Representing Multiple Regions Of Interest with Wavelets*

Andrew T. Duchowski†
andrewd@cs.tamu.edu
Department of Computer Science
Texas A&M University
College Station, TX 77843-3112

ABSTRACT

Gaze-contingent systems minimize bandwidth requirements by displaying high-resolution imagery within eye-slaved Regions Of Interest (ROIs) of limited spatial extent. The parafoveal resolution transition must be sufficiently smooth to render degradation effects imperceptible. In this paper, a wavelet image filtering scheme is presented which preserves high (original) resolution in Regions Of Interest matching foveal vision, and gradually degrades resolution in the periphery. The method permits the representation of multiple ROIs extending previous work based on MIP-mapping to the wavelet domain. Degradation is achieved through wavelet coefficient scaling following Voronoi partitioning of the image plane. Three variants of peripheral degradation are offered, including one matching human visual acuity. Reconstruction examples of images processed with the Haar and Daubechies-6 wavelets are provided. Results from gaze-contingent experiments, conducted to test perceived impairment of degraded image sequences, are summarized suggesting imperceptible degradation effects.

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

1 INTRODUCTION

In order to match Human Visual System (HVS) spatial acuity, a multiresolution method is developed for preserving multiple Regions Of Interest (ROIs) in images.Regions Of Interest are maintained at high (original) resolution while peripheral areas are degraded. The proposed wavelet method for generating images of variable resolution operates by selectively scaling wavelet coefficients. This technique extends previous work based on *MIP-mapping*.[4] MIP-mapping involves smooth subsampling of an image (decomposition) in order to texture map (reconstruct) the image at variable resolution. It is shown that reconstruction of appropriately scaled wavelet coefficients is equivalent to linear interpolation MIP-mapping.

Region Of Interest (ROI) image processing aims at presenting a single high resolution area to the fovea. The goal of limiting high resolution to a foveal "spotlight of attention" is to minimize bandwidth requirements, while matching early vision capabilities of the HVS in the periphery. Typically, peripheral imagery is degraded in order

* This research was supported in part by the National Science Foundation, under Infrastructure Grant CDA-9115123 and CISE Research Instrumentation Grant CDA-9422123, and by the Texas Advanced Technology Program under Grant 999903-124.

† As of 01/98 the author can be reached at *andrewd@cs.clemson.edu*, Department of Computer Science, 451 Edwards Hall, Clemson University, Clemson, SC 29634-1906.

to minimize information content prior to encoding or transmission. Approaches range from luminance attenuation, smoothing, and locally adapted transform coding techniques (e.g., local clamping of DCT coefficients[10]). 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 objective of the method presented here is to provide a foveal ROI and also represent peripheral ROIs as potential future foci of gaze (but not necessarily attention). Moreover, the periphery around each ROI is degraded matching the acuity function of the HVS.

2 DISCRETE WAVELET TRANSFORM

Utilizing the multiresolution analysis (MRA) framework, introduced by Meyer and Mallat,[9] an n-length discrete function at the j^{th} level of resolution,

$$f^j(x) = f^j_\phi(1), f^j_\phi(2), \ldots, f^j_\phi(n), \tag{1}$$

is decomposed through relations:

$$f^{j-1}_\phi(x) = \sum_k h_k f^j_\phi(2x+k), \quad f^{j-1}_\psi(x) = \sum_k g_k f^j_\phi(2x+k), \quad k \in \mathbf{Z},$$

where $\{h_k\}, \{g_k\}$ are one-dimensional low- and high-pass filters, corresponding to the scaling and wavelet functions ϕ and ψ, respectively. The Discrete Wavelet Transform (DWT) is then defined by:

$$\{Wf(x)\}(j-1) = f^{j-1}_\phi(1), f^{j-1}_\psi(2), \ldots, f^{j-1}_\phi(n-1), f^{j-1}_\psi(n). \tag{2}$$

Permuting the terms so that the first $n/2$ elements are the low-pass (scale) coefficients, i.e.,

$$\{Wf(x)\}(j-1) = f^{j-1}_\phi(1), f^{j-1}_\phi(3), \ldots, f^{j-1}_\phi(n-1), f^{j-1}_\psi(2), f^{j-1}_\psi(4), \ldots, f^{j-1}_\psi(n), \tag{3}$$

and relabeling the indices,

$$\{Wf(x)\}(j-1) = f^{j-1}_\phi(1), \ldots, f^{j-1}_\phi(n/2-1), f^{j-1}_\psi(n/2), \ldots, f^{j-1}_\psi(n) \tag{4}$$

the first $n/2$, smooth (or averaged), elements are recursively decomposed. The fully transformed function contains the global average as the first element, the next 2^j elements contain the detail (or difference) information at each resolution level j. Except for the average value, the transformed elements comprise the so-called wavelet coefficients of the function.

To reconstruct the function, the terms at each resolution level are repermuted so that the average and wavelet coefficients are interleaved, as per Equation (2). Introducing the $\bowtie$ operator denoting element interleave, the $j-1$ level coefficients can be arranged into an intermediate representation for reconstruction at level j:

$$f^{j-1}_{\phi \bowtie \psi}(2x+p) \quad = \quad (1-p) f^{j-1}_\phi(x) + (p) f^{j-1}_\psi(x),$$

for $p \in \{0,1\}$. Reconstruction at level j, with $p \in \{0,1\}$ is then written as:

$$f^j_\phi(2x+p) = (1-p) \sum_k \widetilde{h}_k f^{j-1}_{\phi \bowtie \psi}(x-k) + (p) \sum_k \widetilde{g}_k f^{j-1}_{\phi \bowtie \psi}(x-k)$$

which gives the original function $f^j(x)$ in (1). Note that the variable p is used as a selection variable, that is, in the dyadic wavelet reconstruction, the element at position $2x+p$ is the result of filtering the lower pyramidal elements with either dual reconstruction filter $\{\widetilde{h}_k\}$ or $\{\widetilde{g}_k\}$ (discussed below).[12] The permutation function $f^j_{\phi \bowtie \psi}$ serves as an alternate method of reconstruction used instead of traditional supersampling.[3]

Multidimensional extensions of the DWT rely on the use of multidimensional bases, constructed by obtaining the tensor product of unidimensional bases. The computational realization of the 2D DWT described here is an instance of the well-known pyramidal multiresolution representational framework first proposed by Tanimoto and Pavlidis.[13,7] The pyramidal model stipulates a hierarchical processing paradigm known as the *coarse-to-fine* (resolution) strategy. The pyramidal wavelet transform in particular is related to the Laplacian pyramid introduced by Burt and Adelson for image coding.[1] In two dimensions, a $\log_2 N$ level pyramid is constructed from an $N \times N$ image, where the bottom level of the pyramid (level $j = \log_2 N - 1$) contains the finest resolution, and the top level ($j = 0$) contains the coarsest. Note that the original image is considered the finest level of resolution (level $j = \log_2 N$), however it is not contained within the pyramid itself.

Given 1D scaling and wavelet filters H, G associated with ϕ, ψ, respectively, 2D filters, corresponding to the 2D wavelet bases $\phi\phi$, $\phi\psi$, $\psi\phi$, $\psi\psi$, are generated by the non-standard 2D wavelet basis construction using tensor products:

$$HH = H \otimes H^T, \quad HG = H \otimes G^T, \quad GH = G \otimes H^T, \quad GG = G \otimes G^T.$$

Taking dyadic filter translation into consideration, if the 1D filters are mutually orthogonal, then the 2D filters are also mutually orthogonal in x- and y-directions. In general, multidimensional orthogonal filters are also *separable*, i.e., satisfying

$$h(k, m) = h(k)h(m),$$

due to their tensor product construction. Note that the DWT is neither necessarily translationally nor rotationally invariant.

The multidimensional tensor product filters are useful for visualizing spatiotemporal properties of the multidimensional wavelet transform, however direct implementation with multidimensional filters is inefficient. Instead, relying on the separability of the filters, the wavelet transform can be implemented by processing each dimension separately. The decomposition relations describing the non-standard decomposition of the spatial average image at level j are:

$$\begin{aligned}
f_{\phi_r}^{j-1}(x,y) &= \textstyle\sum_k h_k f_{\phi\phi}^{j}(x, 2y+k) & f_{\psi_r}^{j-1}(x,y) &= \textstyle\sum_k g_k f_{\phi\phi}^{j}(x, 2y+k) \\
f_{\phi\phi}^{j-1}(x,y) &= \textstyle\sum_k h_k f_{\phi_r}^{j-1}(2x+k, y) & f_{\phi\psi}^{j-1}(x,y) &= \textstyle\sum_k h_k f_{\psi_r}^{j-1}(2x+k, y) \\
f_{\psi\phi}^{j-1}(x,y) &= \textstyle\sum_k g_k f_{\phi_r}^{j-1}(2x+k, y) & f_{\psi\psi}^{j-1}(x,y) &= \textstyle\sum_k g_k f_{\psi_r}^{j-1}(2x+k, y)
\end{aligned}$$

where $\{h_k\}, \{g_k\}$ are the one-dimensional low- and high-pass filters. The non-standard DWT first involves subsampling along the rows of the lower pyramidal level spatial average image (denoted by $f_{\phi\phi}^{j}$) to generate the temporary upper pyramidal level images $f_{\phi_r}^{j-1}$ and $f_{\psi_r}^{j-1}$. Due to dyadic downsampling, these images are half the width of the image at the lower resolution level. Subsampling then along the columns generates the four subimages denoted by the subscripts $\phi\phi$, $\phi\psi$, $\psi\phi$, and $\psi\phi$. The smooth (or averaged) subimage $f_{\phi\phi}^{j-1}$ is recursively subsampled at each stage of the decomposition. The transformed image contains the global average at the top of the pyramid, the lower layers contain the detail (or difference) information at each pyramid level. These lower layers comprise the so-called wavelet coefficients of the transformed image. An example of the 2D DWT applied to an image is shown in Figure 1. The transformed image has been processed for display purposes.

In traditional pyramidal approaches, where the pyramid contains only smoothed multiscale versions of the original image (e.g., texture-mapping applications), subimages at each level provide the pixel intensity values for reconstruction usually involving interpolation. In the wavelet transform, the image is synthesized by a recursive process of adding detail information to the average (smoothed) subimages in order to reconstruct the next level's average subimage. Rows and columns are interleaved[12] prior to filtering instead of the traditional null row and column padding.[2] Generally, row and column padding (supersampling) is used under monadic convolution. Dyadic

(a) Original *cnn* image. Reprinted with permission from Turner Broadcasting System, Inc.

(b) 2-level DWT (processed by histogram equalization with subsampled image inset).

Figure 1: Non-standard 2D DWT.

convolution precludes the need for padding, but instead requires that rows and columns be interleaved prior to reconstruction filtering. Introducing the $\bowtie_r$ and $\bowtie_c$ operators denoting row and column interleave, respectively, the reconstruction is obtained with the use of the following intermediate relations:

$$f^j_{\phi\phi \bowtie_r \psi\phi}(2x+p,y) = (1-p)f^{j-1}_{\phi\phi}(x,y) + (p)f^{j-1}_{\psi\phi}(x,y), \quad (5)$$

$$f^j_{\phi\psi \bowtie_r \psi\psi}(2x+p,y) = (1-p)f^{j-1}_{\phi\psi}(x,y) + (p)f^{j-1}_{\psi\psi}(x,y), \quad (6)$$

$$f^j_{\phi_r \bowtie_c \psi_r}(x,2y+q) = (1-q)f^j_{\phi_r}(x,y) + (q)f^j_{\psi_r}(x,y), \quad (7)$$

where $p,q \in \{0,1\},\ x,y,k \in \mathbf{Z}$ and

$$f^j_{\phi_r}(2x+p,y) = (1-p)\sum_k \widetilde{h}_k f^j_{\phi\phi \bowtie_r \psi\phi}(2x-k,y) + (p)\sum_k \widetilde{g}_k f^j_{\phi\phi \bowtie_r \psi\phi}(2x-k,y),$$

$$f^j_{\psi_r}(2x+p,y) = (1-p)\sum_k \widetilde{h}_k f^j_{\phi\psi \bowtie_r \psi\psi}(2x-k,y) + (p)\sum_k \widetilde{g}_k f^j_{\phi\psi \bowtie_r \psi\psi}(2x-k,y),$$

so that

$$f^j_{\phi\phi}(x,2y+q) = (1-q)\sum_k \widetilde{h}_k f^j_{\phi_r \bowtie_c \psi_r}(x,2y-k) + (q)\sum_k \widetilde{g}_k f^j_{\phi_r \bowtie_c \psi_r}(x,2y-k). \quad (8)$$

Equation (8) roughly states that $f^j_{\phi\phi}$ is reconstructed from the expansion of the f^{j-1} functions at resolution level $j-1$ by the dual (bi-orthogonal) filters $\{\widetilde{h}_k\}, \{\widetilde{g}_k\}$ representing scale and wavelet bases generated by $\widetilde{\phi}$, $\widetilde{\psi}$. Since $f^j_{\phi\phi}$ was originally projected by the 2D bases functions $\phi\phi$, $\phi\psi$, $\psi\phi$, and $\psi\psi$, represented by the tensor products of $\{h_k\}$ and $\{g_k\}$, producing the four lower resolution level functions $f^{j-1}_{\phi\phi}$, $f^{j-1}_{\phi\psi}$, $f^{j-1}_{\psi\phi}$, and $f^{j-1}_{\psi\psi}$, $f^j_{\phi\phi}$ is faithfully reconstructed provided $(\widetilde{\phi},\widetilde{\psi})$, (ϕ,ψ) satisfy (bi-)orthogonality conditions.

3 WAVELET INTERPOLATION

The Discrete Wavelet Transform can be used to reconstruct images at variable resolution by selectively scaling wavelet coefficients. Provided appropriate wavelet filters can be found, reconstruction exactly matches linear *MIP-mapping*.[1] MIP-mapping is a well known texture mapping algorithm used extensively in computer graphics.[15,14]

[1] The acronym MIP, introduced by Williams, is from the Latin phrase *multum in parvo* meaning "many things in a small place".

The algorithm involves preprocessing an image at several resolution levels (decomposition) in order to texture map (reconstruct) an image at variable resolution. In this sense, MIP-mapping falls under the classical pyramid framework for early vision.[7] In this section it is shown that linear interpolation of scaled wavelet coefficients is equivalent to linear interpolation under MIP-mapping. A description of the MIP-mapping approach itself is omitted, but can be found elsewhere.[4] Although the present discussion is limited to two-dimensional images, the wavelet coefficient scaling method is applicable to multidimensional signals for multiresolution representation.

3.1 Variable Resolution Reconstruction Through Wavelet Coefficient Scaling

To obtain interpolation results identical to MIP-mapping, an intuitive approach would be to maintain reconstructed scaled subimages produced by successive steps of the IDWT, then to perform the interpolation step. Although this approach would yield identical results due to the equivalence of subsampling filters and the DWT's perfect reconstruction (due to the orthogonality of the filters), it is memory-intensive. What is perhaps not obvious is that identical interpolation results can be obtained by scaling wavelet coefficients prior to reconstruction. Scaling of the wavelet coefficients prior to reconstruction results in the attenuation of the signal with respect to the average (low-pass) signal. Full decimation of the coefficients (scaling by 0) results in a lossy, subsampled reproduction of the original. Conversely, scaling wavelet coefficients by 1 preserves all detail information producing lossless reconstruction. Selectively scaling the coefficients by a value in the range $[0, 1]$, at appropriate levels of the wavelet pyramid, produces a variable resolution image upon reconstruction. This approach is equivalent to MIP-mapping reconstruction with linear interpolation of pixel values.

In MIP-mapping, the value of the interpolant p is determined by an arbitrary mapping function which specifies the desired resolution level l. The two closest pyramid resolution levels are then determined by rounding down and up to find subimage levels $j-1$ and j. The interpolant value is obtained by the relation:

$$p = l - \lfloor l \rfloor .$$

Note that the slope of the mapping function should match the resolution hierarchy of the pyramid, i.e., if resolution decreases eccentrically from some reference point, the parameter l should also decrease eccentrically. If it does not, its value may be reversed by subtracting from the number of resolution levels, i.e., $n - l$. To scale wavelet coefficients, p is set to either 0, 1, or the interpolant value at particular subbands according to the following relations dependent on l:

$$p = \begin{cases} 1, & j \leq \lfloor l \rfloor; \\ l - \lfloor l \rfloor, & j = \lceil l \rceil; \\ 0, & j > \lceil l \rceil. \end{cases} \tag{9}$$

For example, if at some particular pixel location (x, y), $l = 1.5$, then wavelet coefficients would be preserved (scaled by 1) at levels $j \leq 1$, scaled by .5 at level $j = 2$, and decimated (scaled by 0) at levels $j > 2$ at the appropriate pixel location in the subimages.

THEOREM 3.1. *Wavelet coefficient scaling is equivalent to linear pixel interpolation under MIP-mapping.*
Proof: Consider the interpolation step in the latter,

$$f = (1-p)f^{j-1} + (p)f^{j},$$

where pixel coordinates are made implicit. When there is no need for interpolation, i.e., the desired resolution level at a pixel falls on a mapped resolution boundary, $l - \lfloor l \rfloor = 0$, and $j = \lceil l \rceil = l$, then

$$f = \begin{cases} f^{j-1} = f^{l-1} & \text{if } p = 0; \\ f^{j} = f^{l} & \text{if } p = 1; \end{cases}$$

or simply $f = f^l$ meaning that the resolution at the given pixel will match the resolution level of the subimage at the l^{th} pyramid level. Simplifying the IDWT reconstruction Equation (8) and expanding,

$$f^j_{\phi\phi} = (pf^{j-1}_{\psi} + \ldots + (pf^2_{\psi} + (pf^1_{\psi} + (pf^0_{\psi} + f^0_{\phi\phi}))) \ldots), \tag{10}$$

where f^j_{ψ} collectively represents subimages containing wavelet coefficients, $f^j_{\psi\phi}$, $f^j_{\phi\psi}$, $f^j_{\psi\psi}$, with implied pixel coordinates. Note that in Equation (10) the symbol p is now the interpolant and not the binary selection variable as used in Equation (8). When $l - \lfloor l \rfloor = 0$ as above,

$$f = (0 \ldots + (f^l_{\psi} \ldots + (f^2_{\psi} + (f^1_{\psi} + (f^0_{\psi} + f^0_{\phi\phi}))) \ldots) \ldots).$$

Since $f^{j+1}_{\phi\phi} = f^j_{\psi} + f^j_{\phi\phi}$ at each level of reconstruction, the resultant image will contain resolution matching the contents of subimage f^{l+1}. The subimage $f^{l+1}_{\phi\phi}$ contains the average (lowpass) component at level $l+1$, or equivalently, it contains the entire reproduced image at level l. In this case, the reconstructed image will contain resolution no finer than that found in subimage f^l, i.e., $f = f^l$, as with MIP-mapping.

When $l - \lfloor l \rfloor \neq 0$, interpolation is required since the desired resolution level falls between resolution mapping boundaries. In the IDWT, with $\lfloor l \rfloor < l < \lceil l \rceil$,

$$f = (0 \ldots + (0 + (pf^{\lceil l \rceil}_{\psi} + (f^{\lfloor l \rfloor}_{\psi} \ldots + (f^2_{\psi} + (f^1_{\psi} + (f^0_{\psi} + f^0_{\phi\phi}))) \ldots))) \ldots).$$

That is, resolution at the given pixel location is no finer than what is reproduced at subimage $f^{\lceil l \rceil + 1}_{\phi\phi}$, since

$$\begin{aligned} f^{\lceil l \rceil + 1}_{\phi\phi} &= pf^{\lceil l \rceil}_{\psi} + f^{\lfloor l \rfloor + 1}_{\phi\phi} \\ &= pf^j_{\psi} + f^j_{\phi\phi} \\ &= f^{j+1}_{\phi\phi} \\ &= f^j, \end{aligned} \tag{11}$$

where $\lfloor l \rfloor + 1 = \lceil l \rceil$, and $j = \lceil l \rceil$ from Equation (9). Equation (11) shows that the reconstructed resolution will be no finer than the resolution contained in f^l matching the finest resolution level produced by MIP-mapping.

What remains to be shown is that the resolution gain obtained by scaling wavelet coefficients (at level j) is equivalent to the gain obtained by the MIP-mapping interpolation. That is, the scaling operation $pf^j_{\psi} + f^j_{\phi\phi}$ given in Equation (11) above is equivalent (in terms of resolution gain) to the interpolation $(1-p)f^{j-1} + (p)f^j$. To prove this, recall the reconstruction relation,

$$\begin{aligned} f^j_{\phi\phi} &= f^{j-1}_{\psi} + f^{j-1}_{\phi\phi}, \text{ or} \\ f^{j+1}_{\phi\phi} &= f^j_{\psi} + f^j_{\phi\phi}, \text{ giving} \\ f^j_{\psi} &= f^{j+1}_{\phi\phi} - f^j_{\phi\phi}. \end{aligned} \tag{12}$$

Substituting Equation (12) into the wavelet scaling equation,

$$\begin{aligned} pf^j_{\psi} + f^j_{\phi\phi} &= f^j_{\phi\phi} + pf^j_{\psi} \\ &= f^j_{\phi\phi} + p(f^{j+1}_{\phi\phi} - f^j_{\phi\phi}) \\ &= f^j_{\phi\phi} - pf^j_{\phi\phi} + pf^{j+1}_{\phi\phi} \\ &= (1-p)f^j_{\phi\phi} + pf^{j+1}_{\phi\phi} \\ &= (1-p)f^{j-1} + pf^j \end{aligned}$$

completes the proof showing that scaling wavelet coefficients by p at pyramid level j is equivalent to linearly interpolating pixel values of scaled subimages f^{j-1}, f^j through MIP-mapping. □

4 RESOLUTION MAPPING

As stated previously, ROI-based reconstruction of the image from its wavelet transformation relies on the choice of a mapping function. The mapping function, denoted by l, maps resolution from the multiresolution pyramid to *image space*. The choice of a mapping function is a precursor to reconstruction of the image and is crucial to the the final representation of the image. It is important to note that resolution information in the pyramid is distributed nonlinearly (by decreasing powers of 2 if the multiresolution pyramid is dyadic in nature). Since reconstruction is carried out in image space (dependent on the pixel location (x, y) in the final image), the resultant percent resolution distribution is obtained by taking the inverse of the constant 2 raised to the mapping function, i.e., % resolution$/100 = 1/2^l$.[2] In the current implementation, three mapping functions are utilized: linear, nonlinear, and empirical HVS acuity-matching. 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 and equal. Although this is not a restriction imposed by the image reconstruction, the size of each ROI is maintained consistently across mapping functions so that different peripheral degradation methods could be readily compared. The specification of the three mapping functions is omitted here, but can be found elsewhere.[4,6]

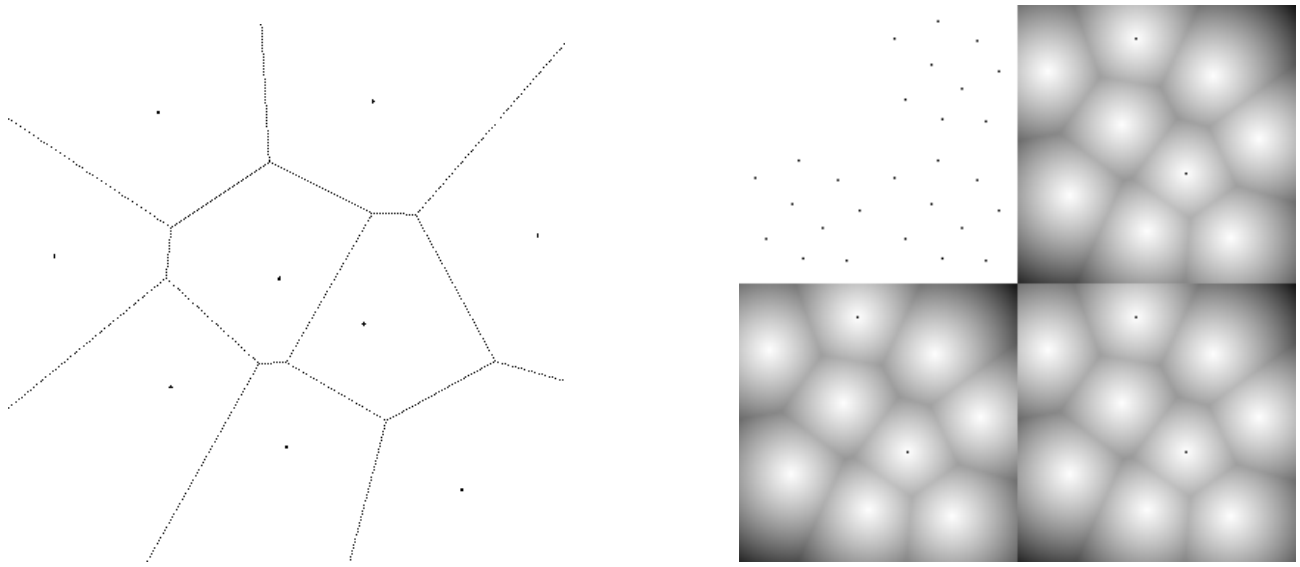
4.1 Multiple ROI Image Segmentation

To include multiple ROIs within the reconstructed image, the image is partitioned into multiple regions. Image filtering is performed on a per-pixel basis, where the desired resolution at each pixel location is determined by the mapping function, relative to the center of only one of multiple ROIs. To select the appropriate ROI, each pixel is subjected to a membership test. This test involves measuring the distance from the pixel location to each ROI center. Using the Euclidian distance metric, the resolution level of the pixel is determined by the mapping function with respect to the closest ROI center.

Formally, the set $S = \{p_1, \ldots, p_n\}$ of n points in the plane, defined by ROI centers, defines a partition of the plane into n regions $V_1, \ldots, V_n$ such that any pixel in the region V_i is closer to the point p_i than to any other $p_j \in S$. This definition of the planar partitioning specifies the *Voronoi diagram* where each V_i is a convex polygonal region called the *Voronoi polygon* of the point p_i in S (an alternate definition and construction of the Voronoi diagram is also available[11]). Voronoi diagrams have found diverse purposes in a number of disciplines. For example, in archaeology, Voronoi polygons are used to map the spread of the use of tools in ancient cultures, and in ecology, the Voronoi diagram of various territorial animals is used to investigate the effects of overcrowding.[11] In image processing, the Voronoi diagram has recently been employed to order codewords in a principal component analysis codebook search algorithm for a vector quantization based coding system.[8] An example of the Voronoi diagram is shown in Figure 2(a).

A graphic representation of wavelet coefficient scaling (as discussed in §3.1) of an arbitrary image at two resolution levels is shown in Figure 2(b). White regions represents coefficients scaled by constant 1, black regions represent coefficient decimation (scaling by 0), and intermediate regions are scaled by linearly interpolated values in the interval $(0, 1)$. Note that the boundaries between linearly interpolated regions, i.e., boundaries between ROIs, are (by construction) Voronoi edges. In other words, Regions Of Interest, where wavelet coefficients are scaled producing spatially degrading resolution, constitute Voronoi polygons about the ROI centers. Although Voronoi partitioning may have little to do with the Human Visual System, it is nevertheless an elegant and natural way to partition the image plane.

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



(a) Voronoi planar partitioning.

(b) Two-level wavelet coefficient scaling.

Figure 2: Example of Voronoi partitioning.

5 RESULTS

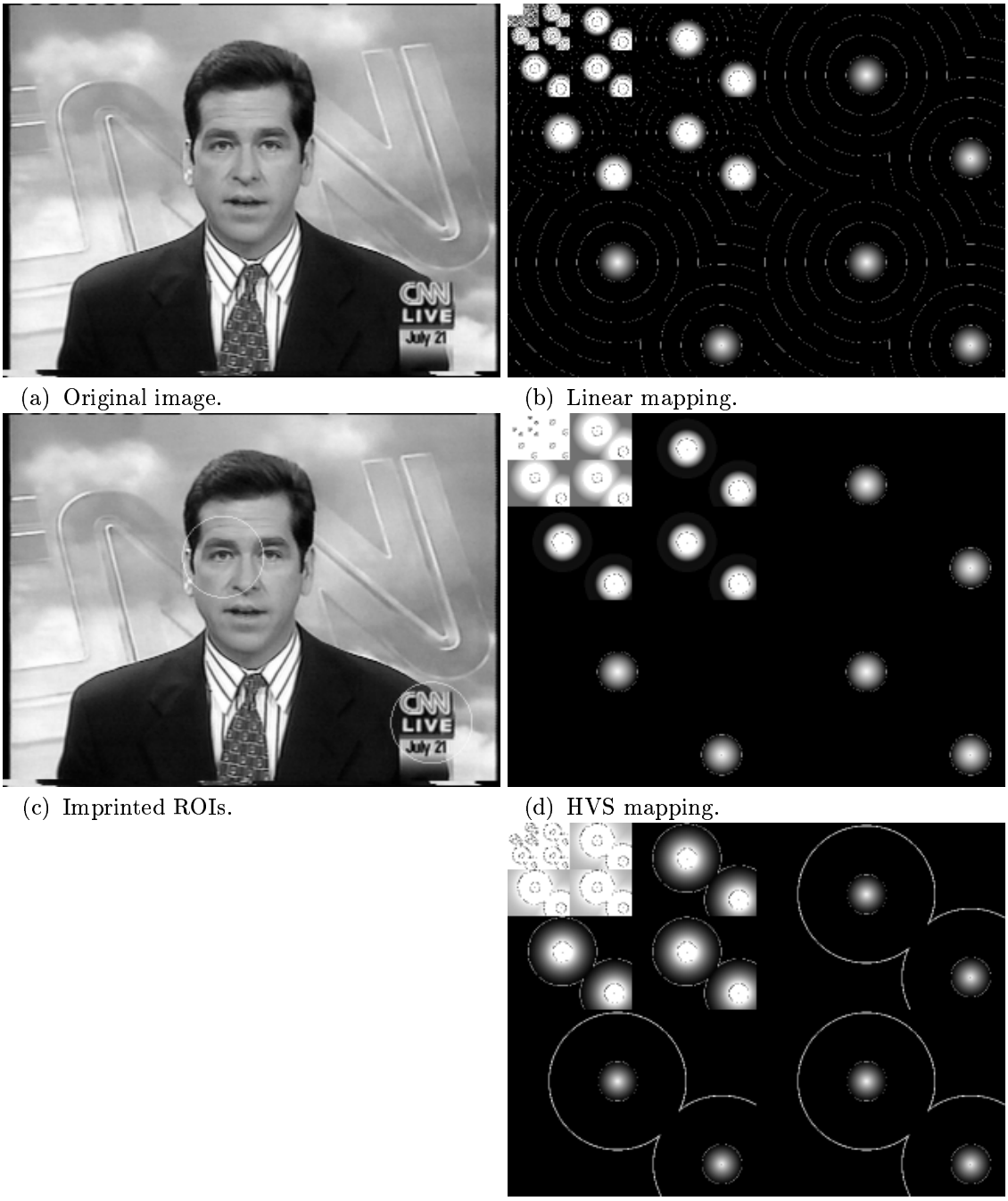
Variable wavelet reconstruction results are presented through example images and subjective quality impairment analysis. In the next section, an image from the *cnn* video sequence is showcased with two artificially placed Regions Of Interest. Images processed with Haar and Daubechies-6 wavelets are shown for comparison. In the following section, analytical results are summarized from a gaze-contingent video display experiment performed with human subjects.

5.1 Multiple ROI Image Reconstruction Examples

Examples of the variable resolution wavelet scaling technique are shown in Figure 3. The *cnn* image was processed with two artificially placed ROIs over the anchor's right eye and the "timebox" found in the bottom right corner of the image. Figure 3(a) contains the original image and Figure 3(c) shows circles imprinted over the ROIs. Figure 3(b), (d), and (e) show the extent of wavelet coefficient scaling in frequency space. Notice the distribution of the concentric resolution bands. In the linear mapping, resolution bands are brought together to generate sharp degradation with respect to ROI centers. The nonlinear mapping spreads out resolution bands resulting in gradual degradation.

The upper left quadrants of the wavelet space images show the wavelet scaling within the lower frequency bands of the wavelet transform representation of the image. Notice the extent of the black regions symbolizing the degree of wavelet coefficient decimation. In the linear mapping, coefficients are decimated at levels 1 (the bottom of the pyramid, or highest frequency bands), 2, 3, and 4. In the nonlinear mapping, few coefficients are decimated at level 3 suggesting that more information is preserved.

Reconstructed images are shown in Figure 4. Note the texture of the anchor man's tie, and the resolution of the



(a) Original image.

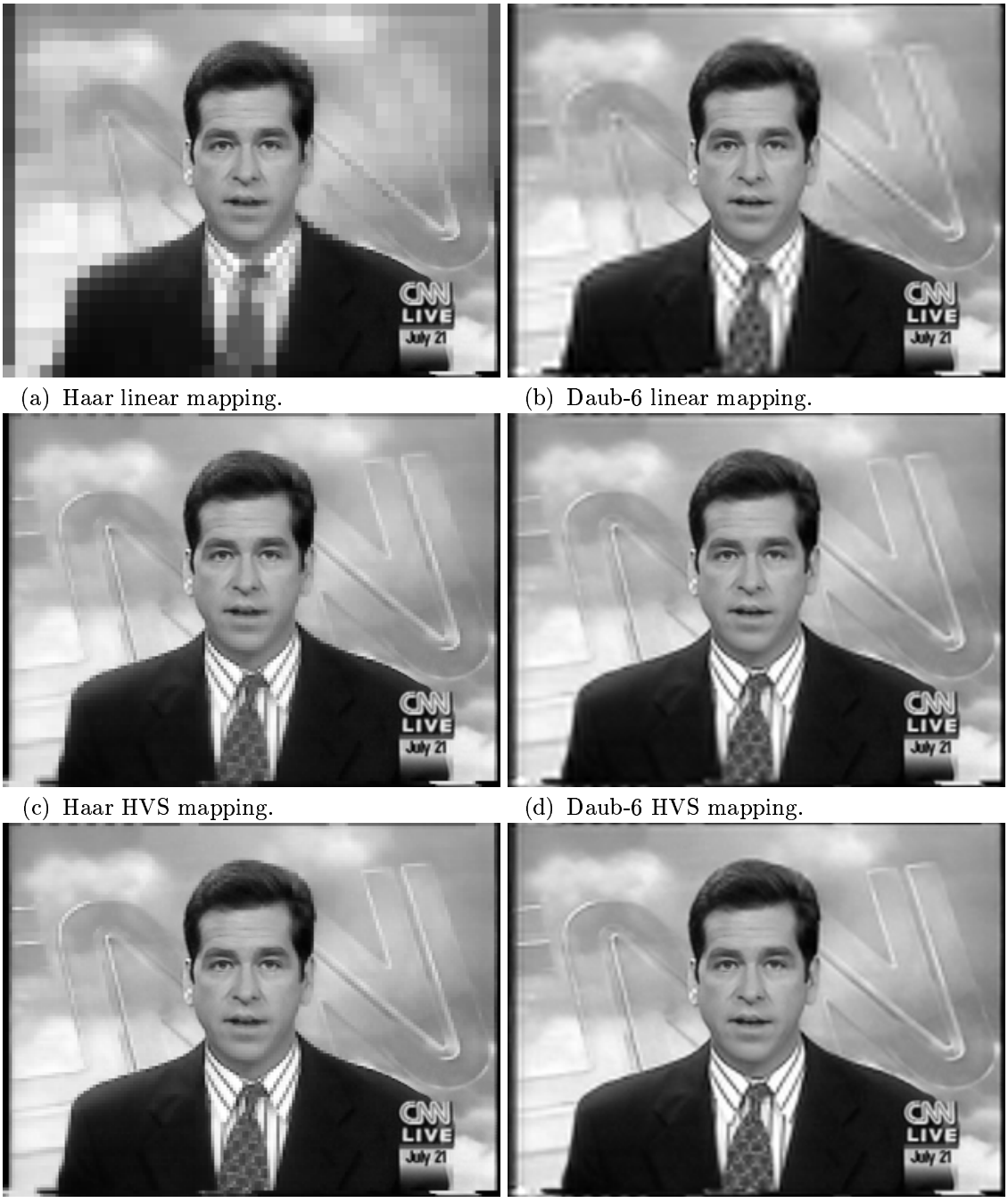
(b) Linear mapping.

(c) Imprinted ROIs.

(d) HVS mapping.

(e) Nonlinear mapping.

Figure 3: Wavelet coefficient resolution mapping (assuming 50dpi screen resolution).



(a) Haar linear mapping. (b) Daub-6 linear mapping.

(c) Haar HVS mapping. (d) Daub-6 HVS mapping.

(e) Haar nonlinear mapping. (f) Daub-6 nonlinear mapping.

Figure 4: Image reconstruction (assuming 50dpi screen resolution).

anchor's right shoulder. Due to the length of the Daubechies-6 wavelets, more information is contained within wavelet coefficients generating a smoother resolution reconstruction. Conversely, blocking artifacts are easily detected in the images processes with the length-2 Haar wavelets. The Haar images are provided as a reference when comparing resolution degradation in the images processed with Daubechies-6 wavelets.

5.2 Perception Impairment Analysis

Eye tracking experiments were carried out in the Virtual Environments Laboratory, Department of Computer Science, Texas A&M University. Video sequences (8-second, grayscale), subject to three resolution degradation mapping functions (linear, human visual system, and unprocessed, denoted by LIN, HVS, and ORG, respectively) were shown to viewers in random order. Each trial consisted of three presentations of one video sequence processed by the LIN, HVS, or ORG mapping, each sequence processed with Daubechies-6 wavelets.

The specific hypothesis tested was whether the spatial resolution degradation of the HVS mapping is imperceptible under *visual tracking* and *free viewing* ROI placement strategies. The former condition presents a single ROI over an easily identifiable target, determined *a priori*. The latter condition displays multiple ROIs over regions fixated previously by subjects under the same condition over the unprocessed sequence. Subjective impairment ratings were scrutinized for (1) overall accuracy of the eye tracker, (2) accuracy of the eye tracker during the viewing task, and (3) accuracy of gaze position over the intended Regions Of Interest. Details of the analysis are reported elsewhere.[5]

Perception impairment analysis over all conditions suggests no variability across resolution mappings (LIN, HVS, ORG), but indicates strong evidence ($p < 0.0000$) of variability across both viewing conditions (visual tracking and free viewing). Analysis within viewing conditions suggest no significant difference between subjective quality ratings under the visual tracking paradigm. In the free viewing case, a significant difference ($p < 0.01$) between subjective ratings suggests the LIN mapping differs considerably from both the HVS and ORG mappings. A comparison between each pair of mapping effects supports this finding.

As expected, results suggests the HVS resolution mapping produced imperceptible video degradation under the visual tracking paradigm. Surprisingly, the linear mapping (LIN) also generated imperceptible results within the same condition. Linear mapping effects were noticeable under the natural viewing condition. It should be noted that, within this condition, it is plausible the method of ROI placement is at least partially responsible for generating perceptible image impairments. The ROI placement method depended on an aggregate account of multiple viewers' fixation locations collected in a previous experiment. Since such a historical record of eye movements is known to be a poor predictor of future fixation locations, this method is likely to be at least partially responsible for the cause of perceptible effects. Since no significant difference was found between the HVS and ORG mappings under the free viewing condition, a plausible explanation for perceptible effects in this experiment was not so much the resolution mapping strategy, but the inadequate method of gaze prediction.

6 CONCLUSION

In this paper, a wavelet image filtering scheme is presented which preserves high (original) resolution in Regions Of Interest matching foveal vision, and gradually degrades resolution in the periphery. The method permits the representation of multiple ROIs extending previous work based on MIP-mapping to the wavelet domain. Degradation is achieved through wavelet coefficient scaling following Voronoi partitioning of the image plane. Three

variants of peripheral degradation are offered, including one matching human visual acuity.

Evidence from subjective experiments suggests that under the visual tracking condition, a significant amount of information may be withheld in gaze-contingent visual displays with little perceptible effect. Surprisingly, results suggest that effects of both linear and acuity-matching resolution mapping are imperceptible. Dyadic linear mapping results in roughly 50% resolution degradation over 97% of the image (50% resolution beyond the 105×105 foveal ROI over a 640×480 image frame). This information reduction has significant implications for gaze-contingent image and video compression.

7 REFERENCES

[1] Peter J. Burt and Edward H. Adelson. The Laplacian Pyramid as a Compact Image Code. *IEEE Transactions on Communications*, 31(4):532–540, April 1983.

[2] Kenneth R. Castleman. *Digital Image Processing.* Prentice-Hall, Inc., Englewood Cliffs, NJ, 1996.

[3] Charles K. Chui. *An Introduction to Wavelets.* Academic Press, Inc., San Diego, CA, 1992.

[4] Andrew T. Duchowski and Bruce H. McCormick. Simple Multiresolution Approach for Representing Multiple Regions of Interest (ROIs). In *Visual Communications and Image Processing*, pages 175–186, Taipei, Taiwan, May 1995. SPIE.

[5] Andrew T. Duchowski and Bruce H. McCormick. Gaze-Contingent Video Resolution Degradation. In *Human Vision and Electronic Imaging II*, Bellingham, WA, January 1998. SPIE.

[6] Andrew Ted Duchowski. *Gaze-Contingent Visual Communication.* PhD thesis, Texas A&M University, College Station, TX, 1997.

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

[8] Yih-Chuan Lin and Shen-Chuan Tai. Dynamic windowed codebook search algorithm in vector quantization. *Optical Engineering*, 35(10):2921–2929, October 1996.

[9] Stephane G. Mallat. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 11(7):674–693, July 1989.

[10] 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.

[11] Franco P. Preparata and Michael Ian Shamos. *Computational Geometry: An Introduction.* Springer-Verlag, New York, NY, 1985.

[12] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, Cambridge, 2nd edition, 1992.

[13] S. Tanimoto and A. Klinger, editors. *Structured Computer Vision.* Academic Press, New York, NY, 1980.

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

[15] Lance Williams. Pyramidal Parametrics. *Computer Graphics*, 17(3):1–11, July 1983.