EFFICIENT IMPLEMENTATION OF FOVEATION FILTERING

Sanghoon Lee, Alan C. Bovik and Brian L. Evans

Laboratory for Image and Video Engineering
Department of Electrical and Computer Engineering
The University of Texas at Austin, Austin, TX 78712-1084 USA
{slee,bovik,bevans}@ece.utexas.edu
http://signal.ece.utexas.edu

ABSTRACT

In the human eye, the maximum detectable frequency is a function of the photoreceptor density and eccentricity (visual angle). The maximum detectable frequency is obtained by using an exponential model of the spatial sampling grid on the fovea in the retina and choosing the model parameters such that the human eye cannot discriminate the foveated image from the original when assuming a given fixation point in the image being observed. It can also be used as the cutoff frequency for obtaining a foveation-filtered image, which is a nonuniform resolution image with spatial resolution at its peak near the fixation/foveation point and falling off away from this point. Foveation filtering can be implemented using a bank of lowpass filters. In order to reduce the number of operations, we use separable, even-symmetric or circularly symmetric lowpass filters. The number of taps in each filter is adaptively changed according to the magnitude of the cutoff frequency. The overall design is amenable to incorporation into implementations of foveated H.263 compression algorithms on digital signal processors.

1. INTRODUCTION

When the human eye focuses on a surface point, a variable-resolution image is transmitted into the brain via the retinal photoreceptors. The density of cones and ganglion cells, which is a major parameter of visual resolution, exponentially decays with respect to the *fovea*. Where the photoreceptors are widely spaced, the visible bandwidth is so narrow that the human eye cannot discriminate high frequency surface features. Where the photoreceptors are spaced close to one another, the surface details become visible. Hence, if a foveated image is artificially created by removing the undetectable frequencies of an original image (by presupposing a point of foveation), then the foveated image will appear the same as the original image.

A foveated image may be interpreted as having spatially varying local bandwidth. Once local spatial frequency information has been obtained relative to a foveation point, a variable-resolution image having the same information content as a foveated image may be obtained by applying a *foveation filter*, which consists of a bank of low pass filters having variable cutoff frequencies. The data reduction has allowed us to demonstrate high performance video compression with reduced computational redundancy for motion estimation or transmission delay for visual communications [1, 2, 3].

In this paper, we introduce a human visual model using an exponential function of spatial sampling density based on the photoreceptor density and viewing geometry. The model is used to obtain the local bandwidth of foveated images. The foveated image is obtained by using a filter bank where the local bandwidth are used for the cutoff frequencies. In order to reduce the computational complexity, symmetric filters are utilized for real-time processing.

2. LOCAL BANDWIDTH ACQUISITION

After obtaining local bandwidth using human visual modeling, we use the local bandwidth for the cutoff frequencies of the filter bank. In order to quantify the distance from the fovea to any other point on the retina, define the *eccentricity* to be the angle between the fovea and the point with respect to the nodal point of the optics in the human eye. Given the eccentricity, any point in the image formed on the retina(foveated image) can be mapped into a point in the original image.

In order to obtain a position varying band limited filter, the local bandwidth must be calculated based on the eccentricities of pixels on the input image. In Fig. 1, the viewing parameters are shown for a standard viewing geometry: the image size i_p (pixels), the image size i_d (units of length), the distance v_d from the viewer to the image (units of length), and the distance $d_{\mathbf{x}}$ from the foveation point in the image to a pixel at $\mathbf{x} = (x_1, x_2)$ (pixels). The eccentricity $e_{\mathbf{x}}$(degrees) is calculated by the Euclidean distance from the foveation point and is given at a point $\mathbf{x}$ by

$$e_{\mathbf{x}} = \tan^{-1}\left(\frac{i_d d_{\mathbf{x}}}{v_d i_p}\right). \tag{1}$$

Since the density of cones is a dominant factor for determining human visual resolution, it might be possible to derive the detectable frequency from the spatial mosaic of cones. For example, two adjacent cone photoreceptors are spaced about 2.5 μm at the fovea and the distance from the nodal point to the retina is 17 mm. Using a trigonometric relationship, the cone density is calculated to be 119 (per degree) at the fovea. According to the sampling theorem, the signal whose bandwidth is less than the half of the sampling rate can be reconstructed. The local bandwidth $f_{p_{\mathbf{x}}}$(cycles/pixel) in the discrete domain at the position $\mathbf{x}$=(x_1,x_2) is also derived from the maximum detectable frequency according to the viewing distance v_d to the image. However, the obtained frequency is so high that the foveated image is the same as the original image over a large area. The human eye cannot distinguish an appropriate foveated image from an original image even if the frequency is less than the frequency obtained by using the sampling density of cone photoreceptors. Therefore, we measure the maximum detectable frequency $f_{d_{\mathbf{x}}}$(cycles/degree) at $e_{\mathbf{x}}$ by using the model

$$f_{d_{\mathbf{x}}} = \frac{\gamma}{e_{\mathbf{x}} + \eta} - \zeta \tag{2}$$

where γ, η and ζ are parameters that control the spatial frequency decay. Experimentally, the parameters γ, η and ζ are chosen such that the human eye cannot discriminate the foveated image from the original image for a given fixation point. Fig. 2 shows the detectable frequencies $f_{d_{\mathbf{x}}}$ based on the photoreceptor density and the human visual model where $\gamma = 18$, $\eta = 0.2$, $\zeta = 0.0$, $f_{min} = 0.07$, $v_d = 30$ cm and $i_d = 9$ cm. We approximate $f_{d_{\mathbf{x}}}$ by using an exponential model that has the same shape as the photoreceptor density profile.

Since a position in the image is measured in pixels, a conversion factor from pixels to degrees must be calculated. To obtain the relationship between pixels and degrees of visual angle, we use a geometric approximation. Suppose each pixel forms a square with the length of each side $\epsilon = \frac{i_d}{i_p}$. Then, the conversion factor $\beta_{\mathbf{x}}$ is approximately

$$\beta_{\mathbf{x}} = tan^{-1}(\frac{\epsilon}{2v_d} + \frac{d_{\mathbf{x}}}{v_d}) - tan^{-1}(\frac{-\epsilon}{2v_d} + \frac{d_{\mathbf{x}}}{v_d}). \tag{3}$$

In (3), $\beta_{\mathbf{x}}$ is maximized at $d_{\mathbf{x}} = 0$ (the foveation point).

The maximum detectable frequency is used to design the local foveation filter. Suppose that the position $\mathbf{x}$ corresponds to the n^{th} pixels in the image. Then, $f_{p_{\mathbf{x}}} = f_{p_n}$. Since the normalized frequency of a signal in the discrete domain is less than 0.5, it is thresholded below at a minimum value of 0.5 so that

$$f_{p_n} = \min[\beta_{\mathbf{x}} f_{d_{\mathbf{x}}}, 0.5]$$

in cycles/pixel. If there exists a lower bound on the local bandwidth like the distribution of the photoreceptors, f_{p_n} is also thresholded by the minimum bandwidth f_{min} (here 0.7). Hence,

$$f_{p_n} = \max[f_{p_n}, f_{min}]$$

Fig. 3 shows the obtained local bandwidth f_{p_n} from $f_{d_{\mathbf{x}}}$.

3. EFFICIENT IMPLEMENTATION

To indicate the n^{th} pixel, we use the subscript n or the position vector $\mathbf{n} = (n_1, n_2)$. From the discrete original image $I(\mathbf{m})$, the foveated image $\tilde{I}(\mathbf{n})$ can be obtained by

$$\tilde{I}(\mathbf{n}) = L^*[I(\mathbf{m}), f_{p_n}] \tag{4}$$

where $\mathbf{m} = (m_1, m_2)$ is position vector and $L^*(., f_{p_n})$ is an ideal lowpass filter with the cutoff frequency f_{p_n}.

3.1. Filter design error

According to Parseval's theorem, the total energy in a discrete signal is equal to the integral of the energy-density spectrum. Let $h^*(i)$ be an ideal low pass filter with the cutoff frequency ω_c and $H^*(e^{i\omega})$ be the Fourier transform of $h^*(i)$. Then,

$$\sum_{i=-\infty}^{\infty} h^{*2}(i) = \frac{1}{2\pi}\int_{-\pi}^{\pi} |H^*(e^{i\omega})|^2 d\omega = \omega_c/\pi$$

where

$$H^*(e^{i\omega}) = \begin{cases} 1, & |\omega| < \omega_c \\ 0, & \omega_c < |\omega| \leq \pi. \end{cases}$$

The impulse response of $h^*(i)$ becomes

$$h^*(i) = \frac{\omega_c}{\pi} sinc(\omega_c i) \quad -\infty \leq i \leq \infty.$$

If ω_c is changed according to each pixel such as $w_c = 2\pi f_{p_n}$, then $H^*(., f_{p_n})$ becomes the ideal low pass filter $L^*(., f_{p_n})$ for creating foveated images.

In a practical implementation, the filter length is finite. Let $h(i)$ and $H(e^{i\omega})$ be the Fourier transform pair of a low pass filter with the filter length N. Then, the total energy of the error signal between $h^*(i)$ and $h(i)$ becomes

$$\begin{aligned} e &= \sum_{i=-\infty}^{\infty} [h^*(i) - h(i)]^2 \qquad (5) \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} |H^*(e^{i\omega}) - H(e^{i\omega})|^2 d\omega \end{aligned}$$

In order to minimize e, $h(i)$ must be

$$h(i) = \begin{cases} h^*(i), & \text{if } -N/2 \le i \le N/2 \\ 0, & \text{otherwise} \end{cases}$$

The error ratio τ relative to the total energy becomes

$$\tau = \frac{e\pi}{\omega_c}$$

Given e, τ is inversely proportional to the cutoff frequency ω_c. As ω_c decreases, the value N must increase to have a small error ratio. Decide a minimum N while τ is less than a constant. Then, the filter length varies according to position with reference to the fixed value τ because the local bandwidth is a function of position. Thus, the filter length N becomes a position dependent variable N_n.

3.2. Separable even symmetric filters

For smoothing the frequency response of the low pass filter, the Hamming window $w(i_1, i_2)$ is employed. Then, the designed foveation filter $l(i_1, i_2)$ can be expressed by $l(i_1, i_2) = h(i_1, i_2)\ w(i_1, i_2)$. Since the cutoff frequency depends on position, $l(.)$ is also a function of n such as $l_n(.)$. Then, the foveated image can be obtained by

$$\tilde{I}(n_1, n_2) = \sum_{i_1=-N_n/2}^{N_n/2} \sum_{i_2=-N_n/2}^{N_n/2} I(n_1 - i_1, n_2 - i_2) l_n(i_1, i_2) \qquad (6)$$

To reduce the number of multiplications, we use a separable even symmetric low-pass filter $l(i_1, i_2)$:

$$l_n(i_1, i_2) = \begin{cases} l_n(i_1) l_n(i_2), & \text{if} - N_n/2 \le i_1,\ i_2 \le N_n/2 \\ 0, & \text{otherwise} \end{cases}$$

and $l_n(i_1) = l_n(-i_1)$. Then,

$$\tilde{I}(n_1, n_2) = \sum_{i_1} l_n(i_1) \sum_{i_2} I(n_1 - i_1, n_2 - i_2) l_n(i_2)$$

For the separable even symmetric filter, the number of operations is reduced to $2 * (N_n/2 + 1)$ for additions and multiplications for each pixel.

3.3. Circularly Symmetric Filters

Using circularly symmetric filters, we can obtain more symmetric frequency response associated with the local bandwidth. To reduce the number of multiplications, we use the octal symmetry of circularly symmetric filters $l_n(i_1, i_2) = l_n(\pm i_1, \pm i_2) = l_n(\pm i_2, \pm i_1)$. Then, (6) becomes

$$\tilde{I}(n_1, n_2) = \sum_{\substack{i_1,\, i_2 \ne 0 \\ i_1 = 0, 1, \ldots, N_n/2 \\ i_2 = 0, 1, \ldots, i_1}} l_n(i_1, i_2) \Big[I(n_1 \pm i_1, n_2 \pm i_2) + I(n_1 \pm i_2, n_2 \pm i_1) \Big] + I(n_1, n_2) l_n(0, 0) \qquad (7)$$

Using (7), we can compute the number of additions and multiplications that are required for implementing this circularly symmetric filter. The number of operations is $7\ (N_n/2+1)(N_n/2+2)/2-6$ for additions and $(N_n/2+1)(N_n/2+2)/2-1$ for multiplications. Thus, due to the octal symmetry, the number of multiplications required for implementing circular symmetric filters has been reduced by an approximate factor of 8 for $N_n \to \infty$.

4. SIMULATION RESULTS

Suppose that the distribution of foveation points is a Gaussian function according to the centered point in the image. The probability density function is

$$p(n_1, n_2) = \alpha e^{-2\pi^2 \sigma^2 r^2 / i_p^2} \qquad (8)$$

where $r = \sqrt{(n_1 - i_p/2)^2 + (n_2 - i_p/2)^2}$, $0 \le n_1,\ n_2 \le i_p - 1$, and α is a constant. Here, we assume that the number of pixels in horizontal/vertical line is i_p. Usually the half-peak radius r_c is a factor to select σ by

$$e^{-2\pi^2 \sigma^2 r_c^2 / i_p^2} = 1/2 \qquad (9)$$

giving

$$\sigma = \left(\frac{i_p}{\pi r_c}\right) \sqrt{log\sqrt{2}} \approx 0.19 \left(\frac{i_p}{r_c}\right). \qquad (10)$$

The value α is decided in order that the sum of $p(n_1, n_2)$ for all pixels is equal to 1. Then, the average number of operations is obtained by

$$\bar{O}_T = \sum_{n_1=1}^{i_p} \sum_{n_2=1}^{i_p} p(n_1, n_2) O(N_n) \qquad (11)$$

where $O(N_n)$ is the total operation when the foveation point is at the n^{th} pixel. Table 1 shows the average number of multiplications at each pixel. Fig. 4 shows the original *lena* image, and Figs. 5 - 7 show the foveated images according to the filter length and filter coefficients. In the results, it is demonstrated that the visual qualities of foveated images are similar with each other while reducing the computational complexity.

5. CONCLUSIONS

In this paper, we implement foveation filtering using low pass filters with continuously varying cutoff frequencies. After obtaining local bandwidth based on human visual modeling, we use the local bandwidth for the cutoff frequencies of the filter bank. For a real-time implementation, we reduce the number of operations by using separable even symmetric or circularly symmetric filters. We further reduce the computational complexity by changing the filter length (number of taps) according to the magnitude of cutoff frequencies. We obtain foveated images with good visual quality while reducing computational complexity. Foveation-filtered images and video streams are anticipated to have application in low bitrate image and video communication systems, since foveation-filtered images can be compressed significantly more than standard compressed images, while maintaining compliance with standards, e.g. MPEG-IV and H.263. We expect to demonstrate these methods on C62X digital signal processors.

6. REFERENCES

[1] S. Lee, M. S. Pattichis, and A. C. Bovik, "Optimal rate control for real-time, low bitrate foveated video coding," *IEEE Trans. Image Processing, submitted.*

[2] S. Lee and A. C. Bovik, "Motion estimation and compensation for foveated video," in *Proc. IEEE Int. Conf. on Image Processing*, (Kobe, Japan), Oct. 1999.

[3] S. Lee, A. C. Bovik, and Y. Y. Kim, "Low delay foveated visual communications over wireless channels," in *Proc. IEEE Int. Conf. on Image Processing*, (Kobe, Japan), Oct. 1999.

	separable even symm.			circularly symm.		
	τ			τ		
	0.15	0.1	0.05	0.15	0.1	0.05
σ=0.38	7.41	10.15	20.93	8.40	16.17	69.03
σ=0.57	7.21	9.84	20.08	7.96	15.24	63.94

Table 1: Average number of multiplications in foveation filtering.

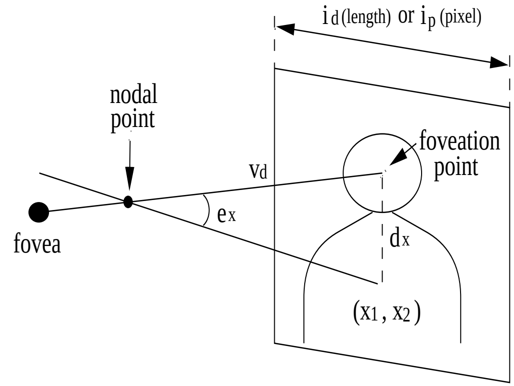

Figure 1: Parameter definition in the viewing geometry.

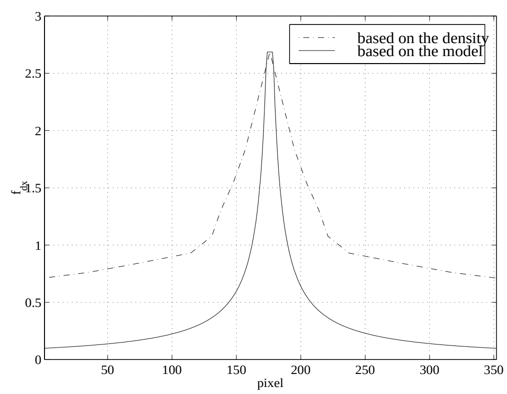

Figure 2: Human visual detectable frequencies derived based on the photoreceptor density and the human visual model.

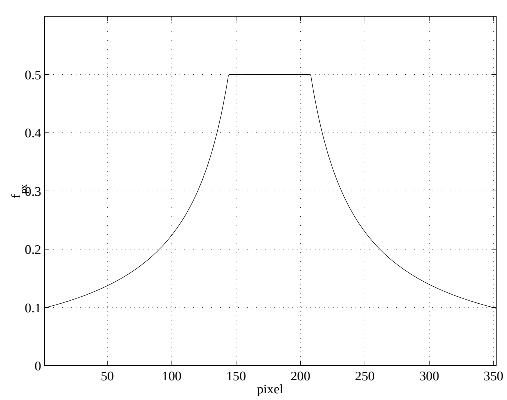

Figure 3: Local bandwidth (in cycles/pixel) which is also used as cutoff frequencies in the filter bank.

Figure 4: Original *lena* image

Figure 6: Foveated *lena* image using separable even symmetric filters with adaptive N_n and $\tau = 0.1$.

Figure 5: Foveated *lena* image using the circularly symmetric filters with $N = 31$.

Figure 7: Foveated *lena* image using circularly symmetric filters with adaptive N_n and $\tau = 0.1$.