A Wavelet Approach to Foveating Images

Ee-Chien Chang and Chee K. Yap*
Courant Institute of Mathematical Sciences
New York University
New York, NY 10012
U.S.A.

Abstract

Motivated by applications of foveated images in visualization, we introduce the **foveation transform** of an image. We study the basic properties of these transforms using the multiresolution framework of Mallat. We also consider practical methods of realizing such transforms. In particular, we introduce a new method for foveating images based on wavelets. Preliminary experimental results are shown.

1 Introduction

Conventional images have uniform resolution. **Foveated images** which have non-uniform resolution arise in biological vision. In figure 1 we show a foveated version of an uniform image. The process of going from an uniform image to figure 1 is called "foveation".

One of the most interesting forms of foveated images is based on the complex logarithm function. Such **logmap images** were studied by Rojer and Schwartz [10] and others. The complex logmap is a model consistent with empirical data on the mapping from primate retina to the visual cortex [11, 12]. This neuro-physiological model goes back to the pioneering (and the Nobel prize) work of Hubel and Wiesel [5, 6]. Rojer [9] demonstrated many favorable properties of such images. Perhaps the most striking fact is that the data density in such images grows logarithmically with the diameter of the visual field (as opposed to quadratically in the case of uniform images). Such low density images have been exploited in applications such as video phones [14]. In this paper, we use the term "foveated" to refer to any variable resolution image, not just those with the geometry of the logmap or having a definite "foveal region".

Visualization Applications. Foveated images have been exploited in computer vision, (e.g., [4, 13]), especially in the context of **active vision**[1, 2]. But they are also useful in **visualization**, although this aspect is less explored. For visualization, we must compensate for the loss of peripheral resolution in foveated images by rapidly presenting images that are foveated at points of user interest. We could use eye-tracking technology [7] to follow the eye and to present the corresponding foveated images. In principal, if the foveation uses the correct parameters and eye-tracking is good, a user should not be able to detect any difference from viewing a uniform image. A system with such properties has been reported by [8]. But in our visualization applications, we can dispense with eye-tracking — the visualizer may be counted upon to cooperate with the image server. For instance, the visualizer can indicate moving areas of interest in an image via a standard mouse input. Because of the rapid eye activity in normal visualization, we call this approach **active visualization**.

Thinwire Visualization. We describe one application, viz., in "thin-wire" models of computation. Here, we have a **server** who acts as an image-server and a **client** who wishes visualize images from the server. The images may be very large (e.g., a terrain or map) and have high resolution. The server has large computational resources but there is only a thin wire (such as the Internet) connecting the server and client. A basic problem in this scenario is the construction foveated images from uniform images. Our paper focuses on this problem.

2 Foveation Theory

There are various practical methods to foveate an image. We would like to provide the mathematical formulation of this process. This would allow us to evaluate the various methods.

As usual, one can begin with one-dimensional images or functions. Given an image $I(x)$, we would like to define a foveated image $I_0(x)$ A **foveation** is determined by a **weight function** and a **scaling function**,

$$w : \mathbb{R} \to \mathbb{R}_{\geq 0}, \quad s : \mathbb{R} \to \mathbb{R}.$$

We require the weight function to be positive except at finitely many values (so $w(x)^{-1} < \infty$ at all but finitely many x). The scaling function is analogous to the "father wavelet" usually denoted $\phi(x)$ in wavelet theory. We define

$$I_0(x) := \int I(t) C^{-1}(x) s\left(\frac{t-x}{w(x)}\right) dt,$$

where $C(x) := \|s\left(\frac{\cdot - x}{w(x)}\right)\|$. The **weighted translation** (or,

*`yap@cs.nyu.edu`, supported by an National Science Foundation Grant #CCR-9402464.

Figure 1: Blended image with 2 gaze points (on the extreme left and extreme right faces with rates $\frac{1}{20}$ and $\frac{1}{30}$ respectively). The image is obtained using the definition of foveation and blending of two standard weight functions.

Figure 2: Same as figure 1 but is is obtained using wavelet foveation.

w-translation) of s by x is defined to be

$$s_x(t) := C^{-1}(x) s\left(\frac{t-x}{w(x)}\right). \quad (1)$$

Now the foveated image could be rewritten as

$$I_0(x) = \langle I, s_x \rangle.$$

A **standard weight function** is one of the form

$$w_{\text{std}}(t) = w_{\alpha,\beta,\gamma}(t) := \alpha|t-\gamma| + \beta,$$

where α, β, γ are fixed constants and $\alpha, \beta \geq 0$. We call α the **rate**, β the **foveal resolution** and γ the **gaze point**. Intuitively, this weight function defines the point $t = \gamma$ as the fovea with maximum resolution β. The resolution decreases linearly with the distance from the gaze point, at the rate α.

We can obtain new weight functions by combining several given weight functions. The following definition corresponds to an effective method for blending images: given weight functions w_1 and w_2, their **blended weight function** w is the pointwise minimum of the two functions: $w(x) = \min\{w_1(x), w_2(x)\}$.

3 Weighted Distance

For a function I, we want to measure how far I is from the foveation I_0. It is not useful to use the usual $\|I - I_0\|$ norm as we want to place less emphasis on $I(x) - I_0(x)$ when x is far away from the "foveal region". Define for a function f,

$$\|f\|_w = \left(\int (w(t))^{-1} |f(t)|^2 \, dt\right)^{1/2}. \quad (2)$$

Then the w-**distance** between f and g is defined to be $d_w(f,g) := \|f-g\|_w$. Let $L(t)$ be a function such that $dL(t) = \frac{dt}{w(t)}$. If $L^{-1}(t)$ is well-defined then

$$\int |f(t)|^2 (w(t))^{-1} dt \quad = \quad \int |f(L^{-1}(y))|^2 dy. \quad (3)$$

$L(t)$ and $L^{-1}(t)$ for the standard weight functions with $\beta = 1$ is are given by

$$L(t) = \text{sg}(t) \cdot \ln(\alpha|t| + 1), \qquad L^{-1}(t) = \text{sg}(t)(e^{\alpha|t|} - 1).$$

4 Multiresolution Foveation

For any weight function w and $n \in \mathbb{Z}$, let $w_n(t) := 2^{-n} w(t)$. (An alternative definition is $w_n(t) := w(2^{-n}t)$). Define I_n to be the foveation using the weight function w_n. Then $\{I_n\}$ is now a sequence of increasingly better foveation as n increases.

5 Practical Foveation Methods

Computing I_0 from its definition is computationally inefficient. We consider practical methods to approximate I_0. There are two general approaches. The first approach is based on **sampling**. Previous papers on foveation can be put under this approach. This is computationally efficient (usually implemented with lookup tables), but lack flexibility when different weight functions are required. In this paper we introduce a second approach. It exploits any multiresolution representation of an image of I: to create a foveated image from I, we 'cut' the necessary pieces from each level of the multiresolution representation and 'glue' them together. There is considerable freedom in choosing the multiresolution representation (e.g., different kinds of wavelet transforms) and in how we cut and paste.

5.1 Foveation via Wavelets

We give an approximation of foveation using wavelets. Let ψ be the mother wavelet and ϕ be the corresponding father wavelet. Let $\psi_{m,n}$ be the translated and dilated version of ψ. Specifically, $\psi_{m,n}(t) = \psi(2^{-m}t - n)$. Assume that $\{\psi_{m,n}\}$ form an orthonormal basis for $L_2(\mathbb{R})$. The approximation is as follows.

$$I^{\text{wav}} = \sum_{(m,n) \in B(w)} \langle \psi_{m,n}, I \rangle \, \psi_{m,n}(x),$$

where $B(w)$ is a discrete set depending on the weight function w,

$$B(w) = \{(m,n) | 2^m \geq w(n2^m)\}.$$

For the standard weight $w = \alpha|t| + \beta$, the set $B(w)$ is $\{(m,n)|2^m \geq \alpha|n|2^m + \beta\}$, which is $\{(m,n)|2^{-m}c_1 + |n| \leq r\}$

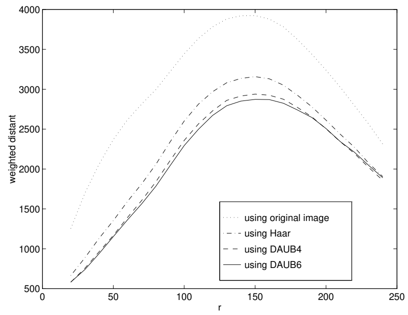

Figure 3: The weighted distance $d_w(I, I_0)$ and $d_w(I^{\text{wav}}, I_0)$ where I^{wav} is computed using different wavelets. The weight function used is $w(t) = \frac{1}{r}\|t\|_\infty$.

where $c_1 = \alpha^{-1}\beta$ and $r = \alpha^{-1}$. This means that we just have to take at most r coefficients from each level. If the foveal resolution β is greater than 2^m, then no coefficients from level m is in $B(w)$ (m could be a negative integer). This simple structure is useful for fast reconstruction.

We use the weighted distance to measure the quality of approximations using different wavelets. Figure 3(a) shows experimental results of this comparison. The two dimensional weighted distance is

$$d_w(f, g) := \int\int w(x, y)^{-2}\|f(x, y) - g(x, y)\|^2 dx\ dy.$$

5.2 Bandwidth reduction

Our techniques show promise for thin-wire visualization. Figure 4 illustrates experimental results on the reduction in number of coefficients and number of bytes (note that a coefficient may take more than one byte) needed to represent foveated images at various rates. The number of bytes is computed after a lossless compression. The original image used is the uniform version of figure 1.

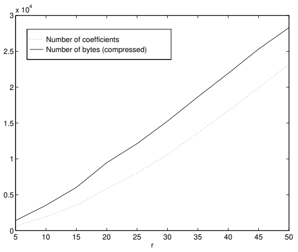

Figure 4: Size of original image is 375×469 pixels with 256 gray levels. When $r = 20$, on average, the image uses 0.43 bit per pixel. When $r = 50$, the image uses 1.28 bit per pixel.

6 Further Work

We are currently looking at the mathematical analysis of multiresolution foveation and comparison of various foveation approaches using our formulation. On the practical side, we are implementing active visualization of a large (for example 4000 by 4000 pixels) image across a network. We are also looking into generalization to 3-d, that is, volume data.

References

[1] J. Aloimonos, I. Weiss, and A. Bandyopadhyay. Active vision. *Intl. J. Computer Vision*, 2:333–356, 1988.

[2] R. Bajcsy. Active perception. *IEEE Proceedings*, 76(8):996–1005, 1988.

[3] Benjamin B. Bederson. *A miniature space-variant active vision system: Cortex-I.* PhD thesis, Computer Science, Courant Institute, New York University, 1992.

[4] C. Colombo, M. Rucci, and P. Dario. Integrating selective attention and space-variant sensing in machine vision. In Jorge L.C. Sanz, editor, *Image Technology: Advances in Image Processing, Multimedia and Machine Vision*, pages 109–128. Springer, 1996.

[5] D.H.Hubel and T.N.Wiesel. Receptive fields and functional architecture of monkey striate cortex. *J. Physiol. (Lond.)*, 195:215–243, 1968.

[6] D.H.Hubel and T.N.Wiesel. Sequence regularity and geometry of orientation columns in the monkey striate cortex. *J. Comp.Neurol.*, 158:267–293, 1974.

[7] R. Jacob. Emerging user-interface media: potentials and challenges – eye tracking. In *Tutorial, SIGGRAPH '89*, 1989.

[8] Philip Kortum and Wilson S. Geisler. Implementation of a foveated image coding system for image bandwidth reduction. In *Human Vision and Electronic Imaging, SPIE Proceedings Vol. 2657*, pages 350–360, 1996.

[9] Alan S. Rojer. *Space-variant computer vision with a complex-logarithmic sensor geometry.* PhD thesis, Computer Science, Courant Institute, New York University, 1989.

[10] Alan S. Rojer and E.L. Schwartz. Design considerations for a space-variant visual sensor with a complex-logarithmic sensor geometry. In *10th ICPR*, volume 2, pages 278–285, 1990.

[11] E.L. Schwartz. The development of specific visual projections in the monkey and the goldfish: outline of a geometric theory of receptotopic structure. *J.Theoret.Biol.*, 69:655–685, 1977.

[12] E.L. Schwartz. Spatial mapping in primate sensory projection: analytic structure and relevance to perception. *Biological Cybernetics*, 25:181–194, 1977.

[13] E.L. Schwartz, D.N. Greve, and G. Bonmassar. Space-variant active vision: Definition, overview and examples. *NeurNet*, 8(7–8):1297–1308, 1995.

[14] R.S. Wallace, B.B. Bederson, and E.L. Schwartz. Voice bandwidth visual communication through logmaps: The Telecortex. In *Workshop on Applications of Computer Vision*, pages 4–10, 1992.