

TECHNICAL RESEARCH REPORT

Compact Image Coding from Multiscale Edges

by S. Dalbague, J. Baras, N. Sidiropoulos

CSHCN T.R. 98-15
(ISR T.R. 98-61)



The Center for Satellite and Hybrid Communication Networks is a NASA-sponsored Commercial Space Center also supported by the Department of Defense (DOD), industry, the State of Maryland, the University of Maryland and the Institute for Systems Research. This document is a technical report in the CSHCN series originating at the University of Maryland.

Web site <http://www.isr.umd.edu/CSHCN/>

Compact Image Coding From Multiscale Edges

S. Dalbague, J.S. Baras and N.D. Sidiropoulos
Center for Satellite and Hybrid Communication Networks
Institute for Systems Research
University of Maryland
College Park, MD 20742

Contents

1	Introduction	1
1.1	Description of the scheme	2
2	The Wavelet Transform	4
2.1	Introduction	4
2.2	Mathematical definitions for one-dimensional signals	4
2.2.1	The Multiresolution Transform	4
2.2.2	The Orthogonal Wavelet Representation	5
2.2.3	Implementation of an Orthogonal Wavelet Representation	5
2.3	Wavelet Transform for Images	6
3	The Wavelet Maxima Representation	8
3.1	Why such a representation ?	8
3.2	Edge detection and Wavelet Transform	9
3.3	Computation of the Wavelet Maxima Representation	11
3.4	The Reconstruction Algorithm	11
3.5	The orthogonal projection on the vector space Γ	12
4	Compact Image Coding	14
4.1	Perceptual Ranking of Information	14
4.2	Construction of the curves	14
4.3	Curve-shifting procedure and curve selection	15
4.4	Selection on the curves	15

4.5	Curve coding	16
4.6	Results	17
5	Measures of the perceptual quality of reconstructed images	17
5.1	Introduction	17
5.2	Weighted pixel-to-pixel differences	18
5.3	The visible differences predictor	19
5.3.1	Introduction	19
5.3.2	Amplitude non-linearity function	19
5.3.3	Contrast Sensitivity function	21
5.3.4	Spatial frequency hierarchy	22
5.3.5	Masking function	23
5.3.6	Psychometric function and probability summation	24
5.4	Perceptual distortion of reconstructed images	25
5.5	Results	25
6	Conclusions	27

List of Figures

1	Global diagram of the compression scheme	3
2	Filter banks used for the Wavelet Analysis/Synthesis	7
3	Finite impulse response of the spline quadratic filters	8
4	An example of the dyadic Wavelet Transform of an image	8

5	The Wavelet Representation	10
6	The Wavelet Maxima Representation at three resolutions	11
7	Curve shifting procedure	15
8	Curve-shifting procedure and curve selection	16
9	Reconstructed images at compression ratios 12:1 and 19:1	18
10	The three main components of the HVS model within the VDP (reprinted with permission of MIT Press from [6])	20
11	The four components of the detection mechanism in VDP (reprinted with permission of MIT Press from [6])	20
12	Amplitude non-linearity and contrast sensitivity functions (reprinted with permission of MIT Press from [6])	21
13	The formation of the cortex filters (reprinted with permission of MIT Press from [6])	22
14	Two probability mappings obtained for 6:1 and 18:1 compression ratios	25
15	Perceptual distortion of reconstructed images from our coding scheme	26
16	Perceptual distortion of reconstructed images from JPEG	26

Abstract

In this paper we investigate image compression from a perceptual viewpoint. We develop a detailed image compression scheme using multiscale edges of the image, based on wavelet maxima representations. We also develop a detailed implementation of the Human Visual System (HVS) performance evaluation model. The HVS model includes all key components of the physiological human visual model, and it provides a performance evaluation of image compression linked to perceptual criteria. Using the HVS we show that image compression based on multiscale edges gives results inferior to standard schemes such as JPEG.

1 Introduction

The inexorable advance of technology for the processing and display of electronic imagery has in recent years offered the possibility of radically new means of communicating visual information. Bandwidth, memory, and computational resources are limiting factors for all these systems and in every instance directly affect their cost, quality, and practicality.

Therefore, image compression is now essential for applications such as transmission and storage in databases. Several successful generic methods have been developed such as JPEG. These methods have some disadvantages. First they do not take into account the multiresolution composition of the image, i.e., do not use such a decomposition to improve the efficiency of the coding scheme. Furthermore, they are not well suited for multicasting because they encode the image as a whole and do not separate the information corresponding to approximations of the image at different resolutions.

Thus, multiresolution schemes have received much interest during the past few years. One of the most promising methods is to use a Wavelet Representation of the image and Vector Quantization of the wavelet coefficients using a bit allocation scheme that matches the human visual system characteristics (See [12], [14], [15] and [16]).

Here we investigate a new method based on multiscale edges of the image, because the structural information required for recognition tasks is in most cases provided by edges. This is particularly well illustrated by our ability to recognize objects from drawings that outline the contours. We therefore use a multiscale edge-based representation of the image to select the important information for coding (See [3], [4] and [5]).

Furthermore, since the human eye is the final judge of visual image quality, it is essential to understand the human visual system in order to understand the important parameters of image quality. One realizes the need for a perceptual measure of image quality, that would fit human judgment in more accurate way than usual mathematical criteria. Such a perceptual criterion would be very useful for the optimization of coding schemes, because it would point out the deficiencies of the schemes according to the eye requirements (See [6]).

That is the reason why we study several perceptual distortion measures, which we test on the reconstructed images from the presented coding scheme.

1.1 Description of the scheme

In this report, we present an image compression scheme based on a multiscale edge-based representation of the image: the Wavelet Maxima Representation (WMR).

This representation is computed from the Wavelet Representation of the image. From the Wavelet Representation, we keep the coarse resolution image; and for each detail image of the different resolution levels, we keep the values of the wavelet coefficients at the locations of the local modulus maxima along the gradient direction. This coding scheme involves two steps: first we select the information that is considered important for visual image quality, then we efficiently code this information only.

We use the Wavelet Maxima Representation to perform the coding of the image because local modulus maxima of the wavelet coefficients correspond to sharp variation points in the image, i.e they correspond to edges. In order to achieve interesting compression ratios, we build curves from the points of the Wavelet Maxima Representation. This allows us to make an easier selection of the information to encode; and the curve structure allows chain coding.

To compress the computed edge information, we will use different coding methods depending on the nature of the curve information we are coding. The main types of curve information will be point locations, modulus and argument values and some reference information such as the location of the first point of each edge etc. Certain types of information are more important than others. We thus use a combination of lossless and lossy methods.

This report begins by reviewing the Wavelet Representation and the extraction of multiscale edges (including the Wavelet Maxima Representation and associated approximate reconstruction). We also review a recently proposed model of the human visual system (HVS), and discuss how it may be used to assess image quality. These background concepts and methods set the stage for the subsequent presentation of a novel HMR edge map coding technique and its performance evaluation using the above HVS model. The structure of the overall image compression scheme is shown in Figure 1.

Lastly, we will present an algorithm whose goal is to predict the difference between the reconstructed image and the original one that will be actually perceived by the eyes of the observer. This algorithm uses an HVS (Human Visual System) model to predict this perceptual difference. We use this algorithm to assess the quality of the reconstructed images of our coding scheme and we make some comparison with other coding methods.

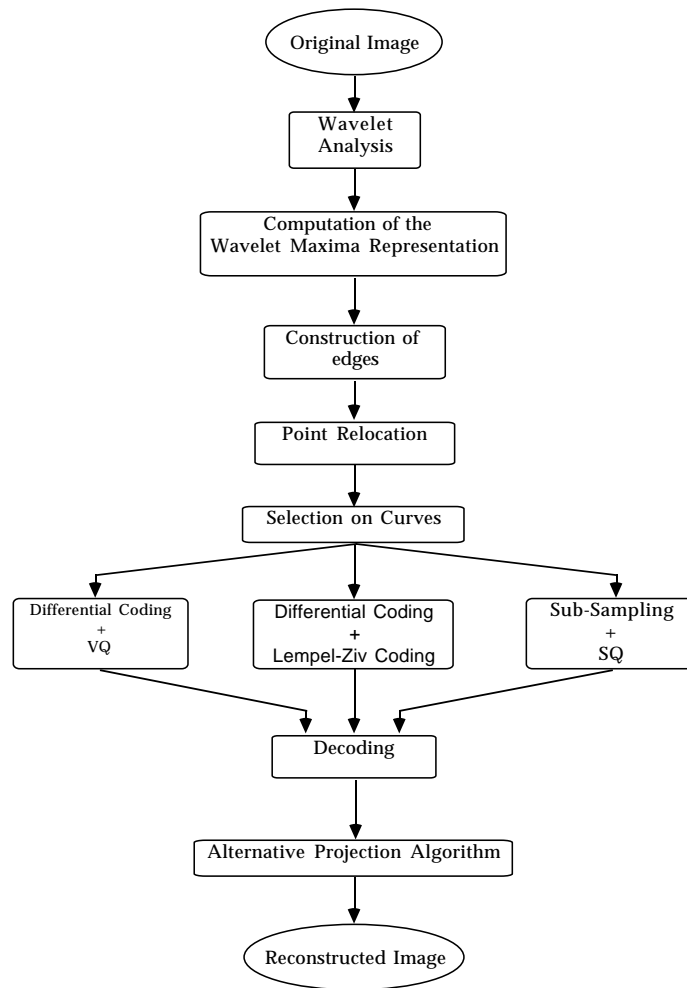


Figure 1: Global diagram of the compression scheme

2 The Wavelet Transform

2.1 Introduction

The Wavelet Transform is a mathematical transform that performs a multiresolution decomposition of a signal or an image. It has already found many fields of applications [1] [2].

The Wavelet Transform reorganizes the signal information into a set of details appearing at different resolutions. It splits the signal into subbands (2 for a 1-D signal and 4 for an image) which are located in the space-frequency domain. Therefore the decomposition of a signal onto an orthonormal wavelet basis gives an intermediate representation between Fourier and spatial representations.

Depending on the subband one considers, the transformed signal will contain more or less details. Since the location information of these details is maintained, the transformed signal subbands will maintain varying degrees (depending on the subband) of resemblance to the original signal. One may iterate the Wavelet Analysis by splitting further the coarse resolution image.

The main properties of the Orthogonal Wavelet Transform are the following:

- It decomposes the signal or the image onto an orthogonal basis and therefore removes redundancy. This is really important if the Wavelet Transform is to be followed by a coding stage. The information is scattered among the different subbands in an optimal way.
- The Orthogonal Wavelet Transform can be implemented using simple FIR filters, this is very convenient.
- We can even use linear phase filters so that we can cascade them for a faster result.

In the following sections, we restrict ourselves to the case of the dyadic wavelet transform, which is used in most practical applications. The dyadic Wavelet Transform involves scales of the form 2^j and therefore splits a signal onto octave subbands.

2.2 Mathematical definitions for one-dimensional signals

2.2.1 The Multiresolution Transform

Let A_{2^j} be the linear operator which approximates a signal at resolution 2^j . A_{2^j} is the orthogonal projection on the vector space V_{2^j} which is the set of all possible approximations at resolution 2^j of functions in $L^2(\mathbb{R})$ (Square integrable functions). When moving from $f(x)$ to $A_{2^j}f(x)$ some information is lost but

$$\lim_{j \rightarrow \infty} A_{2^j}f(x) = f(x)$$

We can build an orthonormal basis of the vector space V_{2^j} by dilating and translating a unique function $\phi(x)$ which is called a scaling function. Thus $(\sqrt{2^{-j}} \phi_{2^j}(x - 2^{-j}n))_{n \in \mathbb{Z}}$ is an orthonormal basis of V_{2^j} , with $\phi_{2^j}(x) = 2^j \phi(2^j x)$. The orthonormal projection on V_{2^j} can be computed by decomposing the signal on this basis:

$$A_{2^j}f = (A_{2^j}f(n))_{n \in \mathbb{Z}} = ((f(u) * \phi_{2^j}(-u))(2^{-j}n))_{n \in \mathbb{Z}}$$

$(A_{2^j}^d f(n))_{n \in \mathbb{Z}}$ is the discrete approximation of $f(x)$ at resolution 2^j .

And all the approximations $A_{2^j}^d f$ for $j < 0$ can be computed from $A_1^d f$ by repeating this process.

2.2.2 The Orthogonal Wavelet Representation

The Orthogonal Wavelet Transform extracts the difference of information between two approximations at resolutions 2^{j+1} and 2^j by decomposing the signal onto an orthonormal wavelet basis. This difference of information is called the detail signal at resolution 2^j . It also corresponds to the difference between the two orthogonal projections on V_{2^j} and $V_{2^{j+1}}$ and can be considered as the projection on O_{2^j} where $O_{2^j} \oplus V_{2^j} = V_{2^{j+1}}$. To compute the projection on O_{2^j} of $f(x)$, we need an orthonormal basis of O_{2^j} . Such a basis can be built by scaling and translating a function $\psi(x)$ which is called the mother Wavelet: $(\sqrt{2^{-j}} \psi_{2^j}(x - 2^{-j}n))_{n \in \mathbb{Z}}$ is an orthonormal basis of O_{2^j} , with $\psi_{2^j}(x) = 2^j \psi(2^j x)$. We can then compute the discrete detail signal at the resolution 2^j

$$D_{2^j} f = (D_{2^j} f(n))_{n \in \mathbb{Z}} = \langle f(u), \psi_{2^j}(u - 2^{-j}n) \rangle_{n \in \mathbb{Z}} = ((f(u) * \psi_{2^j}(-u))(2^{-j}n))_{n \in \mathbb{Z}}$$

which contains the difference of information between $A_{2^{j+1}}$ and A_{2^j} .

Finally, for any $J > 0$, the original discrete signal $A_1^d f$ measured at the resolution 1 is represented by $(A_{2^{-j}}^d f, D_{2^j} f)_{-J \leq j \leq -1}$ which is called the Orthogonal Wavelet Representation of the signal. It consists of:

- The reference signal at a coarse resolution 2^{-J} : $A_{2^{-J}}^d f$
- The detail signals at resolutions 2^j for $-J \leq j \leq -1$

2.2.3 Implementation of an Orthogonal Wavelet Representation

We use a pyramidal algorithm to compute the Orthogonal Wavelet Representation.

We consider:

$$\begin{cases} D_{2^j} f(n) = \langle \psi_{2^j}(t), f \rangle \\ A_{2^j} f(n) = \langle \phi_{2^j}(t), f \rangle \end{cases}$$

We can rewrite these two equations as:

$$\begin{cases} D_{2^j} f(n) = \sum_k g(2n - k) A_{2^{j+1}}(k) \\ A_{2^j} f(n) = \sum_k h(2n - k) A_{2^{j+1}}(k) \end{cases}$$

with $h(n) = \frac{1}{\sqrt{2}} \int \phi(2x - n) \phi(2x) dx$ and $g(n) = \frac{1}{\sqrt{2}} \int \psi(2x - n) \phi(2x) dx$.

Therefore we have $g(n) = (-1)^n h(1 - n)$ and $A_0 f(k) = f(k)$.

G is the mirror filter of H and the filter bank is named Quadrature Mirror Filter bank (QMF). G is a high-pass filter whereas H is a low-pass filter.

The exact reconstruction is computed from the formula:

$$A_{2^{j+1}} f(l) = \sum_k [h(2n - l) A_{2^j} f(n) + g(2n - l) D_{2^j} f(n)]$$

The original image can thus be reconstructed from its Wavelet Representation by iterating this process.

There are some additional conditions that these filters should satisfy in order to obtain a useful decomposition. The wavelet coefficients resulting from the wavelet analysis should be relatively smooth. Several trade-offs should be taken into account:

- In our application, we do not want to introduce any redundancy within the Representation because it is to be followed by a coding step. Therefore ψ and ϕ must give orthogonal basis of V and O so that we get decorrelated representations on these spaces.
- We want the Wavelet Transform to be fast to compute. Therefore we would like to implement it with short compact support filters.
- Lastly, we need Linear Phase filters to be able to use a pyramidal structure, that is to say to be able to cascade them.

The only solution that fulfills these three conditions is the Haar Wavelet ($g(0) = h(0) = h(1) = \sqrt{2} = -g(1)$) which is not smooth enough to give good results. Therefore we have to relax one condition of orthogonality on the wavelets. We keep $V_{2^j} \oplus O_{2^j}$ and we relax $V_{2^j} \perp O_{2^j}$. We then obtain what are called Biorthogonal Wavelets.

This allows us to build wavelets smooth enough (in terms of vanishing moments and discontinuity of the derivative of the wavelet).

In order to maintain the perfect reconstruction property, the orthogonality condition implies that we need to use different filters for analysis and synthesis; the reconstruction formula becomes:

$$A_{2^{j+1}}f(l) = \sum_k [\tilde{h}(2n-l)A_{2^j}f(n) + \tilde{g}(2n-l)D_{2^j}f(n)]$$

We can now compute the Wavelet Representation of the signal with FIR filters which have the property of being symmetric (Linear Phase Filters). The resulting filter banks for analysis and synthesis with biorthogonal wavelets are shown in Figure 2.

2.3 Wavelet Transform for Images

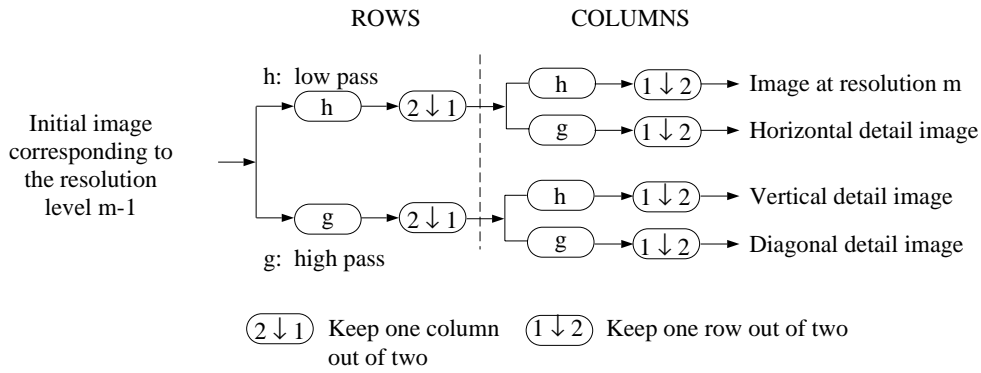
The extension of the wavelet transform to 2-dimensional spaces consists of building 4 wavelet/scaling functions. For practical purposes, and because of the importance of horizontal and vertical directions in man-made images, one typically uses separable wavelets, which can be written as the product of two one-dimensional wavelets:

$$\begin{aligned} \phi(x, y) &= \phi(x)\phi(y) \\ \psi^H(x, y) &= \psi(x)\phi(y) \\ \psi^V(x, y) &= \phi(x)\psi(y) \\ \psi^D(x, y) &= \psi(x)\psi(y) \end{aligned}$$

$\psi^H(x, y)$, $\psi^V(x, y)$ and $\psi^D(x, y)$ correspond respectively to high horizontal frequency, high vertical frequency and high diagonal frequency images.

In order to solve border problems, we introduce symmetries and periodization in the image. We make the assumption that the image is symmetrical with respect to

Wavelet Analysis



Wavelet Synthesis

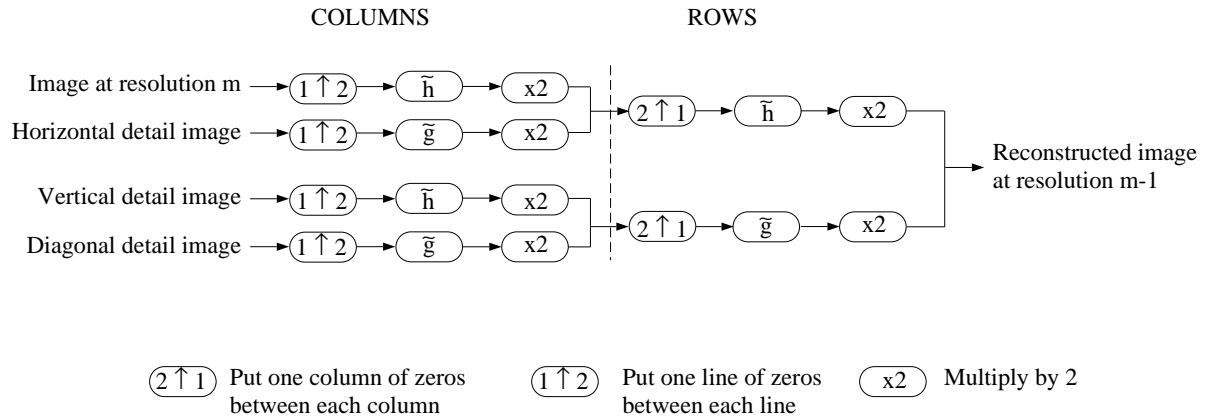


Figure 2: Filter banks used for the Wavelet Analysis/Synthesis

each of its borders, and is periodic with a period equal to twice the number of its pixels in each direction. An example of the wavelet transform of an image is shown in Figure 4.

For our implementation, we use Mallat’s spline quadratic filters [3]. In Figure 3 we show the impulse responses of typical such filters used. They were chosen because of their smoothness and their small compact support.

n	-1	0	1	2
H	0.125	0.375	0.375	0.125
G		-2.0	2.0	

Figure 3: Finite impulse response of the spline quadratic filters

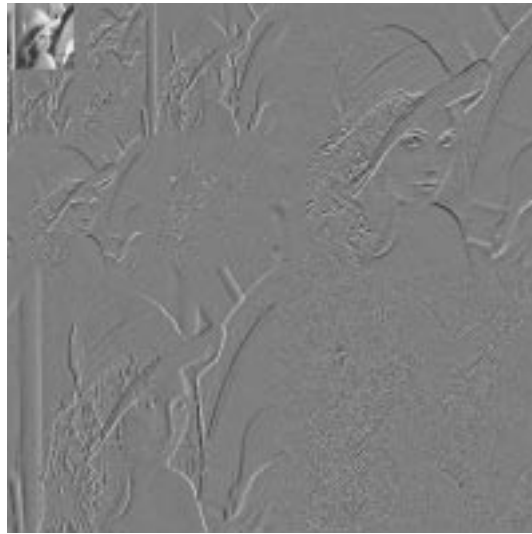


Figure 4: An example of the dyadic Wavelet Transform of an image

3 The Wavelet Maxima Representation

3.1 Why such a representation ?

We use the Wavelet Maxima Representation to select the important information and then encode it because the local modulus maxima of the Wavelet Transform of the image correspond to sharp variation points, that is to say image edges. Therefore if we use this representation, we will code the part of the image content that corresponds to the edges and also to the main structures of the image.

From this representation, one can reconstruct an image that is very close to the original one. In fact, one can obtain a perceptually perfect reconstruction, that is to say, one cannot see the difference between the original image and the reconstructed one [8] [9] [11]. Furthermore this representation is well-adapted to the goals of image

compression, because once the edges have been detected in the image, it is easier to select the important information and then encode this remaining information. We can indeed use the average modulus and the length of the curves built from the Wavelet Maxima Representation as thresholds to make the selection.

Several studies have been conducted on the subject of edge detection based on zero-crossings of the Wavelet Transform. This method uses the second derivative of the smoothed signal at scale s whose zero-crossings correspond to the extrema of the first derivative and to the inflection points of the smoothed signal.

The two methods using respectively the extrema and the zero-crossings of the Wavelet Transform are similar but the one using the extrema exhibits some advantages:

- First it is possible to make a distinction between two types of extrema : it can be a maximum or a minimum of the absolute value. Maxima correspond to sharp variations of the smoothed signal, whereas minima correspond to slow variations of the image. Therefore, with the extrema method, it is easy to keep only the sharp variations of the image if we only keep the local maxima of the Wavelet Transform.
- Second, it is possible to record the absolute value at the maxima locations, which is the absolute value of the derivative at the inflection points of the smoothed image. These values will be very helpful for the subsequent selection of points in the Maxima Wavelet Representation.

3.2 Edge detection and Wavelet Transform

Many edge detection approaches are variants of the following general method: first, smooth the image at different scales. Then, detect sharp variation points from the first or second derivative of the smoothed image at several scales.

Let us see how these detectors and the Wavelet Transform can be related. We call "smoothing function" any function $\theta(x, y)$ such that

$$\iint \theta(x, y) dx dy = 1 \quad \text{and} \quad \lim_{x, y \rightarrow \infty} \theta(x, y) = 0$$

Let $\epsilon_s(x, y) = \frac{1}{s^2} \epsilon(\frac{x}{s}, \frac{y}{s})$ be the dilation of function $\epsilon(x, y)$ by the factor s . Smoothed images at different scales are obtained by convolving the original image with $\theta_s(x, y) = \frac{1}{s^2} \theta(\frac{x}{s}, \frac{y}{s})$ which is the dilation by s of the smoothing function $\theta(x, y)$. Then, for each smoothed image and for each point (x, y) , we can compute the gradient vector $\vec{\nabla}(f * \theta_s)(x, y)$. The direction of the gradient vector indicates the direction in the image plane along which the directional derivative of $f(x, y)$ has the largest absolute value. Therefore edges are defined as points where the modulus value is maximum in the direction given by the gradient vector.

Let us now link this edge detection with the Wavelet Transform of the image. We define two Wavelet functions $\psi^1(x, y)$ and $\psi^2(x, y)$ such that

$$\psi^1(x, y) = \frac{\partial \theta(x, y)}{\partial x} \quad \text{and} \quad \psi^2(x, y) = \frac{\partial \theta(x, y)}{\partial y}$$

Then the Wavelet Transform of $f(x, y)$ at the scale s is defined by the two components

$$W_s^1(x, y) = f * \psi_s^1(x, y) \quad \text{and} \quad W_s^2(x, y) = f * \psi_s^2(x, y)$$

And we finally have

$$\begin{pmatrix} W_s^1 f(x, y) \\ W_s^2 f(x, y) \end{pmatrix} = s \begin{pmatrix} \frac{\partial}{\partial x}(f * \theta_s)(x, y) \\ \frac{\partial}{\partial y}(f * \theta_s)(x, y) \end{pmatrix} = s \vec{\nabla}(f * \theta_s)(x, y)$$

Therefore edge points can be located from the two components of the wavelet transform. They correspond to modulus maxima of the Wavelet Transform of the image along the gradient direction. The modulus and the gradient direction for each scale are given by the two formulae:

$$\begin{cases} \text{Modulus} = \sqrt{W_{1,s}^2 + W_{2,s}^2} \\ \text{Argument} = \arg(W_{1,s} + j W_{2,s}) \end{cases}$$

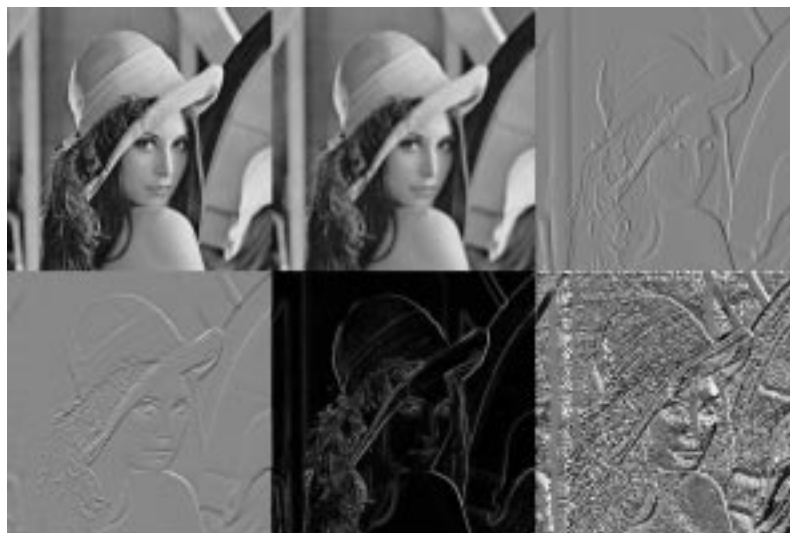


Figure 5: The Wavelet Representation

We illustrate these concepts in the sequence of images presented in Figure 5, of the ubiquitous Lena image. Let us describe the different images from left to right for the top row first and then for the bottom row. The first image is the original Lena image. The second one is the coarse resolution image after one Wavelet Analysis step; it therefore exhibits a resolution that is half the original one. The next two images are the Wavelet transform images that correspond respectively to high horizontal and high vertical frequencies. These three images are part of the Wavelet Transform of the image. The last two images correspond respectively to the modulus and argument images computed from the two previous images. These last two images are used to compute the Wavelet Maxima Representation of the image.

3.3 Computation of the Wavelet Maxima Representation

The Wavelet Maxima Representation is computed from the Wavelet Representation of the image. From this representation, we keep the coarse resolution image. Then for each scale of the Wavelet Representation and, for each pixel in the image, we check whether this pixel is a local modulus maximum along the gradient direction. We compute the modulus and the argument at the pixel from the two formulae above. We discretize the gradient into eight possible directions which correspond to the eight neighboring pixels. Then we check whether the pixel is a local modulus maximum along this direction or not.

Therefore the Wavelet Maxima Representation will be constituted of a coarse resolution image, and for each scale of the Wavelet Representation, of two detail images, one for the modulus, one for the argument. In these images, if the pixel is a local wavelet modulus maximum, the corresponding values will be the modulus and the argument of the Wavelet Representation; otherwise the values will be set to zero.

In Figure 6 we can see the coarse resolution image obtained after three Wavelet Analysis steps and the three remaining pictures show the locations of the Wavelet modulus maxima of the three different scales of the Wavelet Maxima Representation. Observe that, the coarser the resolution is, the fewer modulus maxima we have. The resolution sequence is as follows: right bottom corner coarse, left bottom corner medium, right upper corner fine.

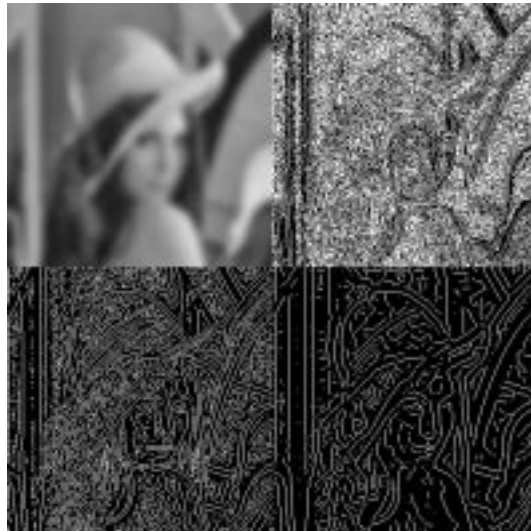


Figure 6: The Wavelet Maxima Representation at three resolutions

3.4 The Reconstruction Algorithm

This algorithm is an extension of the reconstruction algorithm used for one-dimensional signals [3]. We consider an image $f(x, y)$ and its dyadic Wavelet Transform. For each scale 2^j , we have determined the local maxima of $M_{2^j} f(x, y)$ along the direction given by the angle $A_{2^j} f(x, y)$ and stored their locations $(x_v^j, y_v^j)_{v \in R}$ and the values of

$M_{2^j} f(x_v^j, y_v^j)_{v \in R}$ and $A_{2^j} f(x_v^j, y_v^j)_{v \in R}$.

The reconstruction problem is to find the set of functions $h(x, y)$ that satisfy :

- For each scale 2^j at locations $(x_v, y_v)_{v \in R}$ of the maxima:

$$\begin{cases} W_{2^j}^1 h(x_v, y_v) = W_{2^j}^1 f(x_v, y_v) \\ W_{2^j}^2 h(x_v, y_v) = W_{2^j}^2 f(x_v, y_v) \end{cases}$$

- For each scale, the locations of the maxima obtained from $W_{2^j}^1 h$ and $W_{2^j}^2 h$ are $(x_v, y_v)_{v \in R}$

The first constraint means that the Wavelet Transforms of $f(x, y)$ and $h(x, y)$ have to coincide at the maxima locations.

We have to modify slightly the second constraint in order to solve the problem numerically. We thus do not impose that the $(x_v^j, y_v^j)_{v \in R}$ are the only maxima but that $h(x, y)$ minimizes a Sobolev norm:

$$\|h\|^2 = \sum_{j=-\infty}^{+\infty} [\|W_{2^j}^1\|^2 + \|W_{2^j}^2\|^2 + 2^{2j} (\|\frac{\partial W_{2^j}^1 h}{\partial x}\|^2 + \|\frac{\partial W_{2^j}^2 h}{\partial x}\|^2)]$$

This constraint, combined with the first one, will tend to introduce maxima at the $(x_v, y_v)_{v \in R}$ locations. (Due to the partial derivatives in the norm, we will obtain as few spurious maxima as possible. We use a coefficient that is proportional to the scale to take into account the fact that the smoothness of the image grows with its scale.

To minimize the Sobolev norm, we use an alternating projection algorithm. Let Γ be the space of sequences of functions such that for each index j and each location of maximum (x_v^j, y_v^j) :

$$\begin{cases} W_{2^j}^1 h(x_v^j, y_v^j) = W_{2^j}^1 f(x_v^j, y_v^j) \\ W_{2^j}^2 h(x_v^j, y_v^j) = W_{2^j}^2 f(x_v^j, y_v^j) \end{cases}$$

The space Γ is an affine space that is closed in K , the vector space of sequences of finite norm.

Let V be the space of the dyadic Wavelet Transform of all functions of $L^2(R^2)$. The functions that satisfy the first constraint are the elements of K that belong to $\Lambda = V \cap K$. To reconstruct the element of $V \cap K$ that minimizes the Sobolev norm, we use alternating orthogonal projections on V and Γ with respect to the Sobolev norm. The orthogonal projection on V is $P_v = W \circ W^{-1}$. The orthogonal projection on Γ will be described in the next section.

For N^2 pixels images, the implementations of P_v and P_Γ have a complexity of $O(N^2 \log_2 n)$. If we start the iteration with the zero element of K , the algorithm converges strongly to the element of Λ whose norm is minimal [3].

3.5 The orthogonal projection on the vector space Γ

The operator we use for images is actually the one used for a one-dimensional signal, applied to the rows and columns of the image. Let us first explain the one-dimensional algorithm. P_Γ transforms each sequence $(g_j(x))_{j \in Z} \in K$ into the closest sequence

$(h_j(x))_{j \in \mathbb{Z}} \in \Gamma$ for the Sobolev norm. Let $\epsilon_j(x) = h_j(x) - g_j(x)$. We have chosen each function $h(x)$ so that we minimize:

$$\sum_{j=-\infty}^{+\infty} \|\epsilon_j\|^2 + 2^{2j} \left\| \frac{d\epsilon_j}{dx} \right\|^2 \quad (1)$$

For this, we minimize separately each term $\|\epsilon_j\|^2 + 2^{2j} \left\| \frac{d\epsilon_j}{dx} \right\|^2$ in (1). Let x_0 and x_1 be the abscissa of two consecutive maxima of $W_{2^j} f(x)$.

Since $(h_j(x))_{j \in \mathbb{Z}} \in \Gamma$, we have

$$\begin{cases} \epsilon_j(x_0) = W_{2^j} f(x_0) - g_j(x_0) \\ \epsilon_j(x_1) = W_{2^j} f(x_1) - g_j(x_1) \end{cases} \quad (2)$$

In between x_0 and x_1 , minimizing (1) is the same as minimizing

$$\int_{x_0}^{x_1} (|\epsilon_j(x)|^2 + 2^{2j} \left| \frac{d\epsilon_j(x)}{dx} \right|^2) dx .$$

This minimization problem has the following solution:

$$\epsilon_j(x) = \alpha \exp(2^{-j} x) + \beta \exp(2^{-j} x)$$

where α and β are determined from the boundary conditions (2). However, using this solution would be numerically expensive. Therefore we use an interpolation that approximates this solution while requiring fewer computations.

Let N_{oct} be the number of Wavelet decomposition levels. Let us denote $n = x_1 - x_0$. Let $a = \frac{1}{5.8^{2/scale}}$ where $scale$ is of the form $scale = 2^j$ for $j \in 1 \dots N_{oct}$.

Then we use the formula:

$$\text{for } i \in [x_0, x_1], \quad \epsilon_j(i) = \frac{\epsilon_j(x_0)}{1 - a^{2n}} a^i [1 - a^{2n-2i}] + \frac{\epsilon_j(x_1)}{1 - a^{2n}} a^{n-i} [1 - a^{2i}]$$

We introduce the coefficient a to take into account the fact that smoothness has to grow with increasing scale.

Since we only approximate the solution of the Sobolev norm minimization problem, we have to use a follow-up procedure that removes the spurious maxima created by the projection. For this purpose we proceed on the rows and the columns of the image separately. We consider again two consecutive maxima whose abscissae are x_0 and x_1 . We remove the maxima between these two abscissae after checking that the direction of the maximum to be removed is horizontal if we are processing a row and vertical if we are processing a column.

This algorithm allows us to reconstruct a perceptually almost – perfect image with a PSNR above 39 dB. However, as we discuss next, PSNR is not an accurate metric of the perceptual fidelity of reconstructed images.

4 Compact Image Coding

4.1 Perceptual Ranking of Information

From a theoretical point of view, we have to encode several kinds of image information to be able to reconstruct the image from the Wavelet Maxima Representation. In particular:

- For each scale, the locations of the Wavelet maxima and the values of their modulus and arguments.
- The coarse resolution information as a coarse resolution image.

But if we want to obtain a high compression ratio we cannot afford to keep all the information content. We have to remove the components that are not very important from the visual point of view. Therefore we have to rank the information contained in the Wavelet Maxima Representation, and proceed by encoding the different components of suitable levels of information (approximation) loss.

We use a Wavelet Analysis of depth 3. We code only the maxima points contained in the corresponding subbands for the three scales 2, 2^2 and 2^3 . This is justified by the fact that these subbands correspond to more than 90% of the image bandwidth, and, therefore, most of the information content of the image. We also use similarities between these three different scales. We assume that edges from the different scales are spatially close enough so that we can afford to encode only one set of locations for the three different resolutions of the Wavelet transform. And we make the same kind of approximation for the values of the argument: we only store the argument values of the medium resolution. We choose the medium resolution because the argument values are smoother than the values of the finest resolution.

4.2 Construction of the curves

In order to obtain high compression ratios, we build curves from the Wavelet Maxima Representation. These curves are unions of edges. This allows us to select more easily the information that needs to be kept, and the curve-based structure allows chain coding.

To construct curves, we link together points of the Wavelet Maxima Representation whose modulus and arguments are close. We utilize this constraint because we want to obtain smooth curves that correspond to continuous edges in the original image. We use the medium resolution to build the curves because it gives the best trade-off between a low high-frequency noise and a sufficient number of points in the Wavelet Maxima Representation, that is to say a sufficient amount of edges.

To link points together, we use an extended neighborhood. We do not link together only points that are 8-neighbors. If there is no candidate in this close neighborhood of eight surrounding pixels, we look further to find the next point in the chain. We allow a difference of 3 for each coordinate. This procedure permits us to fill gaps in the curves. This gap-filling has two consequences: first it reduces the number of curves and therefore leads to a higher compression rate; second it leads to a reconstructed image that does not have intermittent curves.

As we explained in the previous section, we decided to encode only one set of locations for the three different scales of the Wavelet Transform. These curves are not exactly co-located in these three scales; yet they are very close. We therefore proceed by encoding only one set of locations. In addition, we use a curve-shifting procedure that shifts the curves of the coarsest and the finest resolutions.

4.3 Curve-shifting procedure and curve selection

To shift curves, we consider each pixel of each curve in the medium resolution image, and for this pixel, we try to determine a coarser point and a finer point that will correspond to it, that is to say, we try to determine points that belong to the same curve as the pixel we are considering, but in the two other scales. To determine these two points, we look in the 8 pixel-neighborhood of the pixel under consideration in the finest and coarsest images, and attempt to find a modulus maximum. Among the several potential candidates for each resolution, we choose the one that has the highest modulus value. This is depicted in Figure 7.

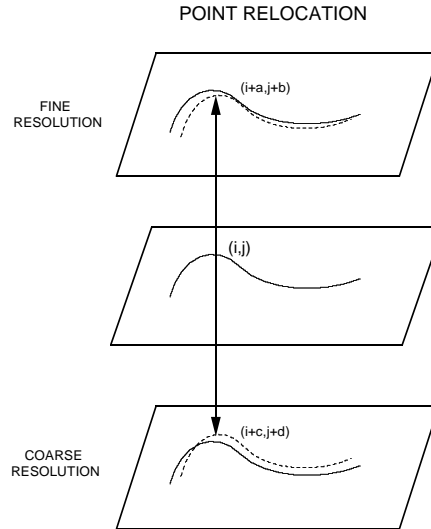


Figure 7: Curve shifting procedure

4.4 Selection on the curves

Once the edges have been constructed, we select only the ones that correspond to the sharpest variations in the image, i.e., to the main structures. Since the main structures usually correspond to rather long curves, we use a thresholds on the length of the curves. In practice we use a length between 15 and 20 pixels on the example of LENA for a 256x256 pixel image. Furthermore, sharp variation curves correspond to curves with a rather high modulus. We impose thresholds on the average modulus along the curve to remove edges that are not sharp enough.

Once this selection process is complete, only the most important curves remain. This is depicted in Figure 8.



Figure 8: Curve-shifting procedure and curve selection

4.5 Curve coding

Once the curve selection process is complete, it remains to encode the information contained in the surviving curves. We use several coding methods, adapted to the different kinds of information that the curves contain. We use lossless coding to encode the locations of the curves. We cannot afford to modify edge positions because the image quality is very sensitive to edge position. We use a differential coding: we record the increment of the position along each curve.

We tried two different methods for this encoding. First an adaptation of Huffman coding to our particular application. Instead of computing an optimal codebook, we only use 8 different codewords that correspond to 8 different directions. These codewords are shown in Table 1.

We first scan the curve once to determine the most likely moves. We then use the estimated empirical probabilities to build a Huffman codebook for this particular curve. We consider the example of a dominant direction equal to 0. This direction corresponds to a move to the right and the directions are numbered in a clockwise manner. The next table shows the codewords used for each move. For the coordinates, the origin is the upper left corner of the image.

The second method we use is Lempel-Ziv coding on the set of the moves for the entire image. Each move corresponds to half a byte. The second method proves more efficient than the first one. Both methods result in a compression gain because along a curve the same move is typically encountered a large number of times.

To encode the values of the modulus and argument along the curves, we use a lossy coding scheme. We use vector quantization with 8-dimensional vectors. As image quality is more sensitive to the argument value than to the modulus value, we use more bits to code the argument than to code the modulus. We employ full search VQ [13], using 0.5 or 1.0 bits per sample for the modulus, and 1.0 or 1.5 bits per sample for the argument. Furthermore, in order to cluster the values that we quantize and obtain

direction	Δx	Δy	codeword
0	1	0	1
1	1	1	011
2	0	1	0011
3	-1	1	0001
4	-1	0	00001
5	-1	-1	00000
6	0	-1	0010
7	1	-1	010

Table 1: Codewords for the 8 directions used in the encoding

better results with the VQ, we quantize the difference between the actual value (of the modulus or the argument) and the corresponding average along the curve. Thus, the values we quantize are clustered around zero. To encode the coarse resolution image, we use the fact that it is inherently low-pass to subsample it by a factor of 2^3 along the rows and the columns of the image, we then perform scalar quantization using only 6 bpp instead of 8. Finally, we have to encode the first point location of each curve and the averages along the curve for the three moduli (corresponding to the three different scales) and the corresponding arguments. This information can be considered as reference information because of the differential coding. Thus we cannot use a lossy coding scheme to encode it: it would lead to a propagation of the error that could completely degenerate image quality. We use Lempel-Ziv coding for these two types of information.

4.6 Results

Figure 9 depicts two examples of image reconstruction using the proposed coding scheme. These are 256x256 pixel LENA images. The left image corresponds to a compression ratio of 12:1 and the right one to a compression ratio of 19:1. They were both obtained using VQ on the modulus and argument values. As expected, most of the texture information of the image is lost, especially at the 19:1 compression rate. A lot of details are lost in the hat and its plume. As expected, the eyes and the mouth are distorted because they mostly consist of rather short curves. On the other hand, the important edges in the image are well – preserved. We still have sharp edges for the hat, the mirror, and Lena’s shoulder.

5 Measures of the perceptual quality of reconstructed images

5.1 Introduction

Usually, to assess the quality of a reconstructed image obtained from a lossy compression scheme, we use mathematical criteria such as MSE or PSNR. But these criteria do not always fit the visually perceived quality of the reconstructed image: one can easily



Figure 9: Reconstructed images at compression ratios 12:1 and 19:1

build two distorted images from an original one in such a way that these mathematical criteria lead to comparable numerical results, yet one looks perceptually much better than the other.

Therefore, it is clear that we need some perceptual criteria that will be able to assess image quality in a way that captures visual perception. This is not an easy task because the Human Visual System (HVS) is very complex, and there are still many aspects of it that we do not fully understand [26].

We will now present two kinds of perceptual distortion measures. The first one, which exhibits low complexity, uses sums of weighted pixel-to-pixel differences. The second one, which is more complex, uses a complete model of the HVS to predict, for each pixel of the image, a probability of error detection by the eye.

5.2 Weighted pixel-to-pixel differences

The idea is to weight the pixel-to-pixel differences in order to give them more or less importance in the summation, depending on the importance that these differences have for visual perception [7]. This model capitalizes on the so-called “masking effect”, i.e., decreased visibility of a signal due to the presence of a suprathreshold background. For each pixel of the image, we compute the activity of the neighborhood of the pixel, and from this value we determine the weight of the pixel-to-pixel difference within the summation. Let us denote respectively by A_i and W_i the activity and the weight of the pixel i . Then the first perceptual distortion measure is given by :

$$EM_p = \left(\frac{1}{m} \sum_{i=1}^m \left| \frac{e_i}{W_i} \right|^p \right)^{1/p}$$

One can use different types of activity functions:

- $A_1 = |x_{i,j} - x_{i-1,j}| + |x_{i,j} - x_{i+1,j}| + |x_{i,j} - x_{i,j-1}| + |x_{i,j} - x_{i,j+1}|$

- $A_2 = \sum_{i=1}^9 |x_i - \bar{x}|^r$
- $A_3 = \max |x_i - x_j|_{i,j=1,2,\dots,9}$

And from this activity value, we compute the weight of the pixel i using an exponential formula, and scale it in order to obtain weights in the range of 1 to 10 approximately.

To improve this perceptual measure, we can also use the fact that the viewer rates the picture by some weighted average of the worst two or three patches. Therefore, we can divide the image into squares, compute a local perceptual measure of the distortion for each square, and finally obtain a global measure for the entire image by using the average of the two or three largest values.

But, because of its lack of complexity as compared to the HVS, this simple type of measure does not prove much more efficient than classical mathematical measures such as the root mean squared error. The root mean squared error is one of the best mathematical criteria because of its non-linearity, which is similar to the eye non-linearity. Nevertheless, it is clear that we need more elaborate HVS models.

5.3 The visible differences predictor

5.3.1 Introduction

The visible differences predictor (VDP) is an algorithm for modeling the human visual response [6]. Its goal is to predict the probability of error detection by the eye, for each pixel of the entire image. This algorithm uses a complete model of the human visual system: it models the amplitude non-linearity function, the contrast sensitivity function, and several detection mechanisms encountered within the visual system.

Within the overall HVS model [6], the multiple detection mechanisms are modeled with four subcomponents: the *spatial frequency hierarchy*, which models the frequency selectivity of the visual system; the *masking function*, which models the magnitude of the masking effect; the *psychometric function*, which describes the threshold in a detailed manner; and the *probability summation*, which combines the responses of all the detection mechanisms into a unified perceptual response. Figures 10 and 11 (reprinted here with permission of MIT Press from [6]) illustrate the overall architecture of HVS and VDP.

5.3.2 Amplitude non-linearity function

The amplitude non-linearity function describes the sensitivity of the eye as a function of light intensity. It is well-known that visual sensitivity and perception are nonlinear functions of luminance [22] [26] [6]. The local amplitude non-linearity is implemented [6] in the VDP as a function of pixel location (i, j) :

$$\frac{R(i, j)}{R_{max}} = \frac{L(i, j)}{(L(i, j) + c_1 L(i, j))^b}$$

where $\frac{R(i, j)}{R_{max}}$ is the normalized retina response, L is the luminance falling on the retina, b is 0.63 and c_1 is 12.6 for the units of cd/m^2 . The amplitude non-linearly function is shown in Figure 12 (reprinted here with permission of MIT Press from [6]).

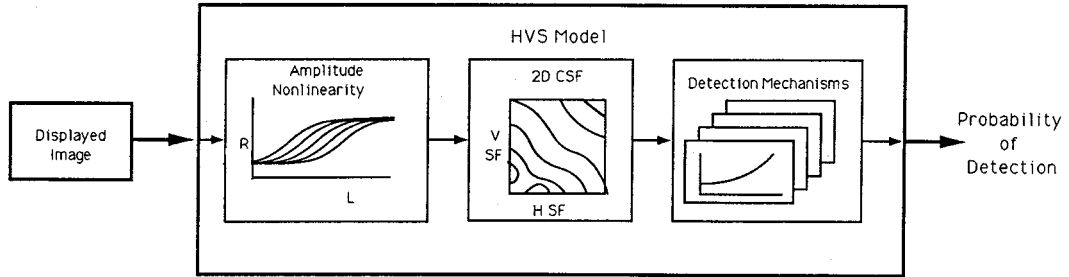


Figure 10: The three main components of the HVS model within the VDP (reprinted with permission of MIT Press from [6])

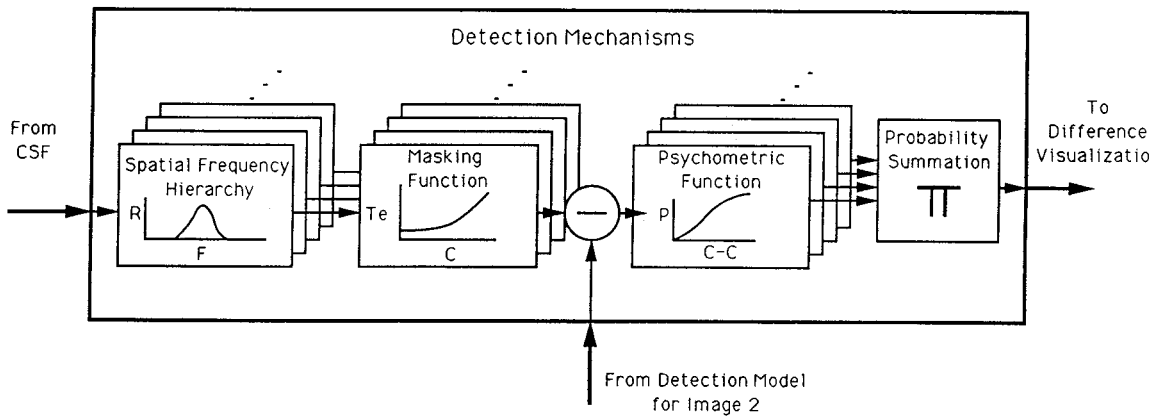


Figure 11: The four components of the detection mechanism in VDP (reprinted with permission of MIT Press from [6])

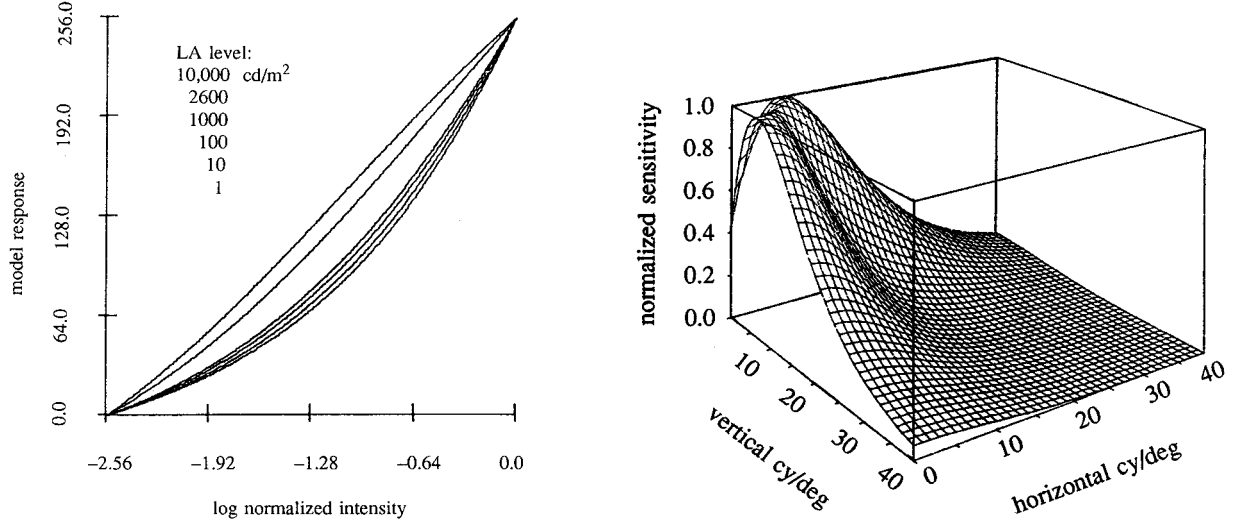


Figure 12: Amplitude non-linearity and contrast sensitivity functions (reprinted with permission of MIT Press from [6])

5.3.3 Contrast Sensitivity function

The contrast sensitivity function (CSF) describes the variations in visual sensitivity as a function of spatial frequency. The CSF is a function of light adaptation, noise, color, accommodation, eccentricity and image size; it is illustrated in Figure 12 (reprinted here with permission of MIT Press from [6]).

The VDP models [6] the sensitivity S as a function of radial spatial frequency ρ in c/deg, orientation θ in degrees, light adaptation level l in cd/m^2 , image size i^2 in visual degrees, lens accommodation due to distance d in meters, and eccentricity e in degrees:

$$S(\rho, \theta, l, i^2, d, e) = P \min[S_1\left(\frac{\rho}{r_a r_e r_\theta}, l, i^2\right), S_1(\rho, l, i^2)]$$

where P is the absolute peak sensitivity of the CSF. In our work, we used $P=250$. The parameters r_a , r_e and r_θ model the changes in resolution due to the accommodation level, eccentricity and orientation, respectively, via the following equations:

- $r_a = 0.856 d^{0.14}$
- $r_e = \frac{1}{1+0.24e}$
- $r_\theta = 0.11 \cos(4\theta) + 0.89$

Remaining to be modeled are the effects of the image size and the light adaptation level:

$$S_1(\rho, l, i^2) = ((3.23 (\rho^2 i^2)^{-0.3})^5 + 1)^{-1/5} A_l 0.9 \rho e^{-(0.9 B_l \rho)} \sqrt{1 + 0.06 e^{0.9 B_l \rho}}$$

with $A_l = 0.801 \left(\frac{1.7}{l}\right)^{-0.2}$ and $B_l = 0.3 \left(\frac{101}{l}\right)^{0.15}$.

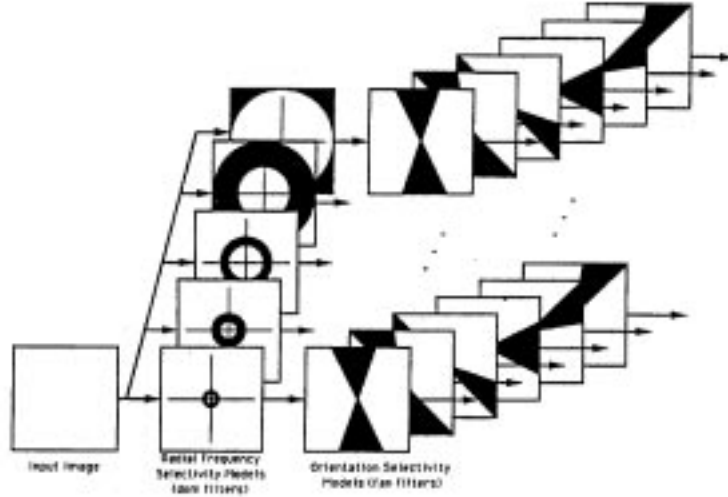


Figure 13: The formation of the cortex filters (reprinted with permission of MIT Press from [6])

5.3.4 Spatial frequency hierarchy

The frequency selectivity of the visual system is modeled [6] by using a hierarchy of filters, called *cortex filters*. Cortex filters are constructed from separate classes of filters, whose effects are cascaded to describe the combined radial and directional selectivity of cortical neurons. The first class of filters is called *dom filters*. The second class is called *fan filters*. These filters are shown in Figure 13 (reprinted here with permission of MIT Press from [6]).

The dom filters [6] are formed as differences of a series of two-dimensional low-pass *mesa* filters, characterized by a flat pass-band, a transition region, and a flat stop-band region. In the VDP algorithm, the transition region is modeled by a Hanning window, so that each mesa filter can be completely described by its half-amplitude frequency $\rho_{1/2}$ and the transition width $t\omega$ as follows:

$$\left\{ \begin{array}{ll} mesa(\rho) = 1.0 & \text{for } \rho < \rho_{1/2} - \frac{t\omega}{2} \\ = \frac{1}{2} \left(1 + \cos\left(\frac{\pi(\rho - \rho_{1/2} + t\omega/2)}{t\omega}\right) \right) & \text{for } \rho_{1/2} - \frac{t\omega}{2} < \rho < \rho_{1/2} + \frac{t\omega}{2} \\ = 0.0 & \text{for } \rho > \rho_{1/2} + \frac{t\omega}{2} \end{array} \right.$$

The radial frequency selectivity is modeled by dom filters (differences of mesa) formed from two mesa filters evaluated with different half-amplitude frequencies. The k th dom filter is given by

$$dom_k(\rho) = mesa(\rho)|_{\rho_{1/2}=2^{-(k-1)}} - mesa(\rho)|_{\rho_{1/2}=2^{-k}}$$

Increased values of k correspond to lower frequency bands in the hierarchical pyramid of filters of the cortex transform.

The lowest-frequency band is referred to as the baseband. To avoid ringing effects that occur if we use the same formulation for this base filter, we use a truncated

Gaussian function for the baseband, as given by:

$$\begin{cases} base(\rho) &= e^{-(\rho^2/2\sigma^2)} & \text{for } \rho \geq \rho_{1/2} + \frac{t\omega}{2} \\ &= 0 & \text{for } \rho < \rho_{1/2} + \frac{t\omega}{2} \end{cases}$$

with $\sigma = \frac{1}{3}(\rho_{1/2} + \frac{t\omega}{2})$ and $\rho_{1/2} = 2^{-K}$.

The equations for the whole set of dom filters are therefore:

$$\begin{cases} dom_k(\rho) &= mesa(\rho)|_{\rho_{1/2}=2^{-(k-1)}} - mesa(\rho)|_{\rho_{1/2}=2^{-k}} & \text{for } k = 1, \dots, K-2 \\ &= mesa(\rho)|_{\rho_{1/2}=2^{-(k-1)}} - base(\rho)|_{\rho_{1/2}=2^{-k}} & \text{for } k = K-1 \end{cases}$$

where K is the total number of radial filters.

In the practical implementation of the algorithm, we used a transition width of the form $t\omega = \frac{2}{3}\rho_{1/2}$. This transition width configuration gives constant behavior on a log frequency axis with a bandwidth of 1.0 octave and symmetrical responses.

The orientational selectivity is modeled with *fan* filters [6]. An integer number of fan filters is used to approximate the nearly continuous orientation selectivity of the visual system. A Hanning window is also used for these filters, which is determined in angular degrees θ in the Fourier plane. The equation for fan l as a function of orientation is

$$\begin{cases} fan_l(\theta) &= \frac{1}{2}(1 + \cos[\frac{\pi|\theta - \theta_c(l)|}{\theta_{tw}}]) & \text{for, } |\theta - \theta_c(l)| \leq \theta_{tw} \\ &= 0.0 & \text{for, } |\theta - \theta_c(l)| > \theta_{tw} \end{cases}$$

where θ_{tw} is the angular transition width and $\theta_c(l)$ is the orientation of the center, or peak, of the fan filter given by $\theta_c(l) = (l-1)\theta_{tw} - 90$.

If we set the transition width equal to the angular spacing $\theta_{\Delta c}$ between adjacent fan filters,

$$\theta_{tw} = \theta_{\Delta c} = \frac{180}{L}$$

where L is the total number of fan filters. The total number of fan filters in our implementation is 6, which gives an orientation bandwidth of 30 degrees.

The cortex filters are formed as the polar separable product of the dom and fan filters as

$$\begin{cases} cortex_{k,l}(\rho, \theta) &= dom_k(\rho) \cdot fan_l(\theta) & \text{for, } k = 1, \dots, K-1; l = 1, \dots, L \\ &= base(\rho) & \text{for, } k = K \end{cases}$$

where the particular cortex filter can be denoted by the dom and fan filter indices, k , l , respectively. The total number of cortex filters in the set is $((K-1)L + 1)$, which is 31 in our implementation ($K = 6, L = 6$).

5.3.5 Masking function

Masking refers to the decreased visibility of a signal due to the presence of a suprathreshold background. The masking function [6] quantifies this effect as a function of the background contrast.

In the algorithm, many frequencies contribute to the normalized mask contrast m_n within a particular band k, l as a function of location $[i, j]$. It is calculated as

$$m_n[i, j] = \mathcal{F}^{-1}(\mathcal{L}[u, v] \cdot \text{csf}[u, v] \cdot \text{cortex}^{k,l}[u, v])$$

where \mathcal{L} is the Fourier transform of the input image processed by the amplitude non-linearity, and u and v are the Cartesian frequency components.

The threshold elevation is implemented as a function of location as follows

$$T_e^{k,l}[i, j] = (1 + (k_1(k_2|m_n^{k,l}[i, j]|)^s)^b)^{1/b}$$

where s corresponds to the slope of the high contrast masking asymptote. Its values range from 1.0 for the baseband to 0.65 for the middle frequencies. For the other parameters we use $k_1 = 1.53 \cdot 10^{-2}$, $k_2 = 392.5$ and $b = 4$.

The band-specific threshold elevation $T_e^{k,l}[i, j]$ is a function of pixel location and is referred to as the *threshold elevation image*. Since we seek to predict the difference between an original image and a distorted one, we have to take into account the fact that the only masking that occurs is that which is mutual to both images. Therefore the masking threshold elevation image is given by

$$T_{em}^{k,l}[i, j] = \min(T_{e1}^{k,l}[i, j], T_{e2}^{k,l}[i, j])$$

where the subscript 1 and 2 refer to the two images input to the algorithm.

5.3.6 Psychometric function and probability summation

The psychometric function [6] describes the increase in the probability of detection as the signal contrast increases. It is given by the following equation

$$P(c) = 1 - e^{-(c/\alpha)^\beta}$$

where $P(c)$ is the probability of detecting a signal of contrast c .

The probability of detection as a function of location is calculated as follows. First the contrast difference as a function of location is given by

$$\begin{aligned} \Delta C_{k,l}[i, j] &= C1_{k,l}[i, j] - C2_{k,l}[i, j] \\ &= \frac{B1_{k,l}[i, j]}{B_K} - \frac{B2_{k,l}[i, j]}{B_K} \end{aligned}$$

where $\Delta C_{k,l}[i, j]$ is the contrast difference for band k, l as a function of pixel location, $B1_{k,l}$ and $B2_{k,l}$ are the two filtered input images for band k, l , and B_K is the baseband mean. Then the probability of detection in band k, l is computed by

$$P_n[i, j] = 1 - e^{-(\Delta C_{k,l}[i, j]/T_{em}[i, j])^\beta}$$

where $\beta = 1.04$ in our implementation.

Once the detection probabilities are computed for each band of the spatial frequency filter hierarchy, these probability images are combined into a single image that describes the overall probability of detecting an error for every pixel in the image:

$$P_t[i, j] = 1 - \prod_{k=1, K; l=1, L} (1 - P_{k,l}[i, j])$$

where $P_t[i, j]$ is the total probability of detection resulting from all bands as a function of location.

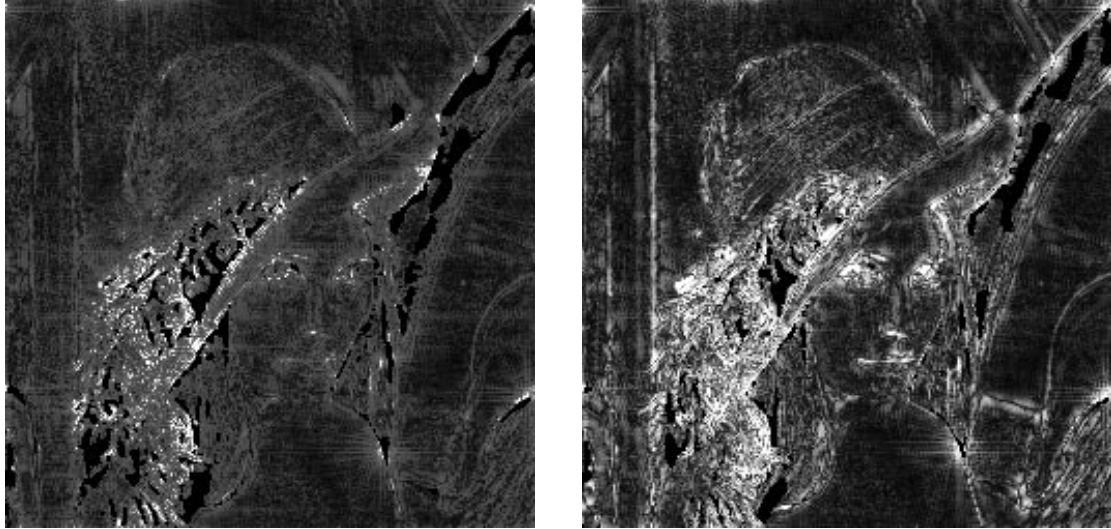


Figure 14: Two probability mappings obtained for 6:1 and 18:1 compression ratios

5.4 Perceptual distortion of reconstructed images

Once the mapping of probabilities of error detection by the eye have been computed for the reconstructed image, it can be useful, in order to compare more easily our coding scheme with some other reference scheme, to convert this mapping into a single number. To do so, we may use the following formula:

$$Dist = \frac{100}{N} \sum_{i,j} p(i,j)^2$$

where $Dist$ will be the perceptual distortion for the entire image, N is the total number of pixel in the image, and $p(i,j)$ is the probability of error detection for the pixel (i,j) .

If we want to obtain meaningful numerical values, we have to normalize this measure. Therefore we subtract the numerical value $dist_0$ obtained for no coding at all, which corresponds to the case of a perceptually perfect image quality for the reconstructed image. We then obtain the following formula:

$$Dist_{norm} = \frac{100}{N} \sum_{i,j} p(i,j)^2 - dist_0$$

We used a factor of 100 to obtain values of a convenient order of magnitude.

5.5 Results

We used this perceptual distortion measure to assess the quality of the reconstructed images obtained from our coding scheme. Typical results of the evaluation are shown in Figure 14. We also performed this quality assessment on the JPEG coding algorithm in order to compare our algorithm versus an established reference algorithm.

The two plots in Figure 15 show the results obtained with and without vector quantization. Image quality at 16:1 with vector quantization is not very different from image quality obtained without vector quantization at 9:1.

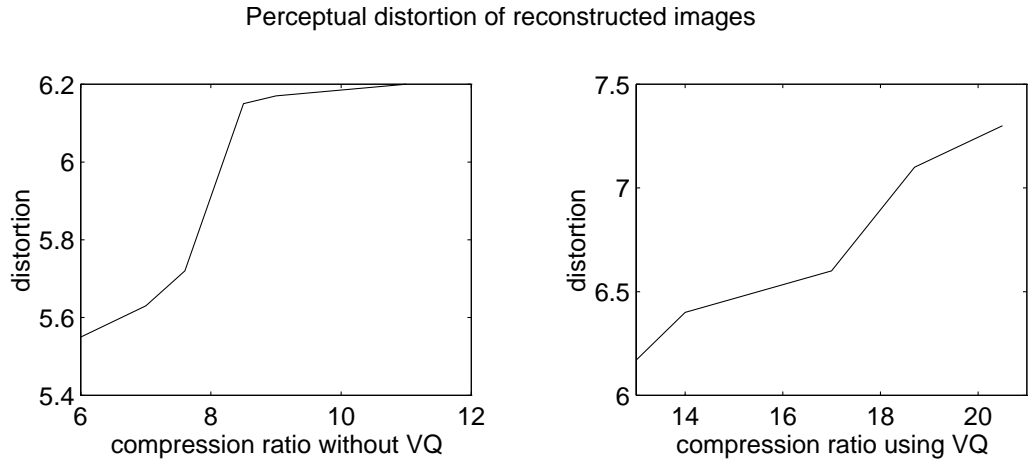


Figure 15: Perceptual distortion of reconstructed images from our coding scheme

However, the comparison with JPEG clearly shows (see Figure 16) that for any compression rate between 5:1 and 20:1, the JPEG algorithm leads to a higher image quality than the proposed multiscale edge-based coding. Actually, the difference in performance is significant. Therefore the method we implemented cannot be considered as a practical alternative to JPEG.

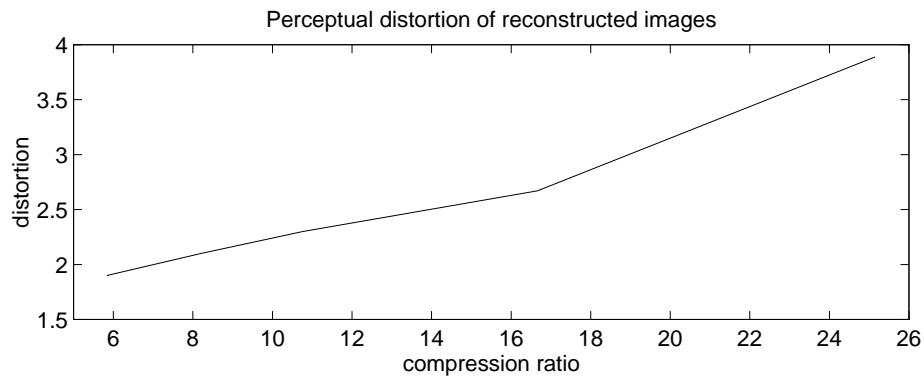


Figure 16: Perceptual distortion of reconstructed images from JPEG

6 Conclusions

As we explained in the previous chapter, the proposed image compression scheme based on multiscale edges does not prove as efficient as other generic methods such as JPEG. The perceptual image quality is noticeably inferior to the one obtained using JPEG, for a wide range of coding gains.

One can conclude from these results that the Wavelet Maxima Representation is not complete enough to be used within such a compression scheme. In fact, edge information is not robust enough to be encoded itself. We must admit that the Wavelet Maxima Representation provides an interesting edge map of the image as a side benefit, and multiscale edges can convey a lot of information about image content. For example, this representation can be used to denoise the image because, from this representation, one can compute the Lipschitz exponents and by putting a threshold on the exponents, remove white noise from a noisy image.

Therefore, one can suggest to use this representation either for purposes other than image compression, or within another coding scheme. This representation can be very useful in pattern recognition algorithms or classification because it extracts important image structure, and classification is often performed using this type of information. Furthermore, one could use this representation within another coding scheme, in order to rank information according to perceptual measures. From the edge-based representation, one can assess the activity of a region of an image and, from this, assign a certain number of bits to this region. It could also be used as part of a more complete coding scheme, coupled with a texture coding, for improved rate-distortion trade-off.

References

- [1] Stephane MALLAT, *A Theory for Multiresolution Signal Decomposition: The Wavelet Representation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 11 No 7, July 1989.
- [2] M.M. LIVSTONE *Wavelets: A Conceptual Overview*, May 1994.
- [3] Stephane MALLAT, *Characterization of Signals from Multiscale Edges*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 14 No 7, July 1992.
- [4] Stephane MALLAT and S. ZHONG, *Compact Image Coding from Edges with Wavelets*, IEEE proceedings ICASSP, 1991, p 2393.
- [5] Stephane MALLAT and S. ZHONG, *Zero-crossings of a Wavelet Transform*, IEEE Transactions on Information Theory, vol 37, No 4, July 1991.
- [6] S. DALY, *The Visible Differences Predictor: an Algorithm for the Assessment of Image Fidelity*, Digital Images and Human Vision, A.P. Watson editor, p 179, MIT Press 1993.
- [7] J. LIMB, *Distortion Criteria of the Human Viewer*, IEEE Transactions on Systems, Man and Cybernetics, vol 9, No 12, December 1979.
- [8] Z. BERMAN, *Generalizations and Properties of the Multiscale Maxima and Zero-Crossings Representations*, ISR Ph.D. Thesis Report #92-9.
- [9] Z. BERMAN and J.S. BARAS, *More about Wavelet Maxima Representations*, Proceedings of the 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. V, pp. 662-665, Minneapolis, Minnesota, April 27-30, 1993.
- [10] Z. BERMAN, J.S. BARAS and C. BERENSTEIN, *The Parametric Wavelet Maxima Representation*, Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis, pp. 59-62, Victoria, British Columbia, Canada, October 4-6, 1992.
- [11] Z. BERMAN and J.S. BARAS, *Properties of the Multiscale Maxima and Zero-Crossings Representations*, IEEE Trans. on Signal Processing, Vol. 41, No. 12, Dec. 1993, pp. 3216-3231.
- [12] M. ANTONINI, M. BARLAUD, P. MATHIEU, I. DAUBECHIES, *Image coding using Wavelet Transform*, IEEE Transactions on Image Processing, Vol 1, No 2, p 205, April 1992.
- [13] Y. LINDE, A. BUZO, R.M. GRAY, *An algorithm for Vector Quantizer Design*, IEEE Proceedings on Communications, Vol 28, No 1, p 84, Jan 1980.
- [14] M. ANTONINI, M. BARLAUD, P. MATHIEU, I. DAUBECHIES, *Image Coding Using Vector Quantization in the Wavelet Transform Domain*, IEEE Proceedings ICASSP, p 2297, April 1990.
- [15] LEWIS, KNOWLES, *Image Compression using 2-D Wavelet Transform*, IEEE Transactions on Image Processing, Vol 1, April 1992, p 244.
- [16] P.C. COSMAN, K.L. OELHER, E.A. RISKIN, R.M. GRAY, *Using Vector Quantization for Image Processing*, Proceedings of the IEEE, No 9, Vol 81, Sept 1993.

- [17] S. OHTSUKA and M. INOUE, *Quality Evaluation of Pictures with Multiple Impairments Based on Visually Weighted Errors*, Proceedings of the SID, vol 30, No 1, 1989.
- [18] M. VETTERLI and C. HERLEY, *Wavelets and Filter Banks: Relationships and New Results*, IEEE Proceedings ICASSP, Albuquerque, April 1990.
- [19] J. WOODS, *Subband Coding of Images*, IEEE ICASSP Vol 34, No 5, October 1986.
- [20] E. RISKIN, E. DALY, R. GRAY, *Pruned Tree-structured Vector Quantization in Image Coding*, IEEE Proceedings ICASSP, Glasgow 1989, p 1735.
- [21] J.C. DUTHOU, *Tree-structured Face Recognition Algorithm using a Multiresolution Scheme*, Rapport de Stage , Ecole Nationale Supérieure des Telecommunications, 1994.
- [22] A. SINGH and V. BOVE, *Image Quality Measure Based on a Human Visual System Model*, IEEE Journal on Selected Areas in Communications, Vol 11, No 1, Jan 1993.
- [23] J. KIM, H. LEE and J. CHOI, *Subband Coding Using Human Visual Characteristics for Image Signals*, IEEE Journal on Selected Areas in Communications, Vol 11, No 1, Jan 93.
- [24] G. KARLSSON and M. VETTERLI, *Three Dimensional Subband Coding of Video*, Proceedings of IEEE ICASSP, 1988, pp. 1100-1103.
- [25] M. PERKINS and T. LOOKABAUGH *A Psychophysically Justified Bit Allocation Algorithm for Subband Image Coding Systems*, Proceedings of IEEE ICASSP 1989.
- [26] N. JAYANT, J. JOHNSTON and R. SAFRANEK, *Signal Compression Based on Models of Human Perception*, IEEE Proceedings, Vol 81, No 10, Oct 1993.
- [27] V. CIZEK, *Discrete Fourier Transforms and their Applications*, Adam Hilger, 1986.