SPIE Proceedings, 2657, 350-360, 1996

# Implementation of a foveated image coding system for image bandwidth reduction

Philip Kortum and Wilson Geisler

University of Texas Center for Vision and Image Sciences. Austin, Texas 78712

## ABSTRACT

We have developed a preliminary version of a foveated imaging system, implemented on a general purpose computer, which greatly reduces the transmission bandwidth of images. The system is based on the fact that the spatial resolution of the human eye is space variant, decreasing with increasing eccentricity from the point of gaze. By taking advantage of this fact, it is possible to create an image that is almost perceptually indistinguishable from a constant resolution image, but requires substantially less information to code it. This is accomplished by degrading the resolution of the image so that it matches the space-variant degradation in the resolution of the human eye. Eye movements are recorded so that the high resolution region of the image can be kept aligned with the high resolution region of the human visual system. This system has demonstrated that significant reductions in bandwidth can be achieved while still maintaining access to high detail at any point in an image. The system has been tested using 256x256 8 bit gray scale images with 20° fields-of-view and eye-movement update rates of 30 Hz (display refresh was 60 Hz). Users of the system have reported minimal perceptual artifacts at bandwidth reductions of up to 94.7% (18.8 times reduction)

KEYWORDS: foveation, field-of-view, gaze contingent, area-of-interest, eye movements, image compression

## 1.0 INTRODUCTION

The human visual system functions as a unique space-variant sensor system, providing detailed information only at the point of gaze, coding progressively less information farther from this point. This implementation is an efficient way for the visual system to perform its task with limited resources; processing power can be devoted to the area of interest and fewer sensors (i.e. ganglion cells and photoreceptors) are required in the sensor array (the eye). Remarkably, our perception does *not* reflect this scheme. We perceive the world as a single high resolution image, moving our eyes to regions of interest, rarely noticing the fact that we have severely degraded resolution in our peripheral visual field.

The same constraints that make this space-variant resolution coding scheme attractive for the human visual system also make it attractive for image compression. The goal in real-time image compression is analogous to that of the human visual system; utilization of limited resources, in this case transmission bandwidth and computer processing power, in an optimum fashion. Transmitted images typically have a constant

resolution structure across the whole image. This means that high resolution information must be sent for the entire image, even though the human visual system will use that high resolution information only at the current point of interest. By matching the information content of the image to the information processing capabilities of the human visual system, significant reductions in bandwidth can be realized, provided the point-of-gaze of the eye is known.

Recently, there has been substantial interest in foveated displays. The US Department of Defense has studied and used so-called "area-of-interest" (AOI) displays in flight simulators. These foveation schemes typically consist of only 2 or 3 resolution areas (rather than continuous resolution degradation) and the center area of high resolution, the AOI, is often quite large, usually between 18° and 40° (see, for example Howard, 1989, Warner, Sefoss and Hubbard, 1993). Other researchers have investigated continuous variable resolution methods using log polar mapping (Weiman, 1990, Juday and Fisher, 1989, Benderson, Wallace and Schwartz, 1992). Log polar mapping is particularly advantageous when rotation and zoom invariance are required, but their implementations have necessitated special purpose hardware for real-time operation. We have developed a preliminary version of a system that accomplishes real-time foveated image compression and display using a square symmetric Cartesian resolution structure, implemented on a general purpose computer processor.

## 2.0 SYSTEM OPERATION

This Cartesian coordinate based Foveated Imaging System (FIS) has been implemented in C (Portland Group PGCC) for execution on an ALACRON i860 processor. This platform has been used merely as a testbed; preliminary tests have shown that a 7 to 10 fold *increase* in performance can be achieved when implemented on a Pentium 90 MHz processor. Figure 1 illustrates the general function of the FIS. The system is initialized and the user is queried for a number of required parameters (half resolution constant, desired visual field of the display, eye movement update threshold). Using these parameters, a space variant arrangement of SuperPixels ( referred to as the ResolutionGrid) is then calculated and stored for display. A SuperPixel is a collection of screen pixels ( where a screen pixel is defined as ScreenPixelSize (degrees) = 60 x display size (degrees) / image size in pixels) that have been assigned the same gray level value. The user is then prompted through an eye tracking calibration procedure in order to account for variations in head position at setup. The system then enters a closed loop mode, in which eye position is determined and compared with the last known eye position. If the eye has not moved more than some predetermined amount (specified during initialization), the pixel averaging subroutine is executed and eye position is checked again. However, if the eye has moved more than this threshold amount, then a new eye fixation location is calculated and the pixel averaging subroutine is executed, creating the gray levels for each of the SuperPixels in the ResolutionGrid. These

SuperPixels are then sent to the display device, at which time eye position is checked again.

Only the closed loop portion of the program is required to run in real time. Initialization, calibration and calculation of the space-variant resolution grid take place prior to image display. However, because of the simplicity of the resolution grid structure (which is described in greater detail below), it can also be re-calculated in real time, if desired.

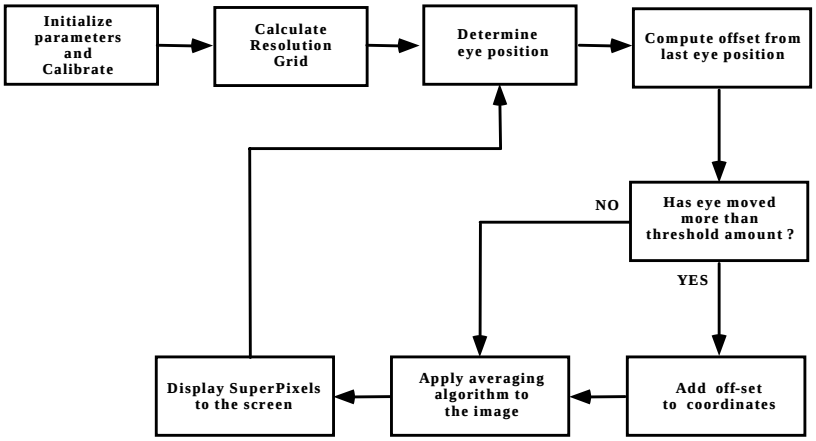



Figure 1: General flow diagram of the operation of the foveated imaging system.

## 2.1 Resolution fall-off calculations

The need for real time performance requires a resolution structure that results in high computational efficiency in algorithmic implementations. The square symmetric pattern, shown in Figure 2, is one such configuration. Because the resolution structure is specified in Cartesian coordinates, and each of the SuperPixels is square, pixel locations can be represented with a set of corner coordinates. This allows implementation of operations such as scaling and translation to occur using only addition.

Starting in the south-west corner of the north-east pixel in ring i (the pixel at location $x_i$, $y_i$), the size of a SuperPixel is calculated according to the formula,

$$W_i = \frac{W_0}{\sqrt{2}}\left(1 + \frac{\sqrt{x_i^2 + y_i^2}}{\varepsilon_2}\right) \qquad (1)$$

where $W_i$ is the size of the SuperPixels in ring i (in pixels), $W_o$ is the size of the central foveal SuperPixel (in pixels), $x_i$ and $y_i$ are the distances along a diagonal from the center of the screen (in degrees), and $e_2$ is the half-resolution constant, expressed in degrees. This function is based on available perceptual data and is also consistent with anatomical measurements in the human retina and visual cortex ( Wilson *et al*, 1990; Geisler and Banks, 1995; Wassle, *et al*, 1992). Specifically, when $e_2$ is between 0.8 and 1.2 the SuperPixel size is approximately proportional to the human resolution limit at each eccentricity. Thus, if $W_o$ is less than or equal to the foveal resolution limit then the foveated image will be *indistinguishable* from the original image (with proper fixation). If $W_o$ is greater than the foveal resolution limit then the foveated image *will* be distinguishable from the original image (note that because $W_o$ is a proportionality constant in equation (1) the SuperPixel size will be above the resolution limit by a constant factor at all eccentricities).

Once the size of the SuperPixel in the NE corner of ring i is determined a three pronged decision tree is entered in order to calculate the size of the remaining SuperPixels in the ring. This is necessary because an integer number of SuperPixels of size $W_i$ may not fit in the space delineated by the square symmetric ResolutionGrid. This means that, while all of the SuperPixels in a ring will have the same size in one direction ($W_i$), they may not have the same size in the other direction. The simplest situation (case 1) occurs when an integer number of SuperPixels of size $W_i$ *can* be accommodated in the specified space. The other two branches of the decision tree (cases 2 and 3) essentially conduct multiple bisections, putting the smallest SuperPixels in the center of the side, increasing SuperPixel size in a symmetric fashion towards the corners, where the SuperPixels are $W_i$ x $W_i$. Case 2 handles situations where only one reduced size SuperPixel is required. Case 3 handles situations in which multiple reduced size SuperPixels are required.

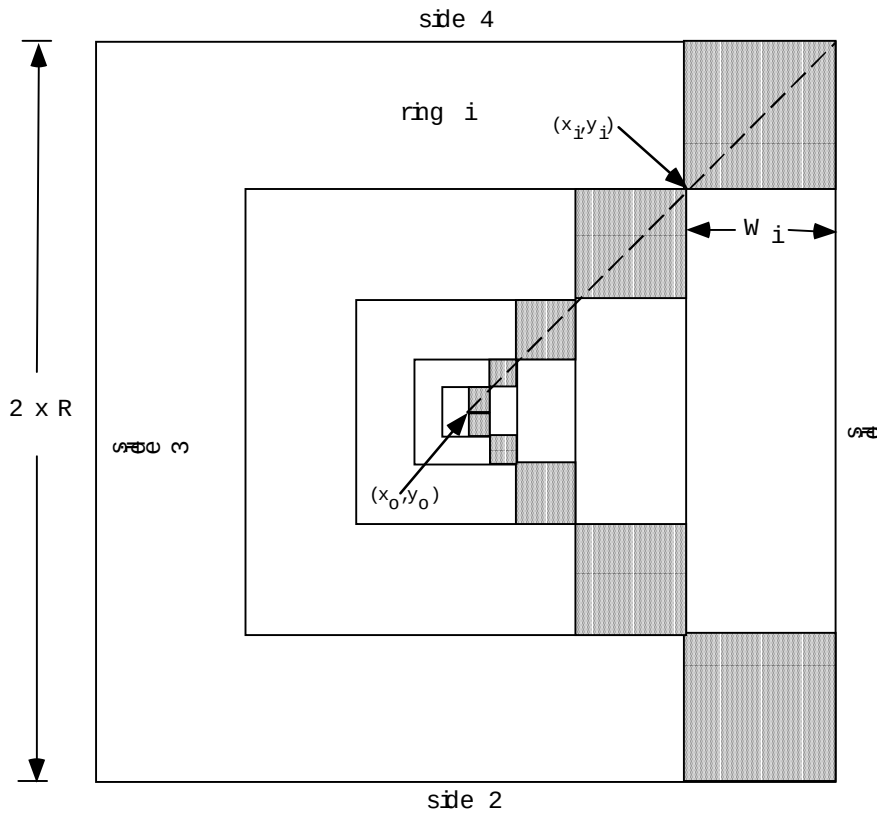



Figure 2: Foveated imaging system SuperPixel pattern arrangement, which is called the ResolutionGrid. SuperPixels of size $W_i$ are arranged in concentric rings (i) about the origin $(x_o,y_o)$. As described in the text, the ResolutionGrid is twice the size (4 times the area) of the viewable screen to allow for simple update as the result of an eye movement.

In order to increase the computational efficiency of the program, this three pronged decision calculation is only carried out for a single side (side 1, as labeled in Figure 2) of the pixel ring i. Each SuperPixel is represented by 4 numbers: the x and y locations of the lower left-hand (SW) and upper right hand (NE) corners. Because of the square symmetric SuperPixel pattern, the SuperPixel coordinates from the single computed side are simply moved through three 90 degree rotations to establish the coordinates of all the SuperPixels in the ring.

Upon completion of a ring, the program checks to see if the coordinates of the NE corner of the NE SuperPixel in the last calculated ring are greater than 2 times the resolution of the image (for reasons explained in detail in the description of the tracking subroutine); if not, the subroutine is run again. If so, the entire set of SuperPixel coordinates (the ResolutionGrid) is written to memory for later use. As mentioned earlier, since the ResolutionGrid is stored for later use, its calculation time does not affect the real time capability of the system.

Once the structure of the ResolutionGrid has been determined, a SuperPixel averaging subroutine averages the gray levels of each of the screen pixels that fall within a SuperPixel boundary and assigns the resulting average gray level to that SuperPixel. If the entire SuperPixel does not fall in the bounds of the display device (recall that the ResolutionGrid is twice the size of the viewable image) then the average includes only displayed pixels. Because the SuperPixel averaging subroutine takes place in real time, computational efficiency is important: therefore, SuperPixels of width 1 are excluded from the averaging subroutine, and their gray levels are passed directly to the display. Once the average gray level is determined for each SuperPixel, its value is added to the ResolutionGrid, which is then passed to the screen for display.

## 2.2 Gaze tracking

After the initial SuperPixel gray level assignment, the program enters a loop in which the position of the eye is measured and compared against the last measured eye position. If the eye has not moved more than the threshold amount (which is specified during the parameter initialization subroutine), the SuperPixel averaging subroutine is called, and the resulting averaged SuperPixels are displayed. This insures that, even with steady fixation, changes in the image (i.e. motion of an object or the video camera) will be reflected in the display. The subroutine then loops back and checks the position of the eye again . If, however, the eye has moved more than the threshold amount, several things happen. First, the amount of the movement (expressed in terms of pixels) is added to the each of the ResolutionGrid coordinates. The result is a change in the position of the high resolution region. Figure 3 shows an example of how the portion of the ResolutionGrid that is displayed changes as fixation changes; here a subject begins with center fixation, and moves his eyes towards the northeast corner of the display device. When this happens, the amount of his eye movements (in the x and y direction) are added to the current ResolutionGrid coordinates, causing the foveated region to offset the same amount. The result is an image that has highest resolution at the point of gaze. Initially calculating a ResolutionGrid that is twice the size of the viewable area (as previously shown in Figure 2) allows us to use this computationally efficient offset method to track eye position and update the display *without* having to recompute the ResolutionGrid each time the eye moves. Figure 4 illustrates how the offset works; adding the eye movement offset to the current ResolutionGrid coordinates is essentially the same as moving the ResolutionGrid to a position that coincides with current fixation location, while keeping the viewable screen in a fixed location. Since the ResolutionGrid is twice the size (4 times the area) of the viewable screen, recomputation of the ResolutionGrid is unnecessary because all eye positions in the viewable screen can be accounted for with a simple offset of the ResolutionGrid. Once the updated SuperPixel configuration on the viewable screen is determined, the SuperPixel averaging subroutine is called, and the resulting averaged SuperPixels are displayed.

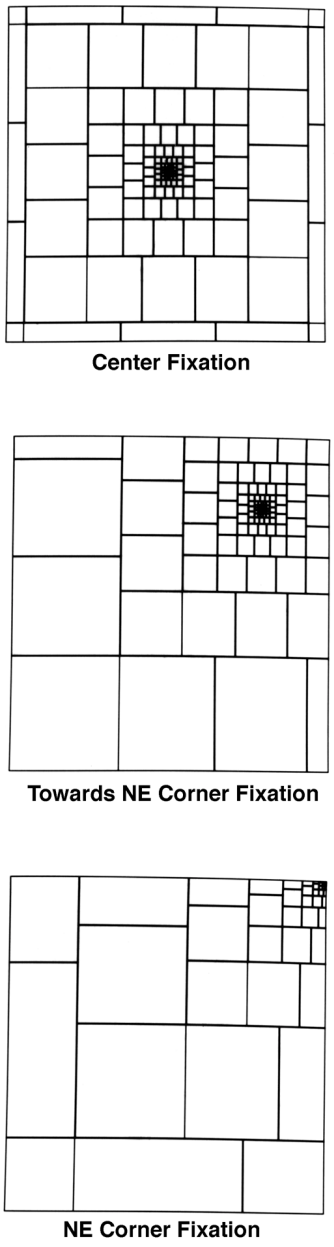



Figure 3: An example of how the portion of the ResolutionGrid that is displayed changes as fixation changes; here a subject begins with center fixation, and moves his eyes towards the northeast corner of the display device. As he does this, the amount of his eye movements (in the x and y direction) are added to the current ResolutionGrid coordinates, causing the foveated region to offset the same amount. Notice how the SuperPixels increase in size as they become further from the point of fixation.

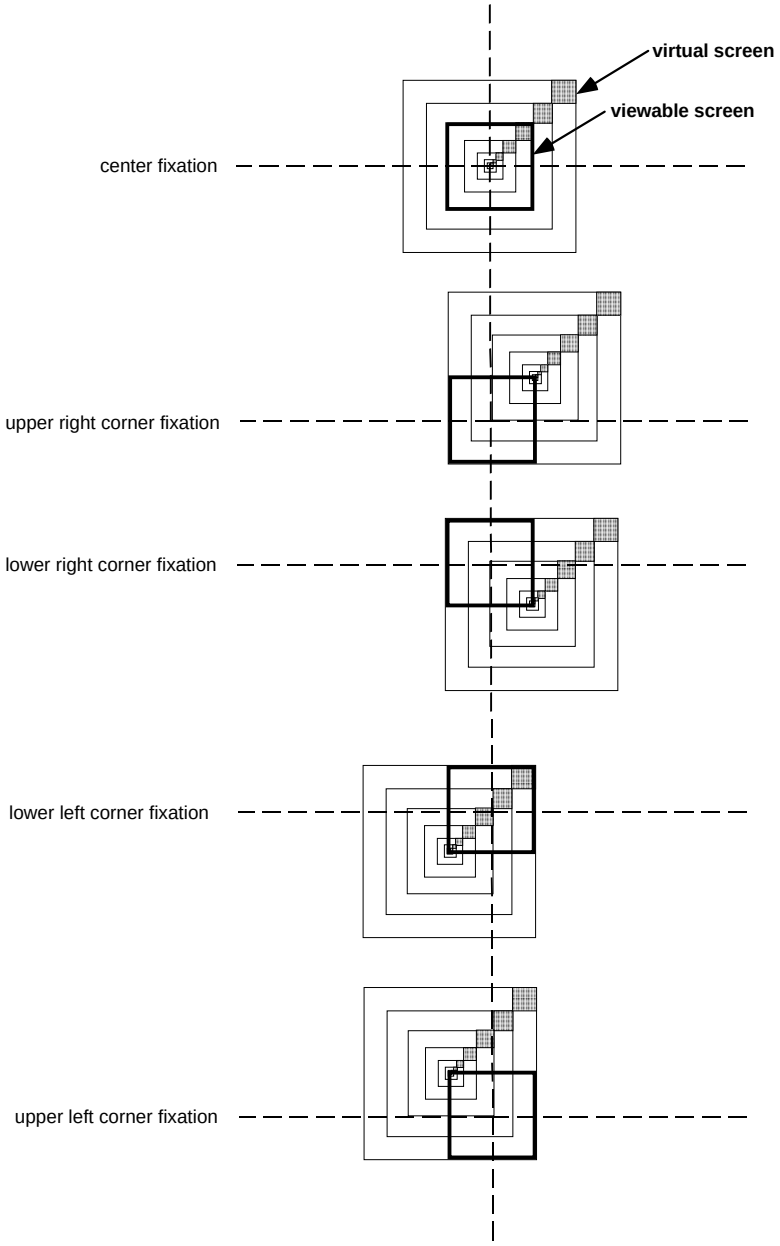



Figure 4: Calculating a ResolutionGrid that is twice the size of the viewable area allows a computationally efficient offset method to track eye position and update the display *without* having to recompute the ResolutionGrid each time the eye moves. The dark outline square is the viewable screen, with the remaining portion being the expanded ResolutionGrid, or so-called virtual screen. Adding the eye movement offset (the amount the eye has moved since the previous measurement) to the current ResolutionGrid coordinates is essentially the same as moving the ResolutionGrid to a position that coincides with current fixation location, while keeping the viewable screen in a fixed location. Since the ResolutionGrid is twice the size (4 times the area) of the viewable screen, recomputation of the Resolution Grid is unnecessary because all eye positions in

the viewable screen can be accounted for with a simple offset of the ResolutionGrid. Several extreme fixation locations are illustrated here to demonstrate this effect.

## 3.0 FIS PERCEPTUAL EVALUATION

The Foveated Imaging System has been evaluated for perceptual performance using a conventional 21 inch VDT. Three images, as shown in Figure 5, were used in the perceptual evaluation of the FIS. The images were 256x256 8 bit images, having a 20° field of view. These images were selected to test the perception of a number of probable image types: a letter chart (evaluation of visual clarity in a reading task), a natural environment scene (evaluation of cluttered, high detail images) and a face (evaluation of telecommunication systems). User reports of the subjective quality of the display were used in the evaluation. More detailed psychophysical performance measurements are currently being undertaken.

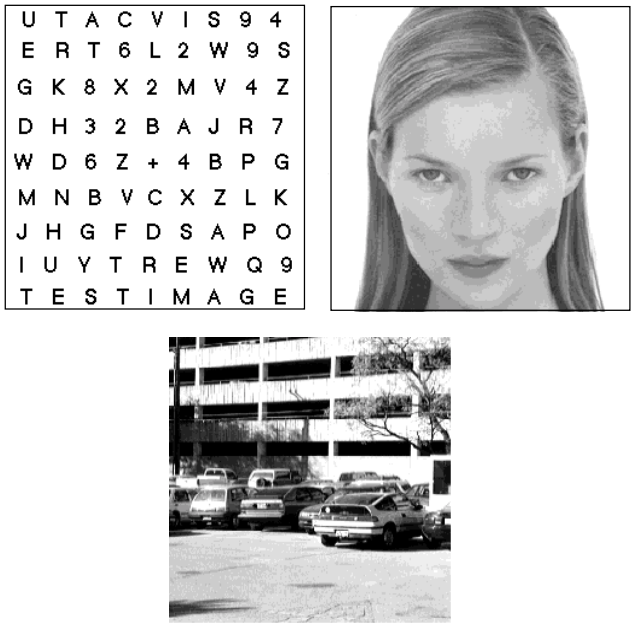



Figure 5: Three images (256x256, 8-bit gray scale) used in the perceptual evaluation of the FIS, chosen to represent the general categories of expected images.

Figure 6 shows, for example, a centrally fixated foveated image for the "letters" image with a half resolution constant ($e_2$) of 1°. This value of $e_2$ reduces the number of transmitted pixels from 65,536 to 3,488 (a factor of 18.8). All subjects reported smooth, accurate tracking and excellent overall functioning of the foveating algorithms. Upon steady fixation , most subjects noted that they were aware of the reduced resolution in the peripheral visual field, but that the effect was minimal and had only a small impact on the perceived quality of the image.

High contrast images (like the letter chart) were also reported to exhibit some reduction in the perceived contrast in the periphery, as compared to an unfoveated image. The effect was significantly reduced in the natural and face images, where contrast changes in the image are typically smoother. Without reference to the unfoveated image (i.e. without switching between the two images), few subjects were even aware of this peripheral reduction in contrast for the natural scene and face images.

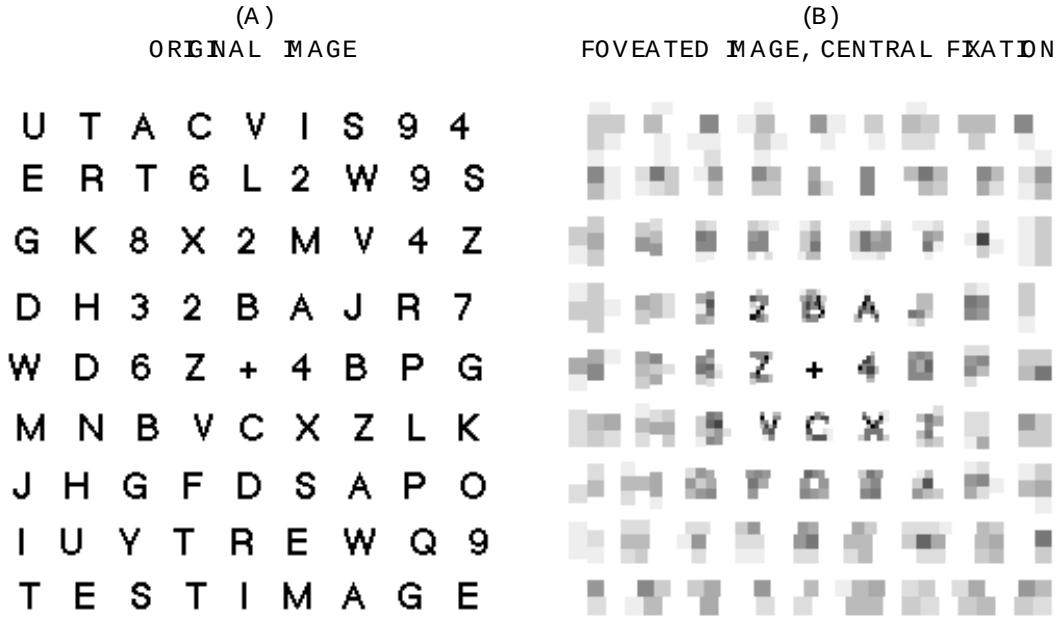



Figure 6: (A) The original 256x256 high contrast letter image. (B) The same image in a foveated viewing scheme, centrally fixated. Notice how the size of the SuperPixels grows as eccentricity increases.

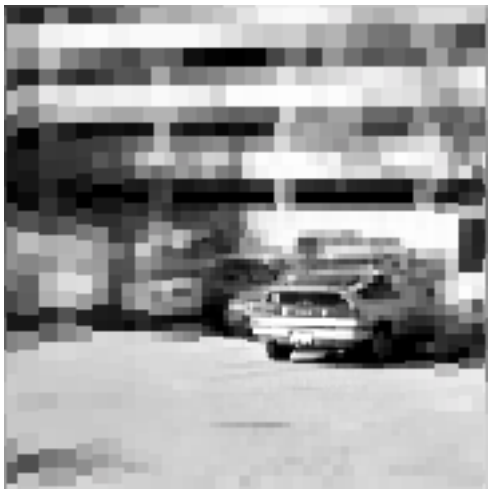
Subjects also reported some apparent motion effects in the periphery (best described as 'image sparkling'), even at steady fixation. This effect was exacerbated by image updates due to small eye tremors and micro saccades, a side effect of using a high precision eye tracking system. Increasing the eye movement threshold parameter eliminated the apparent motion effects at steady fixation. However, eye movements around the image still result in apparent motion effects in the periphery. The effects are not substantial. However, for certain tasks, such as piloting a remote vehicle, the separation of apparent motion from real motion could cause some difficulties. Some of these artifacts can be minimized through the application of post-transmission SuperPixel averaging, resulting in a blurring of the image outside the unaltered point-of-gaze. We have begun to implement and evaluate a number of these post-transmission averaging methods, and preliminary results indicate that there is an improvement in the perceptual quality of the image following this type of image operation.

All of these perceptual artifacts are undoubtedly due, in large part, to the fact that a 256x256 image results in screen pixels that are larger than the resolution limit of the eye. In other words, the smallest possible SuperPixel (1 screen pixel) is resolvable in the

center of the fovea. SuperPixel size grows with eccentricity and hence remains above the human resolution limit (by a fixed proportion) at each eccentricity. When the SuperPixels are above the resolution limit they will produce reduced contrast, motion aliasing and visible SuperPixel edges. Increasing the resolution of the image so the screen pixels are at the resolution limit in the fovea (e.g., a 1024 x 1024 image viewed at 20°) results in foveated images that are virtually indistinguishable from the original (see Figure 7), while still reducing the number of transmitted image pixels by a factor of 18.8.

**(A)**

1024x1024 Foveated Image



**(B)**

256X256 Foveated Image

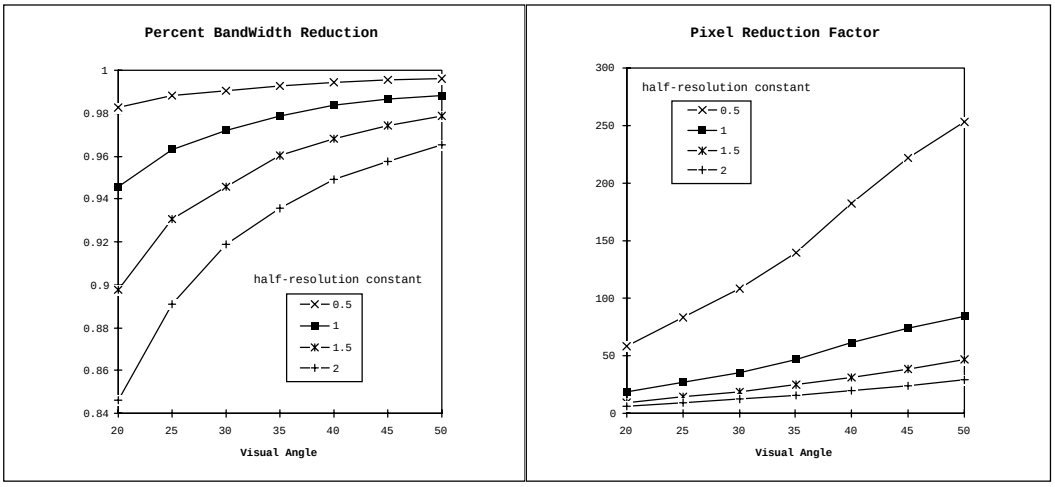
Figure 7: (A) A 1024x1024 foveated image fixated on the license plate of the Honda CRX. Because the center pixel size is below the resolution limit of the visual system (for this size image), the resulting degradations due to foveation are imperceptible with appropriate fixation (notice the blockiness of the tree for verification that it is, indeed, a foveated image). (B) a 256x256 foveated image with the same fixation, for comparison.

## 4.0 BANDWIDTH REDUCTION

Use of the Foveated Imaging System has demonstrated that significant bandwidth reduction can be achieved, while still maintaining access to high detail at any point of a given image. Using a 20° field-of-view, a half resolution constant ($e_2$) of 1° and a foveal SuperPixel size of 1, we are able to achieve bandwidth reductions of 94.7% (18.8 times reduction) for 256x256 8-bit gray scale images at eye-movement update rates of up to 20 Hz (the refresh rate of the display was 60 Hz.). Increasing the field-of-view or decreasing the half-resolution constant will result in greater bandwidth savings. For example, for a 50° field-of-view, bandwidth is reduced by a factor of 96.4. A selected

sampling of bandwidth reductions for different half-resolution constants and fields-of-view is shown in Figure 8.

It is important to note that other image compression schemes, such as DPCM or run length coding, for example, can be easily used in conjunction with the foveated imaging system. The effects of the supplementary compression are multiplicative; a factor of 4 compression on the foveated image described above would yield an overall compression of 75.2. This suggests that extremely high compression ratios are attainable using computationally efficient coding methods.




(a) (b)

Figure 8: Bandwidth reduction for 4 half resolution constants ($e_2$) as a function of the visual angle of the display device, expressed as (a) percent bandwidth reduction compared to an unfoveated display and (b) a reduction factor (i.e. the factor by which the number of pixels is reduced).

## 5.0 CONCLUSIONS

The Foveated Imaging System has demonstrated that real-time image compression can be achieved on general purpose computer processors, with no perceptual loss at the point of gaze and minimal perceptual anomalies in the peripheral visual field. Use of post transmission filtering and expansion to higher resolution images (where the screen pixels are near the resolution limit o the human visual system) should result in highly compressed images that are perceptually indistinguishable from an the constant resolution image from which they were formed.

## 6.0 REFERENCES

Benderson, B.B., Wallace, R.S. and Schwartz, E.L. (1994) A miniature pan-tilt actuator: the spherical pointing motor. IEEE Transactions Robotics and Automation. Vol. 10, 298-308.

Geisler, W.S. and Banks, M.S. (1995) Visual Performance. In Bass, M. (Ed.) Handbook of Optics Volume 1: Fundamentals, Techniques and Design, 2nd Edition. New York: Mcgraw-Hill.

Howard, C.M. (1989) Display Characteristics of Example Light-Valve Projectors. AFHRL-TP-88-44. Operations Training Division, Air Force Human Resources Laboratory, Williams AFB, AZ.

Juday, R.D. and Fisher, T.E. (1989) Geometric Transformations for video compression and human teleoperator display. SPIE Proceedings: Optical Pattern Recognition, Vol. 1053, 116-123.

Warner, H.D., Serfoss, G.L. and Hubbard, D.C. (1993) Effects of Area-of-Interest Display Characteristics on Visual Search Performance and Head Movements in Simulated Low-Level Flight. AL-TR-1993-0023. Armstrong Laboratory, Human Resources Directorate, Aircrew Training Division, Williams AFB, AZ.

Wassel, H., Grünert, U., Röhrenbeck, J., and Boycott, B.B. (1990) Retinal ganglion cell density and cortical magnification factor in the primate. Vision Research, 30, 1897-1911.

Weiman, C.F.R. (1990) Video Compression Via Log Polar Mapping. SPIE Proceedings : Real Time Image Processing II, Vol. 1295, 266-277.

Wilson, H.R., Levi. D., Maffei, L., Rovamo, J. and Devalois, R. (1990). The Perception of Form: Retina to Striate Cortex. In L.S. & J.S. Werner (Eds.), Visual Perception: The Neurophysiological Foundations (pp 232-272). San Diego: Academic Press.