(x) \geq \alpha\}. \tag{2}$$

The blue fringing size is $|A_{BF}|$, the length of the fringing area. We call blue fringing intensity *the value*

$$I_{BF} = \frac{1}{|A_{BF}|}\int_{A_{BF}} d_{BG}(x)\,dx, \tag{3}$$

that is, the mean value of the relative difference of B and G over the blue fringing area.

The same holds for the red fringing. In practice, we noticed that $\alpha = 0.05$ was a good value for usual exposure conditions.

This definition is quite simple but several remarks can be made. First, an excess of green with respect to blue can also be of interest, in which case we replace A_{BF} by $A^*_{BF} = \{x,\ d_{BG}(x) \leq -\alpha\}$. The blue fringing intensity is then a negative value. As noted above, there may be red fringing and blue fringing at the same place, which is typical of achromatic lens focusing the red and the blue at the same position. This is typical of optics optimized on the medium wavelengths. In this case, the apparent color is more purplish for positive blue and red fringing. In case of positive blue fringing and negative red fringing, the dark side of a transition appears cyan. Let us assume that there is no lateral chromatic aberration and that the blue channel is more blurry than the green channel. Then, on a dark/light transition, positive blue fringing appears on the dark side of the transition, while negative blue fringing appears on the bright side of the transition. If the response $G(x)$ and $B(x)$ are linear functions of the illumination, then the amount of positive and negative fringing are the same (up to sign). However, if the bright side of the transition is saturated, which is often the case with a backlit object, only positive blue fringing is observed on the dark side of the transition.

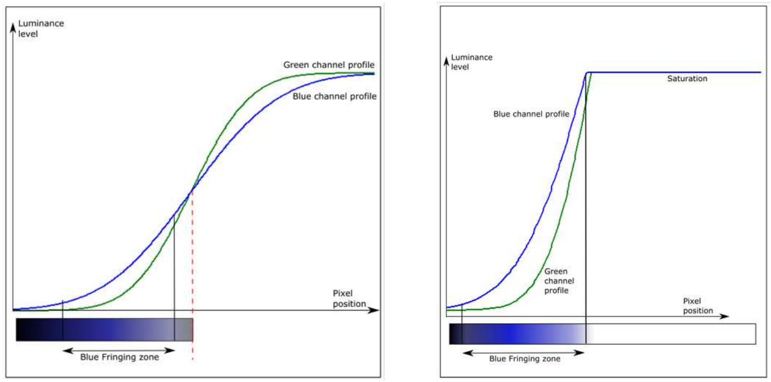

Figure 4. Comparative profile of the green and blue channel across an edge. When the blue channel is more blurry, an excess of blue can be observed on the dark side of the transition (left). The excess of green in the bright part is usually not observed for two reasons. First, the bright part is usually saturated (right), and when it is not the case, the channel differences are attenuated by the tonal curve in the highlights while amplified in the shadows.

3.2 Protocol

In order to obtain images with a high dynamic range, we measure the response to a slanted edge on a transmissive target. The shot must be taken in a dark room, and the only illumination comes from the bright part of the slanted edge. In order to measure the color fringing in the horizontal and vertical directions on a dark/bright and bright/dark transition, the light only emits through a square with slanted sides. The luminance of the device must be uniform. The following parameters must be reported:

- the source luminance (in cd/m^2).
- the illuminant temperature, or better, the spectrum measured with a spectrometer.
- the shooting distance
- the focusing distance
- the exposure time
- the lens aperture
- the ISO setting
- the position in the image field
- the orientation of the transition (horizontal/vertical)
- the type of transition (dark/bright or bright/dark)
- any setting of the Image and Signal Processing (ISP) (as sharpening, denoising, color saturation...)
- the type of the image (RAW or JPG)

In order to measure how much the sensor is saturated, a T4 105 Stouffer target can be set close to the slanted square. This target consists of patches with linear varying density (hence exponential transmittance), and is used to detect the luminance saturating the sensor.

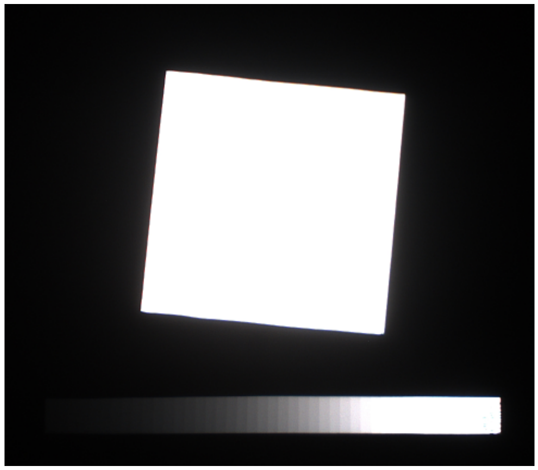

Figure 5. Blue fringing measurement chart. The target is transmissive and allows high illumination ($1500 cd/m^2$ in the present case). The saturation luminance is determined using the Stouffer target below the square. The square is slanted at an angle between $5°$ and $7°$ to obtain profiles with super resolution.

3.3 Direct measurement

This consists in photographing the slanted square source and reporting the size and the intensity of the blue and red fringing. This is adapted for both RAW and JPG and the exposure parameters need to be specified. The T4105 Stouffer target is used to determine the saturation level of the sensor. Remark that on RAW images, the Stouffer target is also used to estimate the white balance correction. For JPG images, white balance is supposed to be corrected, but it might be necessary to adjust it if the difference between channels is too large. Indeed, the fringing measure would be more representative of the white balance bias than of the optical aberration.

On Fig. 6, the profiles of the four RAW channels (R, Gr, B, Gb) of a 2Mpixel camera module are displayed. The blue channel is obviously more blurry than all the other channels. The red channel is as sharp as the green ones. Hence, this sensor exhibits blue fringing (not purple). The profile shown on this figure corresponds to a non-saturated shot. Since the sensor response is linear, all the fringing measurements lead to the same value for non-saturated shots. For saturated shots, the size and the intensity of the color fringing increase with the exposure and empirically stabilize for exposure about 4 times the dynamic of the sensor, although it is dependent on the MTF profile. For a same camera module, measurements can be done in RAW and JPG format. This measurement in JPG is dependent on the color rendering and the possible sharpening applied to the image, whereas the RAW measurement is more representative of the optical aberration. For low exposure levels, the ISP can correct blur on the different channels. Since the correction can be adapted to the channel, blue fringing can be essentially canceled as long as the sensor response is linear. After saturation, such a simple correction is not as efficient, and the JPG values are closer to the RAW values. They still can differ because of the color matrix, the tone curve and the possible filters that are applied in the raw conversion.

3.4 Influencing factors

In order to fully describe the different factors of color fringing, it is necessary to detail some parts of the image acquisition process. The light coming to the sensor has a certain spectral distribution $I(\lambda)$ giving the quantity of energy at the wavelength λ. Note that this light can directly come from a light source and be transformed by the medium it crosses (as the sunlight crossing the atmosphere, making the sky blue), or it can be partially reflected by an object, in which case it is modified by the physical properties of the object. In any case, a color sensor filters out the wavelengths selectively in the different color channels. This process is usually described by the quantum efficiency (QE) of the sensor, which gives the (statistical) number of free electrons generated by a single photon hitting the photosensitive cell with a given wavelength. Each of the color channels R, G and B has

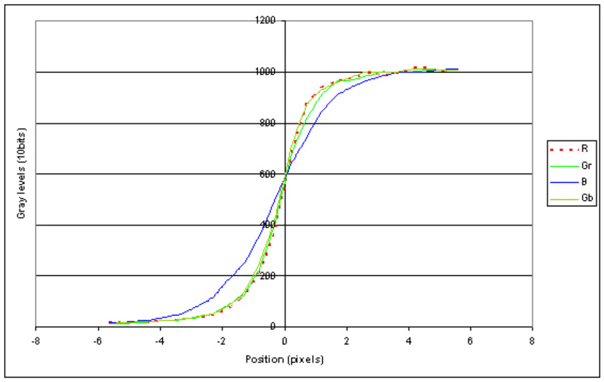

Figure 6. Profile of the four raw channels along a transition. Blue fringing is observed since the red and green channels have the same level of sharpness, contrary to the blue channel which is more blurry.

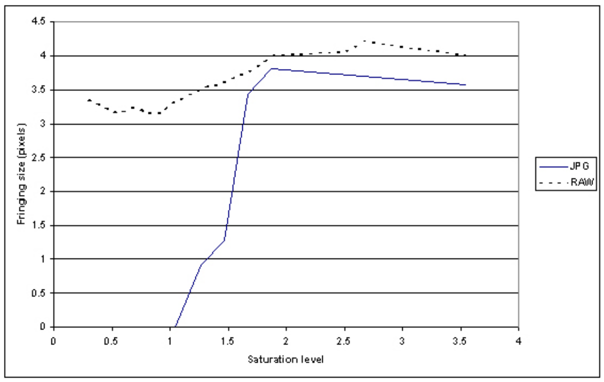

Figure 7. Measurement in RAW vs. JPG for a same camera module. When the image is not saturated, blur can be corrected by linear filters. The ISP corrected the longitudinal chromatic aberration in this case, and the blue fringing equals 0 for the lowest exposures. However, when the image is saturated, the same correction is inefficient and RAW and JPG values show the same behavior. Differences can be due to the ISP, in particular the tone curve and the color matrix (see text).

its own quantum efficiency. Thus, the number of electrons generated by the incoming light is

$$N_e = \frac{1}{hc} \int I(\lambda) QE(\lambda) \lambda \, d\lambda, \tag{4}$$

where h is the Planck constant and c the speed of light. A RAW sensor (which may be CCD or CMOS) basically measures this number of electrons, up to different gain factors, and several types of noise as thermal agitation, analog/digital conversion, etc. If a neutral material is observed, the light coming to the sensor has the same spectral distribution as the illuminant, up to a multiplicative factor depending on the physical properties of the material. In any case, the values of the RGB color channels depend on the illuminant and the quantum efficiencies of the color filters. At least, a measurement of color fringing also has to report the illuminant color temperature, or better its spectral distribution. It is usual to assume that the camera has to compensate the illuminant, through a white balance correction. It makes more sense to measure the color fringing after white balance correction. This makes the measurement more independent of the illuminant (though not completely, especially if the illuminant spectrum has very localized rays as fluorescent tubes). Note that white balance can be tuned a posteriori for digital images. For silver halide photography, the white balance is chosen a priori with the film, and it does not depend on the camera (the variation of the transmission of the lens with wavelength is neglected). If the sensor has a linear response, then the quantity defined in (1) is independent of the exposure: increasing the light intensity, or the exposure time does not change this value, and the color fringing is also constant. (Remark that increasing the aperture increases the quantity of light, but it may change the blur spot, thus color fringing). However, the sensor has a saturation level, and the value measured by (1) depends on the exposure, once the sensor saturates. Hence, the exposure of the sensor has to be given for any measurement. Silver halide films do not have a linear response, especially in highlights and shadows, and this change the value of color fringing. Even after white balancing, the response of a RAW sensor does not correspond to the RGB values in the final image. The most basic color rendering contains at least two more steps:

1. a chromatic correction, that basically consists in mapping the spectral responses of the sensor onto the color matching functions $\bar{x}$, $\bar{y}$, $\bar{z}$. The most simple model is a linear transformation (thus a 3×3 matrix) applied on the raw R, G, B channels. The color matrix usually increases the intensity of color fringing.
2. a tone curve, also called the Opto-Electronic Conversion Function (OECF), operating a non linear mapping on the R, G, B values. A classical tone curve is a power $\frac{1}{2.2}$, but it varies a lot from one camera to the other. However, the tone curve amplifies the contrast in the shadows and compress the contrast in the hightlight. Therefore, it can also increase color fringing on the dark side of a transition.

Every other step of the ISP is also susceptible to change the value of color fringing. For instance, sharpening algorithms change the profile of values across a transition, hence the color fringing. Note that in theory, it is possible to use a deconvolution kernel to inverse the blur on each channel, and to eliminate color fringing. However, this is not feasible in practice. First, deconvolution is an unstable process, very sensitive to noise, and the camera response is non linear due to saturation and tone curve (for JPG images). Demosaicing also mixes the different color channels and attempt to reconstruct information up to the Nyquist frequency, even though the channels are subsampled. Therefore, measurements in RAW or JPG format are expected to be different. The effect of lateral chromatic aberration is twofold: first, since the color channels of the sensor are composed of different wavelengths that are shifted with respect to one another, each color channel is blurry. Second, the central positions of the color channel are also different, and the profiles are globally shifted. This effect is very much dependent on the field position of the illuminant, and on whether the transition is dark/bright or bright/dark.

4. COLOR FRINGING PREDICTION

4.1 From a non saturated shot

This prediction is valid for images in RAW format. Since the response of the sensor is linear (up to the sensor saturation), the profile of the RAW color channels can be predicted for any exposure. In addition, any known part of the ISP can be applied. In particular, the color matrix can have an important impact on the color fringing intensity (slightly less on the size of color fringing). The practical limitation, that prevents to reach very high saturation levels, is the amplification of noise. However, it is usually possible to attain saturation factors of about 8.

4.2 From an optical design

Optical design software programs are able to compute the path of any light beam for an arbitrary wavelength. If the quantum efficiencies of the sensor are known, it is possible (by simple linearity) to estimate the blur spot at any position of the field (taking diffraction blur into account), as well as lateral chromatic aberration. Therefore, it is possible to estimate the ESF for any edge, in any orientation and at any position in the field. Lateral chromatic aberration can be taken into account at will. Moreover, there is no limitation on the exposure, since noise is due to rounding errors and can be arbitrary small. It is also possible to take manufacturing errors into account by introducing tolerance bounds on the design.

5. CONCLUSIONS

The generic term color fringing refers to several different optical phenomena. Contrary to what can sometimes be read, the so-called purple or blue fringing is not an electronic phenomenon exclusive to digital photography. It is not lateral chromatic aberration either. It is due to longitudinal chromatic aberration, or more generally to the fact that the size of the blur spot varies with the wavelength. A protocol to measure purple fringing was

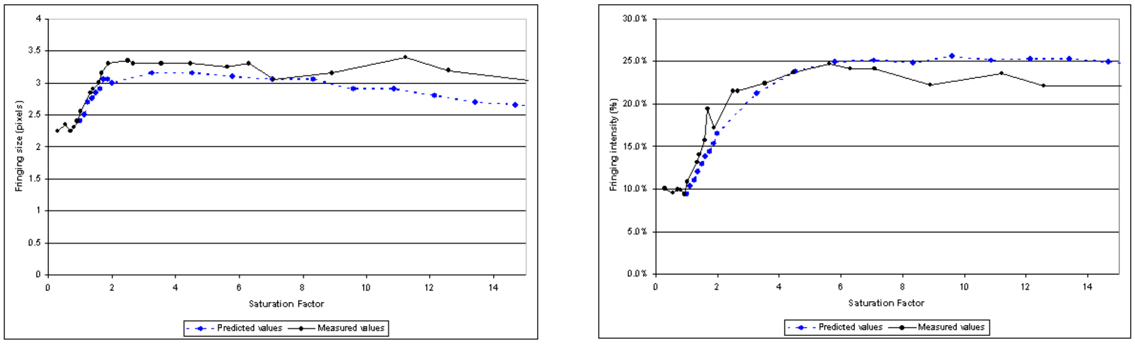

Figure 8. Prediction of blue fringing from a non saturated profile. Left: fringing size; right: fringing intensity. Different values of gain and a clipping were applied to a non saturated profile to simulate a saturated profile at different levels of exposure. A single shot leads to the dashed line. The solid line is obtained from different shots at different exposures. The prediction is good, the inaccuracy coming mostly from noise.

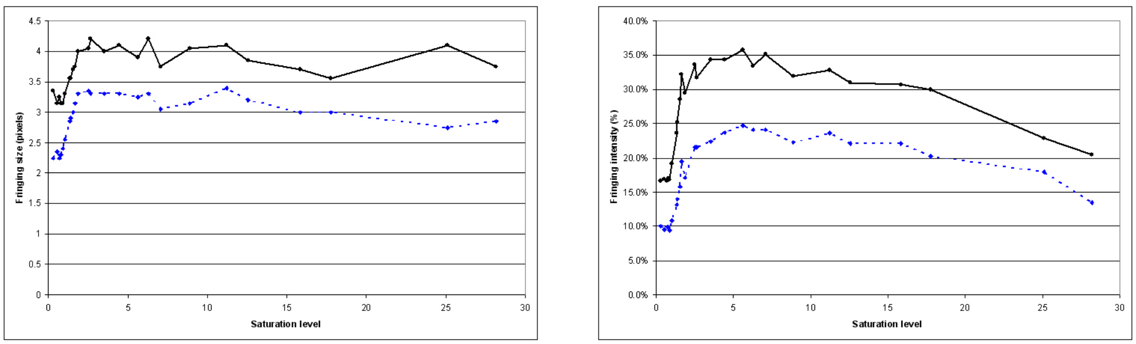

Figure 9. The rendering of colors needs to map the color space of the sensor onto the color space of the output device (or any camera independent color space). This depends on the spectral responses of the sensor, and the illuminant. On the above experiment, the dashed curve corresponds to raw measurements at different exposure levels. The solid curve corresponds to the same measurement after the color matrix has been applied.

proposed. For RAW images, the value of purple fringing can be predicted at any saturation level from a non saturated shot. It can also be predicted from an optical design, although camera module manufacturing errors can bias the measurements.

REFERENCES

1. M. Born and E. Wolf, *Principle of Optics*, Pergamon Press, New York, 1964.
2. W. Welford, *Aberrations of Symmetrical Optical Systems*, Academic Press, New York, 1974.
3. W. Welford, *Aberrations of Optical Systems*, Adam Hilger, Bristol, 1986.
4. W. Smith, *Modern Optical Engineering: The Design of Optical Systems*, Mc Graw Hill, New York, 1990.
5. Commission Internationale de l'Eclairage, *Colorimetry, 3rd edition*, CIE, 2004.
6. ISO, *Photography — Electronic still-picture cameras — Resolution measurements (ISO 12233)*, ISO International Standard, 2003.