Simple Multiresolution Approach for Representing Multiple Regions of Interest (ROIs)

Andrew Duchowski
andrewd@cs.tamu.edu

Bruce H. McCormick
mccormick@cs.tamu.edu

Department of Computer Science
Texas A&M University
College Station, TX, 77843-3112

ABSTRACT

A simple spatial-domain multiresolution scheme is presented for preserving multiple regions of interest (ROIs) in images. User-selected ROIs are maintained at high (original) resolution while peripheral areas are degraded. The presented method is based on the well-known *MIP* texture mapping algorithm used extensively in computer graphics. Most ROI schemes concentrate on preserving a single foveal region, usually attempting to match the visual acuity of the human visual system (HVS). The multiple ROI scheme presented here offers three variants of peripheral degradation, including linear and nonlinear resolution mapping, as well as a mapping matching HVS acuity. Degradation of image pixels is carried out relative to each ROI. A simple criterion is used to determine screen pixel membership in given image ROIs. Results suggest that the proposed multiple ROI representation scheme may be suitable for gaze-contingent displays as well as for encoding sparse images while optimizing compression and visual fidelity.

Keywords: Multiresolution, ROI, Texture Mapping, Image Processing.

1 INTRODUCTION

Region Of Interest (ROI) [1] image processing aims at presenting a single high resolution area to the fovea. The goal of limiting high resolution data to a foveal "spotlight of attention" is to minimize bandwidth requirements, while matching early vision capabilities of the Human Visual System (HVS) in the periphery. Typically, peripheral imagery is degraded in order to minimize information content prior to encoding or transmission. Approaches range from luminance attenuation, smoothing, and locally adapted transform coding techniques (i.e., local clamping of DCT coefficients).

On the other hand, feature detection algorithms that locate visually interesting information (multiple ROIs, essentially) tend to treat the whole image as the periphery and do not typically provide a foveal ROI. The

[1] Also known as the Area or Interest (AOI).

objective of the method presented here is to provide a foveal ROI and also represent peripheral ROIs (pROIs) as potential future foci of gaze (but not necessarily attention). Moreover, the periphery around each ROI is degraded matching the HVS acuity function. For gaze-contingent displays, this type of processing may be more suitable than single-ROI methods since its aim is to preserve preview benefit.[2]

In gaze-contingent (GC) applications (such as flight simulators), emphasis has usually been placed on representing the foveal ROI, while homogeneously degrading the periphery.[5,6] In the Super Cockpit Visual World Subsystem, Kocian considered visual factors including contrast, resolution and color in the design of a head-tracked GC display. In their Simulator Complexity Testbed (SCTB), Longridge et al included an eye-slaved ROI as a major component of the Helmet Mounted Fiber Optic Display (HMFOD). This ROI provided a high resolution inset in a low resolution (presumably homogeneous) field which followed the user's point of regard. The precise method of peripheral degradation was not described apart from the criteria of low resolution. However, the authors did point out that a smooth transition between the ROI and background was necessary in order to circumvent the possibility of a perceptually disruptive edge artifact.

Recently, more sophisticated approaches have been proposed for ROI-based video coding.[7,10] While these schemes concentrate on the representation of the foveal ROI, the periphery is processed by conventional means such as smoothing or quantization of transform coefficients. While the transition from the high-resolution ROI to the periphery may be smooth, it does not necessarily match the HVS acuity function.

Various multiple-ROI image processing schemes have been proposed for feature-detection tasks where features are either preserved or enhanced, while the rest of the imagery is decimated in some way. Of particular relevance are multiresolution, pyramidal schemes.[8,9] Sandon developed a connectionist network model of guided visual attention, while Pölzleitner and Wechsler used distributed associative memories (DAMs) in preattentive mode to find relevant segments in the field of view. In both cases, a Gaussian pyramid was used to subsample the scene into low resolution levels. The purpose of these types of schemes is to locate peripheral ROIs in order to simulate guided visual attention. In contrast to foveal ROI coding schemes, the periphery seems to be of greater importance in these approaches, although again the transition from each ROI to its surround does not necessarily match the HVS acuity function.

The intent of the present research is to propose a scheme to juxtapose the representation of pROIs, as produced by feature-detection tasks, with a foveal ROI as found in gaze-contingent display modalities. Utilizing a multiresolution spatial pyramid, the objective here is not to find visual attractors, but to offer a method of representation suitable for attentive and pre-attentive viewing. The work described here centers on the reconstruction of the image from prefiltered (texture) maps of the original image. The novel aspect of the approach is the ability to maintain several ROIs within the image while gradually degrading the periphery around each ROI. ROI shape is circular (although it need not be) and the peripheral degradation function can be chosen from several variants with respect to spatial distance from the center of the ROI. In this implementation, the periphery can be degraded using a linear, nonlinear, or HVS acuity-matching function. The resolution of an arbitrary peripheral pixel depends on its distance to the closest ROI.

Our approach is a variant of the *MIP* mapping algorithm which can be considered in the class of pyramidal algorithms. Using MIP mapping terminology, the screen scanning procedure is adopted here for texture mapping ROIs. This strategy was chosen for its potential constant time optimization.[3,1] The algorithm, presented here in detail, shows promising results in terms of visual fidelity and compression potential.

2 THE TECHNIQUE

The multiresolution algorithm presented in this paper is a straightforward adaptation of the MIP texture mapping algorithm used extensively in computer graphics.[11] The scheme can also be considered a special case in the classical

pyramid framework for early vision.[4] Our algorithm preserves the original image resolution within multiple user-selected ROIs and gradually degrades the periphery surrounding each ROI according to a specified resolution mapping function.

Given an $N \times N$ image, $I(x,y)$, and assuming N is a power of 2, the image is subsampled and smoothed into $\log_2(N)+1$ subimages, $I_k(x,y)$. The set of subsampled images forms a pyramid. To reconstruct the image, the intensity of each pixel at spatial location (i,j) is calculated as a linear combination of pixel intensities in the pyramid. Resolution information is obtained from the $I_k(x,y)$ subimages, depending on pixel location (i,j). The procedure is detailed below.

Step 1 Prefilter the original image at levels $k = 0,1,2,\ldots,\log_2(N)$ with an averaging box filter of width 2^k. At each level k, the image is reduced by a factor of 4 from the previous level $k-1$ until at level $k = \log_2(N)$ the image is a single pixel. At level k each subimage is of size $\frac{N}{2^k} \times \frac{N}{2^k}$. Schematically,

$$I_k(\left\lfloor \frac{i}{M} \right\rfloor, \left\lfloor \frac{j}{M} \right\rfloor) = \frac{1}{M^2} \sum_{m=1}^{M} \sum_{n=1}^{M} I(i,j)$$

where k is the resolution level, M is the box filter size (2^k). Note that this is a slightly different approach from classical pyramid approaches since each subimage is subsampled directly from the original image $I(x,y)$. Alternatively, each higher subimage in the pyramid can be subsampled from the subimage below. Furthermore, any smoothing filter can be substituted in place of the above simple averaging box filter. In general,

$$I_k(i,j) = \sum_{m=1}^{M} \sum_{n=1}^{M} h(m,n) I_{k-1}(2i+m-z, 2j+n-z)$$

where in this case, M is the (constant) width of the convolution mask, z is a constant, $z = \lfloor (M+1)/2 \rfloor$. The smoothing filter should satisfy the following constraints:[4]

(a) Normalization

$$\sum_{m=1}^{M} \sum_{n=1}^{M} h(m,n) = 1$$

(b) Symmetry

$$h(m,n) = h(M+1-m,n) = h(m,M+1-n) = h(M+1-m,M+1-n)\ \forall m,n$$

(c) Unimodality

$$0 \le h(m,n) \le h(p,q) \text{ for } m \le p < \frac{M}{2} \text{ and } n \le q < \frac{M}{2}$$

(d) Equal contribution to the next level—all pixels of I must contribute the same total weight to each pixel of I_k.

(e) Separability

$$h(m,n) = h(m)h(n)$$

The normalized box filter in this paper satisfies the above constraints. Alternatively, a Gaussian operator can be used. The subsampled images are shown in Figure 1 as obtained by processing the 512×512 *lena* image (subimage I_0 is not shown since it is identical to the original).

Step 2 Select a mapping function, l, from the pyramid to *image space*. This step is a precursor to reconstruction of the image. The choice of mapping function is crucial in determining the final representation of the image. It is important to note that resolution is distributed nonlinearly (by decreasing powers of 2) in the pyramid. Since reconstruction is carried out in image space (dependent on the pixel location (i,j) in the final image),

Figure 1: Subimages, processed by normalized box filter

the resultant percent resolution distribution is obtained by taking the inverse of the constant 2 raised to the mapping function, i.e., [2]

$$\% \text{ resolution} = \frac{1}{2^l}$$

In the current implementation, three different mapping functions are used: a linear mapping, a nonlinear mapping, and an empirical HVS acuity matching mapping function. The linear and nonlinear mapping functions were chosen as approximate lower and upper bounds, respectively, to the HVS matching function, in terms of percent resolution. Each mapping function segments the image into concentric resolution regions, or bands. In all three implementations, resolution within the central 5° of each ROI is consistent. Although this is not a restriction imposed by MIP mapping, we chose to maintain the size of each ROI consistent across mapping functions (by the choice of R) so that different peripheral degradation methods could be examined. All three mapping functions are given below.

(a) Linear mapping

$$l = \frac{d}{R}$$

(b) Nonlinear mapping

$$l = A(1 - e^{-\lambda \frac{d}{R}})$$

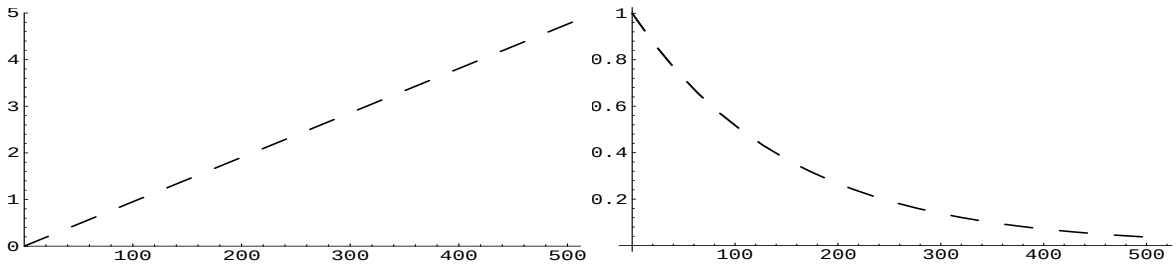
(c) HVS acuity mapping

$$l = -\frac{ln(\text{empirical \% resolution at pixel distance})}{\ln(2)}$$

The parameter d is the pixel distance from the ROI center, and R (set to 105) is the radius of the highest resolution region (foveal region). The derivation of R based on an empirical HVS acuity function is described elsewhere.[2] For the nonlinear mapping function, A is the asymptote approximated at the image boundary (here $A = 2.35$). To consistently preserve resolution within the radius of the highest resolution region, λ is chosen so that $l = 1$ at pixel distance R. That is,

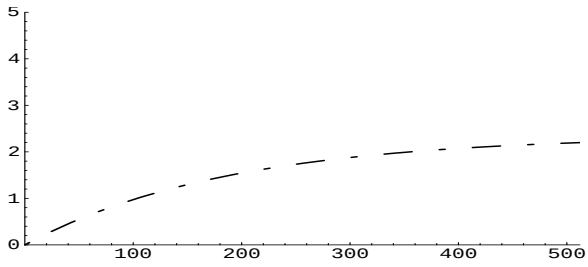
$$1 = A(1 - e^{-\lambda})$$

[2]Percent resolution refers to relative resolution in the reconstructed image assuming 100% resolution in the original.

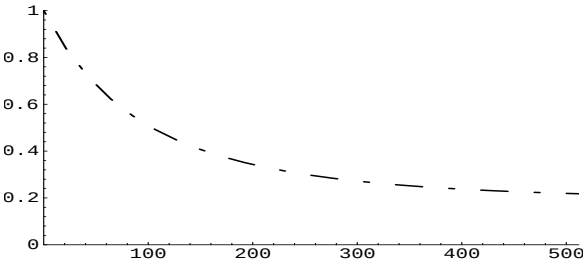


(a) Linear mapping: l vs. d

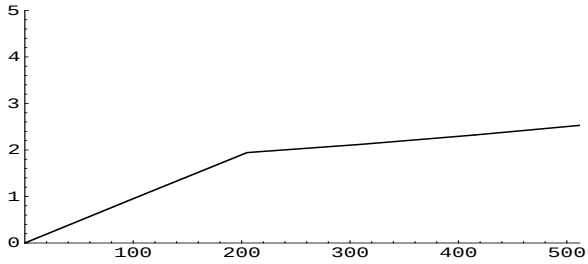
(b) Linear mapping: % resolution vs. d



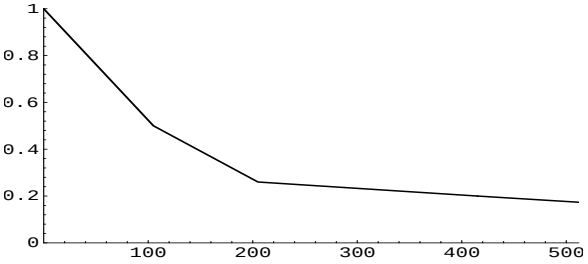
(c) Nonlinear mapping: l vs. d



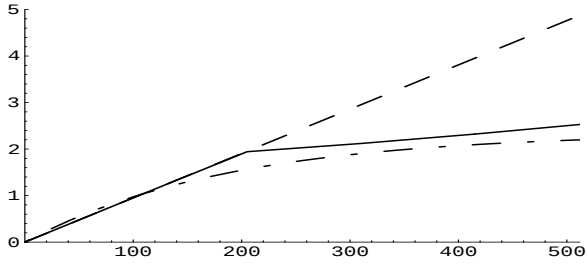
(d) Nonlinear mapping: % resolution vs. d



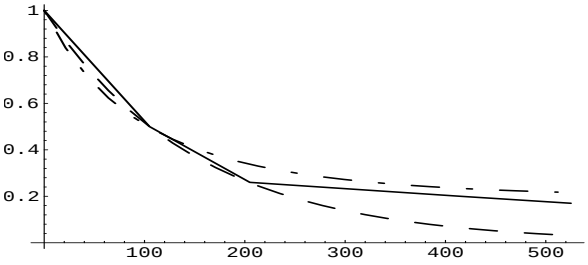
(e) HVS mapping: l vs. d



(f) HVS mapping: % resolution vs. d



(g) Composite plot: l vs. d



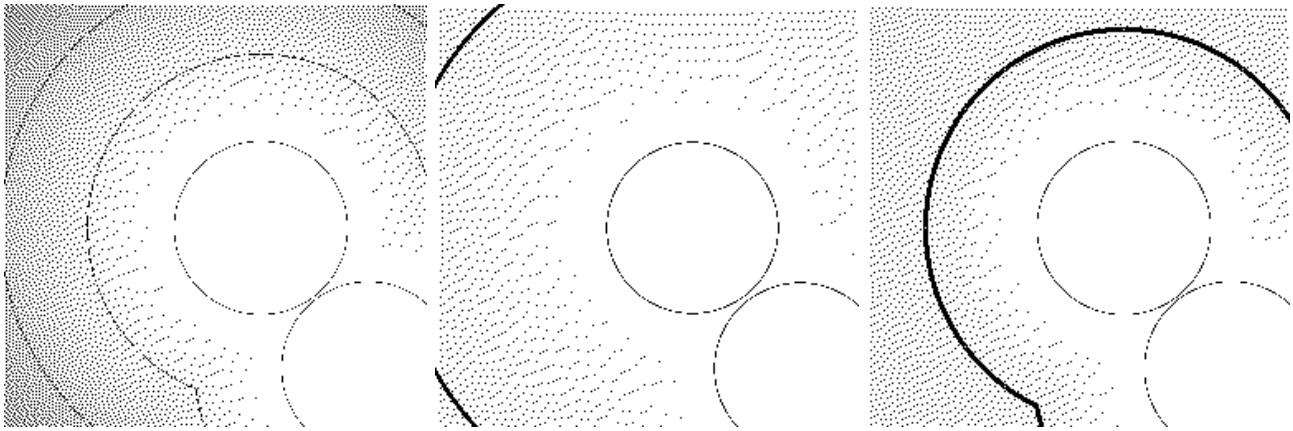
(h) Composite plot: % resolution vs. d

Figure 2: MIP mapping functions

so that

$$\lambda = \ln\left(\frac{A}{A-1}\right)$$

The HVS acuity mapping was originally obtained in terms of percent resolution, that is, it was specified as a function in resolution space and thus the above mapping function (in image space) is the inverse of that empirical function. All three functions are plotted in Figure 2 showing the mapping functions in image space, and the corresponding relative resolution. The concentric resolution bands in image space are shown in Figure 3 for the *lena* image, with 2 ROIs. Lighter areas are reconstructed at higher resolution, black rings are level boundaries.



(a) Linear mapping. (b) Nonlinear mapping. (c) HVS mapping.

Figure 3: Resolution bands in image space.

Step 3 Reconstruct the image. The given mapping function, l, effectively segments the image space into resolution bands bounded by resolution levels $\lfloor l \rfloor$ (or $\lceil l-1 \rceil$), $\lceil l \rceil$. Pixels interior to a band are interpolated between $I_{l-1}(x,y)$ and $I_l(x,y)$. Any pixels lying on the boundary of a resolution band, l, are obtained directly from subimage $I_l(x,y)$.

During the screen scan, check which ROI is closest to the screen pixel (x,y). The metric used here is the Euclidian distance (although any distance metric can be used instead). Define (x_0,y_0) as the center ROI pixel. Based on the pixel's distance d from (x_0,y_0), find the nearest lower and higher resolution bands, $l-1$, l, respectively. Obtain pixel values from subimages I_{l-1} and I_l and interpolate:

$$I(i,j) = (1-interp)I_{l-1}\left(\left\lfloor \frac{i}{\frac{N}{2^{l-1}}} \right\rfloor, \left\lfloor \frac{j}{\frac{N}{2^{l-1}}} \right\rfloor\right) + (interp)I_l\left(\left\lfloor \frac{i}{\frac{N}{2^{l}}} \right\rfloor, \left\lfloor \frac{j}{\frac{N}{2^{l}}} \right\rfloor\right)$$

The interpolation used here is linear, although a weighted interpolation scheme can be substituted for different effects. Furthermore, prior to this inter-map interpolation, various filtering operations can be included. For example, an averaging operation or intra-map interpolation is common.

The algorithm for an 8×8 image with two linearly mapped ROIs is depicted in Figure 4.

3 RESULTS

The 512×512 *lena* image was processed with 2 ROIs, one centered on Lena's eyes, the other in the lower right portion of the image (on the reflection in the mirror). The ROI on the mirror reflection was chosen arbitrarily. It was picked in the lower right portion of the image so that degradation effects can be examined at large distances

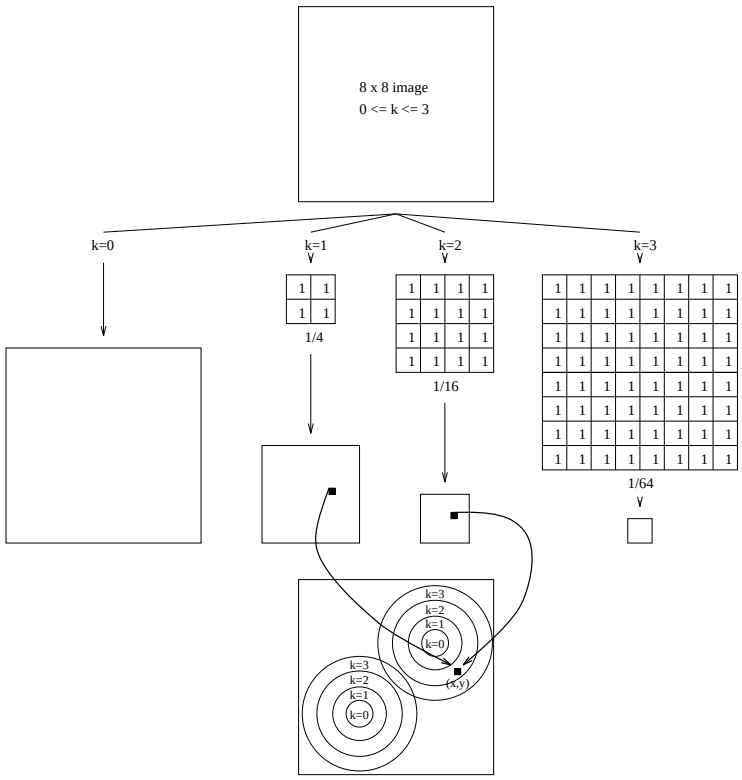


Figure 4: Depiction of algorithm

from both ROIs (relatively far in the periphery). Results (with no intra-map interpolation or smoothing) are shown in Figure 5.

Linear mapping of the periphery gives the worst performance in terms of visual quality. Blocking artifacts are clearly visible. Intra-map interpolation and smoothing options do not offer much help. The result of smoothing and intra-map interpolation produces a blurring effect which brings out subimage boundaries thereby producing artificial image segmentation. Linear mapping, however, provides the best image sources for compression. Non-linear mapping provides the best visual results at the cost of the least compression potential. The HVS-matching mapping function falls in between, as expected.

Although the results look promising, it must be noted that the degradation effects are minimal on small images. Images should be larger than 512 $\times$ 512 in size to notice appreciable degradation effects. Figure 6 shows a 1024 $\times$ 1024 image of Mount Olympus on the planet Mars, with one ROI centered on the circular crater-like feature at the top left of the image. Peripheral degradation is performed by the HVS-matching function. The image is scaled due to the printed page's size constraints, but degradation effects are visible in the lower right portion of the image.

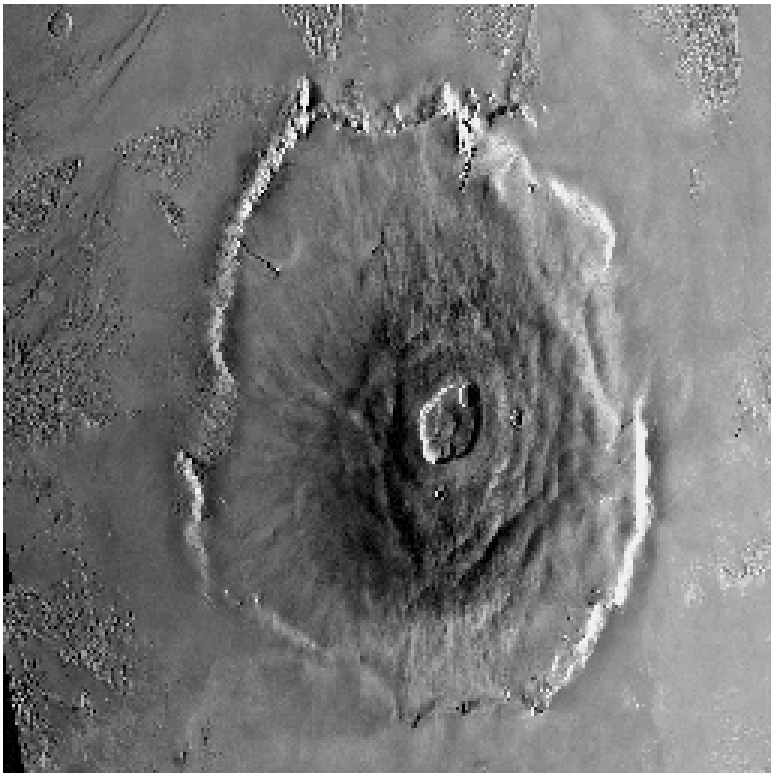
(a) Original image.

(b) HVS mapping, MSE = 140.0.

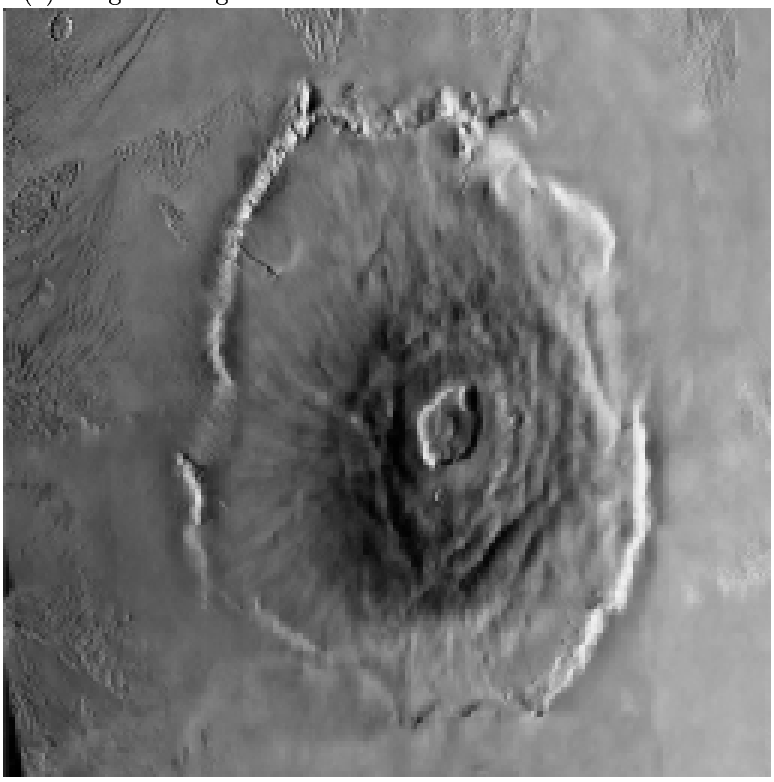
(c) Linear mapping, MSE = 156.7.

(d) Nonlinear mapping, MSE = 131.0.

Figure 5: *Lena* images



(a) Original image.



(b) HVS mapping, MSE = 206.0.

Figure 6: *Mount Olympus* images

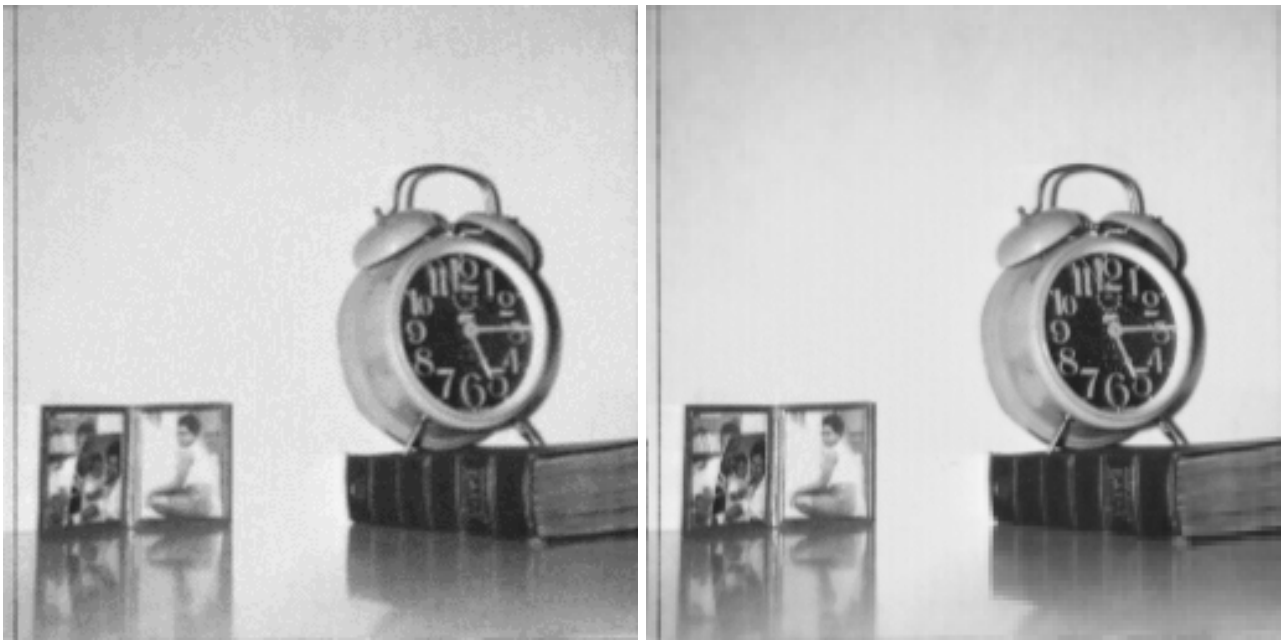
4 DISCUSSION

There is an obvious tradeoff between visual fidelity and compression potential. In the case of sparse images, however, degradation effects in the far periphery are not very noticeable, even under linear mapping. This is not surprising if the peripheral imagery is a homogeneous image field. To demonstrate, a 512×512 *clock* image was processed by the three mapping functions and is shown in Figure 7. Two ROIs are centered on the clock face and image frame. All three mapping functions produce similar results yet the linear mapping function still gives the highest compression potential.

Currently the multiresolution algorithm presented here only handles grayscale images, but this scheme should extend to processing other image components such as contrast and color. The criteria used here for image decimation (box filters), image reconstruction (interpolation from subimages) and effective composition of ROIs (nearest-ROI membership criterion) are limited by their simplicity. The general pyramidal approach, however, is very flexible. Every step of the proposed process is subject to alteration. In Step 1, filters other than the box filter may be substituted. Although the current method operates only in the spatial domain, it can easily be extended to provide frequency analysis. Possible extensions/augmentations should consider wavelet approaches. In Step 2, the obvious flexibility is in the choice of the mapping function. Note that the mapping function can be specified in image space (by defining the distribution of resolution bands) or in the pyramid, as was was done with the HVS acuity-matching function. The mapping function can also be used to vary the form of ROIs. In this implementation all ROIs are of constant size and shape. Future work will consider variable-form pROIs. In Step 3, as suggested, the interpolation need not be linear. To decrease smoothing effects, for example, interpolation may be weighted to favor lower resolution levels where more high frequency information is contained. The ROI-membership criterion may also be weighted to favor the primary (foveal) ROI.

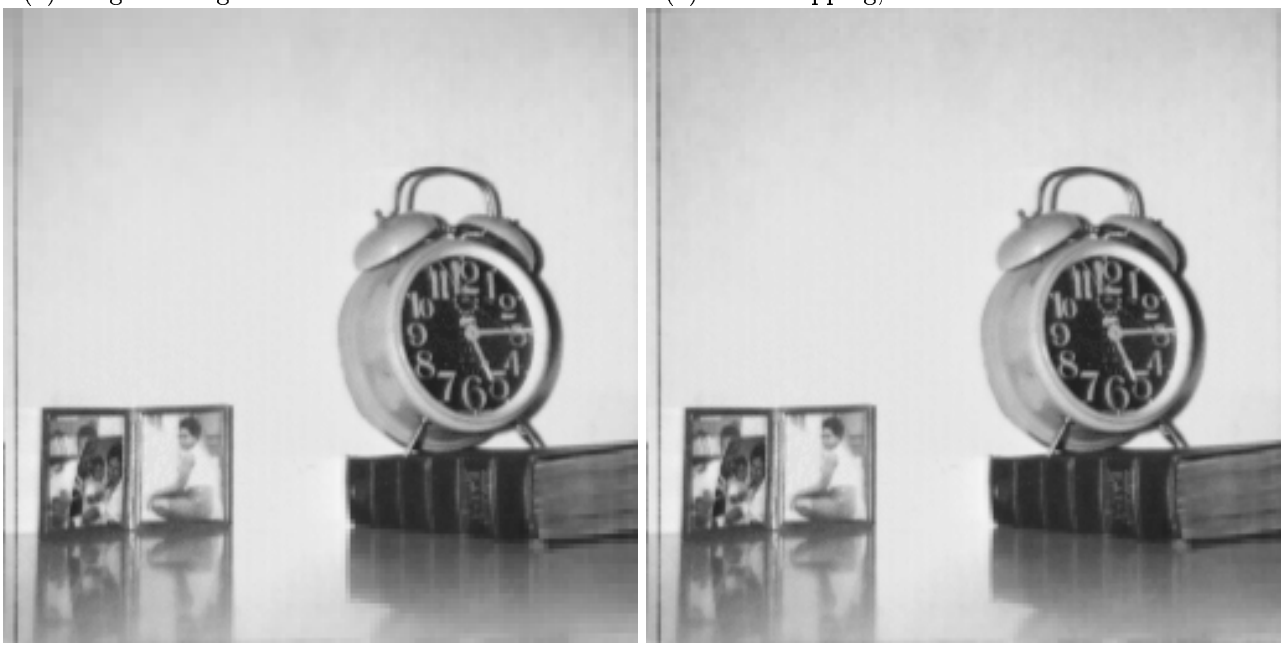
5 CONCLUSION

A simple image processing scheme is used to represent multiple regions of interest in still images, while providing gradual degradation of resolution in peripheral areas. The algorithm presented here provides a prototype for exploration of multiple ROI image processing intended for gaze-contingent displays. The algorithm allows good compression and efficient implementation by providing a flexible framework for specifying ROI form and the resolution degradation function.



(a) Original image.

(b) HVS mapping, MSE = 93.9.



(c) Linear mapping, MSE = 96.5.

(d) Nonlinear mapping, MSE = 95.7.

Figure 7: *Clock* images

6 REFERENCES

[1] Franklin C. Crow. Summed-area tables for texture mapping. In *Computer Graphics*, volume 18, pages 207–212. SIGGRAPH, 1984.

[2] Andrew T. Duchowski and Bruce H. McCormick. Pre-attentive considerations for gaze-contingent image processing. In *Conference on Human Vision, Visual Processing, and Digital Display VI*, San Jose, CA, February 1995. SPIE.

[3] Alain Fournier and Eugene Fiume. Constant-time filtering with space-variant kernels. In *Computer Graphics*, volume 22, pages 229–238. SIGGRAPH, 1988.

[4] Jean-Michel Jolion and Azriel Rosenfeld. *A Pyramid Framework for Early Vision.* Kluwer Academic Publishers, Norwell, MA, 1994.

[5] Dean Kocian. Visual world subsystem. In *Super Cockpit Industry Days: Super Cockpit/Virtual Crew Systems*, Air Force Museum, Wright-Patterson AFB, Ohio, 31 March–1 April 1987. Air Force Systems Command/Human Systems Division/Armstrong Aerospace Medical Research Laboratory.

[6] Thomas Longridge, Mel Thomas, Andrew Fernie, Terry Williams, and Paul Wetzel. Design of an Eye Slaved Area of Interest System for the Simulator Complexity Testbed. In *Interservice/Industry Training Systems Conference*, pages 275–283. National Security Industrial Association, 1989.

[7] E. Nguyen, C. Labit, and J-M. Odobez. A *ROI* Approach for Hybrid Image Sequence Coding. In *International Conference on Image Processing (ICIP)'94*, pages 245–249. IEEE, November 1994.

[8] Wolfgang Pölzleitner and Harry Wechsler. Selective and Robust Perception Using Multiresolution Estimation Techniques. In *11^{th} International Conference on Pattern Recognition, Vol. II: Conference B*, pages 54–57. IEEE, 1992.

[9] Peter A. Sandon. Simulating Visual Attention. *Journal of Cognitive Neuroscience*, 2(3):213–231, 1990.

[10] Lew B. Stelmach and Wa James Tam. Processing Image Sequences Based on Eye Movements. In *Conference on Human Vision, Visual Processing, and Digital Display V*, pages 90–98, San Jose, CA, February 8-10 1994. SPIE.

[11] Alan Watt and Mark Watt. *Advanced Animation and Rendering Techniques.* Addison-Wesley, Reading, MA, 1992.