Model-Based Color Image Sequence Quantization *†

C. B. Atkins, T. J. Flohr, D. P. Hilgenberg, C. A. Bouman, and J. P. Allebach
School of Electrical Engineering, Purdue University, West Lafayette, IN 47907-1285

Abstract

We investigate the display of color image sequences using a model-based approach to multilevel error diffusion. We extend Bouman and Kolpatzik's technique [1] for design of an optimal filter to the temporal dimension. Our model for the human visual system accounts for the spatial and temporal frequency dependence of the contrast sensitivity of the luminance and chrominance channels. We observe an improvement in image quality over that yielded by frame-independent quantization, when the frame rate is sufficiently high to support temporal averaging by the human visual system.

1 Introduction

The desire to display sequences of color images on computer monitors (desktop video) has invoked new attention toward a problem which has been widely addressed in a lower-dimensional context: color image quantization. The term image quantization refers to the reproduction of a continuous-tone image using colors from a (limited) palette with the objective that the quantized image resemble its original as closely as possible. Image quantization procedures can be computationally intensive; and generally the quality of reproduction reflects the computational level of the algorithm. In this work, we combine a universal color palette with an optimized error diffusion scheme in order to quantize sequences of images for video rendering on a computer monitor.

Prior approaches to this problem include work by Mulligan [2] who used an extension of the dispersed-dot screening algorithm where the screen thresholds were distributed spatially and temporally. Gotsman [3] combined a model for the human visual system with a search-based halftoning algorithm for the quantization of image sequences. His model exploits reduced contrast sensitivity to high spatial frequencies, but it does not account for the dependence of contrast sensitivity on temporal frequency.

The palette used in the quantization of an image can be image-dependent or image-independent. An image-dependent palette is better from the standpoint of quantized image quality; such a palette is able to use known statistics of the frequency of occurrence of colors in one image in order to more finely quantize regions of the color space which are more densely populated. Examples of algorithms for image-dependent palette design are given in [4, 5]. Image-dependent palette design is, however, computationally intensive, since the palettization must be done separately for every different image. Although image-independent palette design is suboptimal from the standpoint of tonal reproduction, only one palette design is needed. Some of the loss in tonal reproduction can be

*This work was supported by Hewlett-Packard Company.

†Proceedings of the *SPIE/IS&T Conf. on Human Vision, Visual Processing, and Digital Display V*, vol. 2179, pp. 310-317, San Jose, CA, Feb. 7-10, 1994.

regained via a halftoning technique which redistributes the quantization error to spatial frequencies where it is less noticeable to a human viewer.

A commonly used halftoning technique, introduced by Floyd and Steinberg [6], is known as error diffusion. In this algorithm, the quantization error of previously quantized pixels is filtered and distributed forward to unquantized pixels. Using a model for the human visual system, we can design the filter to attenuate the resulting display error spectrum at frequencies for which contrast sensitivity is high.

Kolpatzik and Bouman [1] showed that the error diffusion filter which minimizes the expected mean-squared error can be computed assuming that the noise resulting from the quantizer is white, and that the monitor and human viewer can be modeled as linear, shift-invariant systems. In this paper, we extend their work to determine an optimal error diffusion algorithm for image sequences. This model-based development involves not only a spatial frequency-dependent characterization of the contrast sensitivity of the human visual system, but one which also will allow us to exploit the reduced contrast sensitivity to stimuli which vary temporally.

The remainder of this paper is organized as follows. Section 2 describes our choice of palette. Section 3 reviews model-based quantization, and describes spatiotemporal error diffusion. Section 4 presents our spatiotemporal contrast sensitivity model for the human visual system. Section 5 describes our results; and Section 6 outlines our conclusions.

2 Palette Design

In this section, we describe the palette which we employed for our quantization scheme. We also describe the procedures which are used to prepare the input image data for quantization, and the quantized images for display. These procedures are depicted in Fig. 1. Let D be the digital value input to the D/A converter which drives any one of the three CRT guns. The displayed luminance Y, scaled to lie between 0 and 1, is then given by

$$Y = \left(\frac{D - offset_{mon}}{255 - offset_{mon}}\right)^{\gamma_{mon}}. \qquad (1)$$

In anticipation of this relationship between the input and output of a monitor, the Society of Motion Picture and Television Engineers (SMPTE) has established a standard $\gamma_{smpte} = 2.2$ and $offset_{smpte} = 0$ for the storage of continuous-tone images.

The first step in the procedure for image quantization is to transform the original image data, which is stored in SMPTE gamma-corrected coordinates D_{smpte}, to coordinates D_{lin} which are linearly related to the luminance displayed by the monitor. This is necessary because the spatial averaging by the human visual system, which results in the perceived tone, does not commute with the nonlinear transformation of the pixel colors. These two coordinate systems are related by

$$D_{lin} = \left(\frac{D_{smpte}}{255}\right)^{\gamma_{smpte}}. \qquad (2)$$

Note that the range for D_{smpte} is 0 to 255 whereas that for D_{lin} is 0 to 1. This transformation is applied pixel-by-pixel to each of three color coordinates in the image data.

The next step shown in Fig. 1 is to quantize and halftone the linear image data. For the quantization of our image sequences, we employed a palette which is separable in the (R, G, B) space of the monitor [7]. More sophisticated palettes are available [4, 5]. We desired that the output levels of the R, G, and B quantizers be perceived as uniformly spaced. As a basis for achieving perceptually equal steps in luminance, we chose the power law relation between displayed luminance Y and lightness L^*, which is intrinsic to the CIE uniform color spaces $L^*u^*v^*$ and $L^*a^*b^*$. Let N be the desired number of quantizer output levels. This results in output levels

$$d_i = \left(\frac{i}{N-1}\right)^{3.0}, \tag{3}$$

for $i = 0, \ldots, N-1$.

After quantizing the image data, we correct for the specific gamma and offset of the monitor on which it is to be displayed. For our display monitor, we measured $\gamma_{mon} = 2.3$ and $offset_{mon} = 40$. The corrected image data D_{corr} is then given by

$$D_{corr} = \tag{4}$$
$$(255 - offset_{mon}) D_{quant}^{\left(\frac{1}{\gamma_{mon}}\right)} + offset_{mon},$$

where $D_{quant} \in \{d_i : i = 0, \ldots, N-1\}$ is the quantized linear image data. When the corrected image data D_{corr} is input to the D/A converters that drive the monitor, the resulting display luminance will be proportional to D_{quant}.

3 Model-Based Quantization

To extend error diffusion to the temporal domain, we modify the region of support of the filter as shown in Fig. 2. This exploits spatial and temporal averaging of the quantized pixel values by distributing the error to the four contiguous, unquantized pixels of the present frame and to the nine contiguous, unquantized pixels of the following frame. In the rest of this section, we develop a method for choosing the weights of the filter.

Central to this analysis are the two assumptions mentioned earlier. First, the number of quantization levels for each quantizer must be sufficiently large to ensure that the quantization error can be modeled as a white, stationary random process; and second, the human visual system is characterized as linear and time- and shift-invariant. With the second assumption, we can represent the human visual system with a filter $W(\omega)$.

A block diagram of the error diffusion algorithm appears in Fig. 3. The input to the system is a sequence of continuous-tone pixels $f_i(n)$. Here the parameter n indexes into the discrete space-time domain of the image sequence, according to the order in which the pixels are scanned. The input pixel values are modified by the addition of past quantization errors $e_q(n)$, filtered in the feedback branch of the diagram. The quantization error is defined as

$$e_q(n) = m(n) - f_o(n); \tag{5}$$

so the input to the quantizer is the modified pixel value

$$m(n) = f_i(n) + [e_q(n) * g(n)], \tag{6}$$

where $g(n)$ represents the filter coefficients and $*$ represents 3-dimensional convolution. We define the display error as the difference between the input and the output of the system:

$$e_d(n) = f_i(n) - f_o(n). \tag{7}$$

It is the display error $e_d(n)$ which is filtered by the human visual system $w(n)$ to produce the perceived error $e_p(n)$.

It can be shown [4] that the display error spectrum is given by

$$E_d(\omega) = E_q(\omega)[1 - G(\omega)]. \tag{8}$$

Since the quantization noise $E_q(\omega)$ is assumed to be white, (8) shows that the filter frequency response $G(\omega)$ directly shapes the spectrum of the displayed error. We would like to use this property of the error diffusion algorithm to design a filter which attenuates the displayed error at frequencies where the human visual system is most sensitive.

We incorporate the model for the human visual system as in Fig. 4, and define an error metric ϵ_p by squaring the output of the filter $W(\omega)$ and integrating. Since the filters $G(\omega)$ and $W(\omega)$ are linear, and time- and shift-invariant, their ordering as shown in Fig. 4 can be switched. Fig. 5 depicts the rearranged system in the space-time domain. Since the filter $g(n)$ is strictly causal, *i.e.* its coefficients are nonzero only for previous inputs $u(m)$, it can be interpreted as a linear predictor for the current value $u(n)$.

The optimal filter coefficients are given by the solution to the optimal linear prediction problem [1]. The random process $u(n)$ is the output of the filter $w(n)$ driven by the white random process $e_q(n)$. Hence the power spectral density of $u(n)$ is given by

$$\Phi_u(\omega) = N_o \left|W(\omega)\right|^2, \tag{9}$$

and the autocorrelation of $u(n)$ is given by

$$R_u(n) = E\{u[m+n]u[m]\} = \mathcal{F}^{-1}\{\Phi_u(\omega)\},$$

where $\mathcal{F}^{-1}\{\cdot\}$ denotes the inverse Fourier transform.

Let

$$\mathbf{g} = [g_{x_1}, \ldots, g_{x_p}]^T$$

be a vector containing the nonzero coefficients of the desired filter impulse response. Here p is the order of the filter; and the indexing sequence $\{x_i\}_{i=1}^{p}$ is selected according to the region of support of the filter, as shown in Fig. 2, where $p = 13$. We similarly define a vector

$$\mathbf{u} = [u_{x_{-1}}, \ldots, u_{x_{-p}}]^T$$

containing past samples of the random process $u(n)$. The filter output which predicts the value of u_{x_0} is formed by the inner product

$$u_{x_0} = \mathbf{g}^T\mathbf{u}. \tag{10}$$

It is shown in [1] that the optimal filter coefficients $\mathbf{g}$ are given by

$$\mathbf{g} = \mathbf{R}^{-1}\mathbf{b}, \tag{11}$$

where $\mathbf{R} = E\{\mathbf{u}\mathbf{u}^T\}$ and $\mathbf{b} = E\{u_{x_0}\mathbf{u}\}$. The elements of $\mathbf{R}$ and $\mathbf{b}$ are values of the autocorrelation $R_u(n)$ defined above.

4 Human Visual System Model

In this section, we develop the human visual system model which we use to determine the optimal error diffusion filter. The human visual system processes color stimuli in an opponent color space; hence for color error diffusion, we filter the quantization error in an opponent color space. We will assume that the spatiotemporal frequency responses of both chrominance channels are identical. The task is then to determine two filters: one for the luminance channel and one for the chrominance channels.

Up to this point, we have expressed spatiotemporal frequency as $\omega \in I\!R^3$. We assume that the contrast sensitivity function is circularly symmetric in spatial frequency. We specify the radial component of spatial frequency as

$$\omega_s = \sqrt{u^2 + v^2} \in I\!R^1, \tag{12}$$

where u and v represent horizontal and vertical spatial frequencies; and we denote temporal frequency as ω_t. Then

$$\omega = (\omega_s, \omega_t) \in I\!R^2. \tag{13}$$

To measure spatiotemporal contrast sensitivity, the subject views a sinusoidal grating

$$B + \Delta B sin(\omega_s x + \omega_t t)$$

traveling horizontally at a constant rate. At a given point in the stimulus field, the luminance of the stimulus varies sinusoidally in time with frequency ω_t; while at a fixed point in time, the stimulus is a (stationary) sinusoidal grating with spatial frequency ω_s. The subject adjusts ΔB until they are able to just discern the moving grating. This yields a "threshold" modulation of $B/\Delta B$. The contrast sensitivity is defined as its reciprocal.

Assuming that spatial contrast sensitivity is circularly symmetric, we can completely characterize the spatiotemporal frequency response as a function of the spatial frequency ω_s and the temporal frequency ω_t. Portions of the 3-dimensional "surface" of luminance contrast sensitivity for the human visual system have been measured by many researchers [8, 9, 10, 11, 12].

Kelly [9] has measured the luminance contrast sensitivity function $W(\omega_t, \omega_s)$ of the human visual system over a wide range of stimulus frequencies. In his study, the stimuli were stabilized on the subject's retina. Since the visual system only responds to a changing stimulus [10], the perception of stationary objects depends on small, jerky movements of the eye called *saccades* which induce movement of the image on the retina. Retinal stabilization of the stimulus effectively cancels out the effects of saccades, severely attenuating measured values of contrast sensitivity for stimuli at very low temporal frequencies. For a given spatial frequency, Kelly's study indicates that the luminance channel temporal frequency response is bandpass. Our stimuli are never fixed on the viewer's retina; hence this model does not perfectly characterize the operation of the visual system when viewing image sequences.

Kelly has fit the data obtained in his experiments with the mathematical function

$$W'(\omega_s, \omega_t) = \left[6.1 + 7.3\left|\ln\left(\frac{\omega_t}{3\omega_s}\right)\right|^3\right] \omega_t\omega_s \exp\left(-\frac{2(\omega_t + \omega_s)}{45.9}\right), \tag{14}$$

where ω_s is in cycles per degree and ω_t is in Hz. To determine the contrast sensitivity function to be used for design of the optimal filter coefficients, we must transform $W'(\omega_s, \omega_t)$ from the continuous-parameter domain at the retina of the viewer to the discrete space-time domain at the display. This transformation depends on viewing distance, display resolution, and frame rate. A plot of the luminance contrast sensitivity function for a viewing distance of roughly 22 inches, a display resolution of 100 pixels per inch, and a frame rate of 60 Hz appears in Fig. 6.

Fig. 7 shows the filter coefficients obtained using the model for the luminance channel for frame rates of 30 Hz and 60 Hz. Note that the coefficients of the filters are nearly identical for the present frame; but that the most significant previous-frame coefficients for the 60 Hz filter have roughly twice the magnitude of their counterparts in the 30 Hz filter. Both filters were designed for the same viewing distance; and as spatial filters they should be nearly the same. This is the reason for the similarity between the filter coefficients in the present-frame plane. The difference between the coefficients of the previous-frame filter plane is due to the greater frame rate of 60 Hz which accommodates temporal averaging of pixels in the quantized image sequence.

Spatiotemporal contrast sensitivity for the chrominance channels has been measured by van der Horst and Bouman [11]. In their study, the stimuli were not stabilized on the subject's retina. Their study indicates that the temporal frequency response is strictly lowpass. They do not provide a mathematical function that represents the contrast sensitivity of the chrominance channels. For the results obtained in this paper, we scaled the temporal axis of the function, which Kelly gave for the luminance channel, by a factor of 7 in order to reflect the lower cutoff temporal frequency of the chrominance channels. While not completely consistent with the data obtained in [11], this strategy allowed us to consistently utilize the structure of a contrast sensitivity function. A plot of the chrominance contrast sensitivity function for a viewing distance of roughly 22 inches, a display resolution of 100 pixels per inch, and a frame rate of 60 Hz appears in Fig. 8.

Fig. 9 gives the filter coefficients obtained using the models for the chrominance and luminance channels for a frame rate of 60 Hz. The coefficients of the filters are not significantly different for the present frame; but the previous-frame coefficients for the chrominance filter are greater than their counterparts in the luminance filter, suggesting that the lower bandwidths of the chrominance channels in the visual system can be exploited by temporal averaging of chrominance information.

A block diagram of our color error diffusion system appears in Fig. 10. Transformation of the R, G, B error data to and from the Y, rg, yb opponent-color space of the filters is performed by the matrix $\mathbf{T}_{R,G,B \to Y,rg,yb}$ and its inverse. The three quantizers $Q(R), Q(G)$, and $Q(B)$ are shown to emphasize the separate quantization of the R, G, and B values of the modified pixels.

5 Results

We viewed color and monochrome image sequences on the monitor of a Hewlett-Packard 755 workstation at frame rates of 30 and 60 Hz. We quantized to 6 levels for the R and G primaries and to 4 levels for the B primary. In the monochrome case, we quantized to 8 gray levels. In addition to the sequences quantized using our optimized algorithm, we viewed sequences quantized using error diffusion with two other filters. In one, we used the filter introduced by Floyd and Steinberg [6]. In the other, we used an optimal 2-dimensional filter proposed by Kolpatzik and Bouman in [1]. Their filter uses a model for contrast sensitivity which is only dependent upon spatial frequency.

The expected result was that temporal averaging would result from quantization using the 3-dimensional optimized filter. This was the case with a frame rate of 60 Hz. The monochrome sequences quantized with the optimized 3-dimensional filter appeared smoother spatially and temporally. The color sequences quantized with the optimized 3-dimensional filter appear marginally smoother, spatially and temporally. Temporal averaging with a frame rate of 30 Hz was not achieved; according to the model which we used, contrast sensitivity is not sufficiently attenuated at temporal frequencies up to 15 Hz. All of the quantized image sequences for the 30 Hz test case appeared comparable in quality.

A comparison of single frames from each of the quantized sequences reveals that our 3-dimensional filters permit more error at lower spatial frequencies than do the 2-dimensional filters proposed in [6] and [1]. This was not surprising, as the 3-dimensional filter was designed for use with image sequences. Our model explicitly takes reduced contrast sensitivity to high temporal frequencies into account; this can only be exploited with frame rates high enough to accommodate temporal averaging of the colors or gray levels in the quantized image sequence.

6 Conclusions

We have observed that spatiotemporal error diffusion does move texture from the spatial domain into the temporal domain.

One possible improvement would be to use a better spatiotemporal model. Specifically, a model which is strictly lowpass in temporal frequency response would more accurately characterize the human visual system.

Another possible improvement would be to perform quantization in an opponent-color space. Van der Horst and Bouman [11] have noted that luminance information coding by the neural system takes place with much greater bandwidth than does coding of color information by either of the visual system's opponent-color channels. Hence it would seem that quantization is more naturally performed in an opponent color space, and to finer resolution along the luminance axis in that color space.

The advantage of spatiotemporal quantization over 2-dimensional methods is realized only when reduced contrast sensitivity to high temporal frequencies can be exploited. This is possible when sequence frame rates that allow temporal averaging are used.

References

[1] B. Kolpatzik and C. A. Bouman, "Optimized Error Diffusion for Image Display",

J. of Electronic Imaging, **1**(3), pp. 277–292, 1992.

[2] J. B. Mulligan, "Methods for Spatio-temporal Dithering", *SID 93 Digest*, pp. 155–158, 1993.

[3] C. Gotsman, "Halftoning of Image Sequences", *The Visual Computer*, **9**, pp. 255–266, 1993.

[4] M. T. Orchard and C. A. Bouman, "Color Quantization of Images", *IEEE Trans. Signal Proc.*, **39**(12), pp. 2677–2690, 1991.

[5] R. Balasubramanian, C. A. Bouman, and J. P. Allebach, "Sequential Scalar Quantization of Color Images", *J. Electronic Imaging*, **3**(1), pp. 45–59, 1994.

[6] R. Floyd and L. Steinberg, "An Adaptive Algorithm for Spatial Greyscale", *Proc. SID*, **17**(2), pp. 75–77.

[7] G. Goertzel and G. R. Thompson, "Digital Halftoning on the IBM 4250 Printer", *IBM J. Res. Develop.*, **31**(1), pp. 2–15, 1987.

[8] D. H. Kelly, "Motion and Vision. I. Stabilized Images of Stationary Gratings", *J. Opt. Soc. of Am.*, **69**(9), pp. 1266–1274, 1979.

[9] D. H. Kelly, "Motion and Vision. II. Stabilized Spatio-temporal Threshold Surface", *J. Opt. Soc. of Am.*, **69**(10), pp. 1340–1349, 1979.

[10] D. H. Kelly, "Visual Processing of Moving Stimuli", *J. Opt. Soc. of Am. A*, **2**(2), pp. 216–225, 1985.

[11] G. J. C. van der Horst and M. Bouman, "Spatio-temporal Chromaticity Discrimination", *J. Opt. Soc. of Am.*, **59**(11), pp. 1482–1488, 1969.

[12] F. L. van Nes, J. J. Koenderink, H. Nas, and M. Bouman, "Spatio-temporal Modulation Transfer in the Human Eye", *J. Opt. Soc. of Am.*, **57**(9), pp. 1082–1088, 1967.

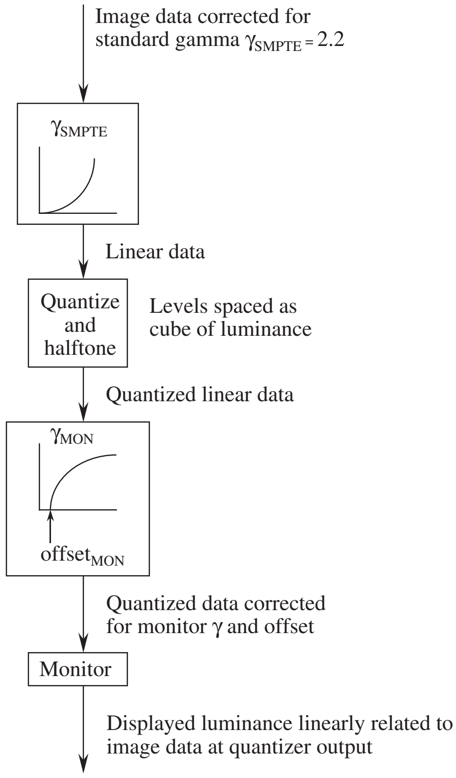


Figure 1: Flow diagram of the quantization process.

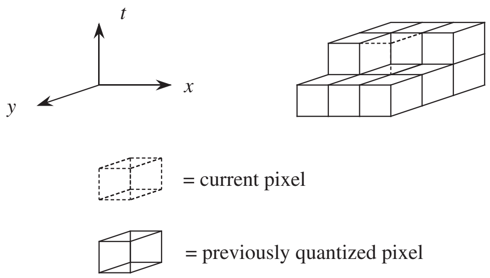


Figure 2: Region of support for the error diffusion filter.

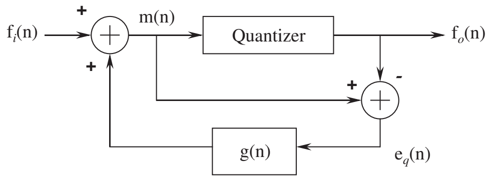


Figure 3: Block diagram of the error diffusion algorithm.

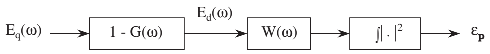


Figure 4: Display error is shaped by filter frequency response $G(\omega)$.

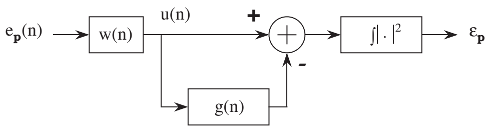


Figure 5: The filter $g(n)$ acts as a linear predictor for the random process $u(n)$.

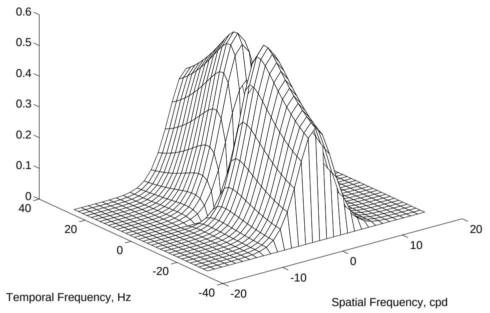
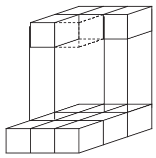


Figure 6: Spatiotemporal contrast sensitivity for the luminance channel.



30 Hz

-0.7110	0.8363	0.0067
0.8462		

-0.0419	0.0516	-0.0053
0.0521	-0.0640	0.0064
-0.0056	0.0064	0.0000

60 Hz

-0.7118	0.8359	0.0077
0.8471		

0.0947	-0.1114	0.0005
-0.1127	0.1320	0.0024
0.0013	0.0024	-0.0123

Figure 7: Luminance channel filter coefficients for 30 and 60 Hz.

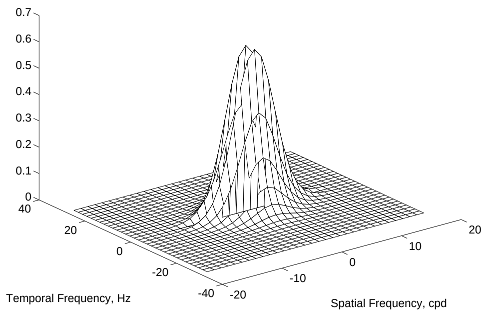
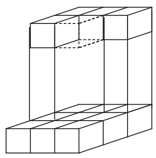


Figure 8: Spatiotemporal contrast sensitivity for the chrominance channel.



Chrominance

-0.7437	0.8469	0.0148
0.8697		

0.6814	-0.7779	-0.0088
-0.7984	0.9196	-0.0036
-0.0046	-0.0037	-0.0036

Luminance

-0.7118	0.8359	0.0077
0.8471		

0.0947	-0.1114	0.0005
-0.1127	0.1320	0.0024
0.0013	0.0024	-0.0123

Figure 9: Chrominance and luminance channel filter coefficients for 60 Hz.

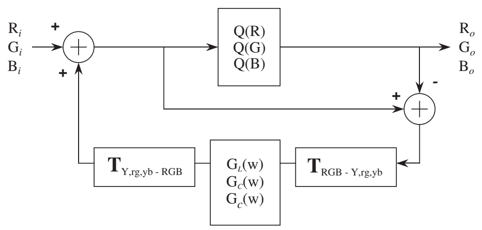


Figure 10: Block diagram of the color error diffusion algorithm.