

Multiscale Coding of Images

by

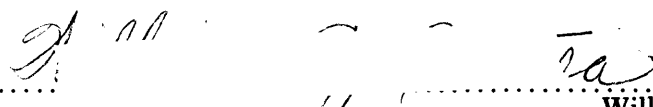
William J. Butera

S.B., Electrical Engineering
Massachusetts Institute of Technology
Cambridge, Massachusetts
1982

SUBMITTED TO THE MEDIA ARTS AND SCIENCES SECTION
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS OF THE DEGREE OF
MASTER OF SCIENCE
AT THE MASSACHUSETTS INSTITUTE OF TECHNOLOGY
September 1988

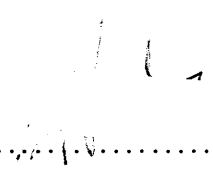
©Massachusetts Institute of Technology 1988
All Rights Reserved

Signature of the Author



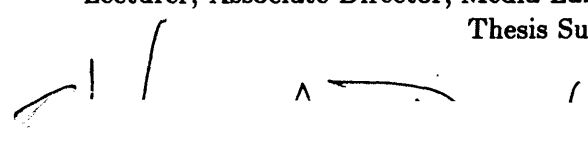
William J Butera
Media Arts and Sciences Section
July 19, 1988

Certified by



Andy Lippman
Lecturer, Associate Director, Media Laboratory
Thesis Supervisor

Accepted by



Stephen A. Benton
Chairman
Departmental Committee on Graduate Students

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

NOV 03 1988

LIBRARIES
DATE

Multiscale Coding of Images

by

William J. Butera

Submitted to the Media Arts and Sciences Section on July 19, 1988 in partial fulfillment of the requirements of the degree of Master of Science at the Massachusetts Institute of Technology

Abstract

A video coding system is described which uses adaptive quantization of spatio-temporal subbands to enable transmission of near NTSC resolution color image sequences over a 1.5 megabit channel. The sequence is first decomposed into spatio-temporal spectral components using quadrature mirror filters. Simple heuristics are used to determine the local importance of a given component's energy. These components are then coded in one of three ways; either set to zero (coarsest quantization), vector quantized or scalar quantized. The representation of an image sequence as set of spatio-temporal coefficients is shown to offer great efficiencies to the task of vector quantization, specifically in minimizing the size of the codebook.

Thesis Supervisor: Andy Lippman

Title: Lecturer, Associate Director, Media Laboratory

This work was supported in part by Columbia Pictures Entertainment Incorporated, Paramount Pictures Corporation and Warner Brothers Incorporated.

Contents

1	Introduction	5
2	Visual & Coding Principles	9
2.1	Nonuniform Response of the HVS	10
2.1.1	Results from Physiology and Psychophysics	11
2.1.2	Results from Image Coding	14
2.2	Subband Coding	15
2.3	Vector Quantization	18
3	Preprocessing for Perceptual Sensitivity	21
3.1	Color	22
3.1.1	Spatial Chromatic Resolution	22
3.1.2	Temporal Chromatic Resolution	23

3.2	Partitioning of the Luminance Spectrum	24
3.3	Bandsplitting Techniques	31
3.4	Visual Sensitivity to Energy Distribution in the Subbands	40
4	Vector Quantization of Subbands	55
4.1	Orientation Independent Codebooks	57
4.2	Cascading of Codebooks	61
4.3	Multiscale Codebooks	67
4.4	Temporal VQ of the Subbands	70
5	Coding Example	71
6	Conclusions	77
6.1	Suggestions for Future Work	77
6.2	Conclusions	79
A	Test Images	80
B	Acknowledgments	87

Chapter 1

Introduction

This thesis describes a coding scheme which combines spatio-temporal band splitting with vector quantization (VQ) to enable transmission of color image sequences over low bandwidth channels. Separable 3D FIR filters are used to divide the image sequence into component spectral bands. The vector coder is used to reduce the redundancy in each subband in a manner which is well matched to the performance of the human visual system (HVS). The subband representation is shown to enable

a significant improvement in the performance of the vector coder.

Most common image compression techniques compress images by either exploiting inherent redundancies in the 1D, 2D or 3D signals or by confining the quantization distortion into 'areas' (either spatial or spectral) of reduced visual sensitivity. Coding schemes have been proposed which combine one or more coding techniques from the two categories into a hybrid coder [Pratt 1978] [Schreiber 1986].

The technique explored in this thesis also segments the coding task into two; grouping the energy of the image into psychovisually distinct categories, or bands, and then separately controlling the relative amount of quantization distortion introduced into each of these bands by VQ. This process is illustrated in Figure 1.1 . The intensities in local regions of the color image sequence are transformed into spatially localized bands of color and oriented luminance energy. The amount of relative distortion introduced into these bands is determined by varying the VQ parameters such as the block dimensions or the size of the codebooks. No claims are made concerning to the optimality of VQ as a tool for distributing or shaping the quantization distortion. However the representation of the luminance signal as oriented spectral bands is shown to offer great efficiencies in reducing the size of the VQ codebook.

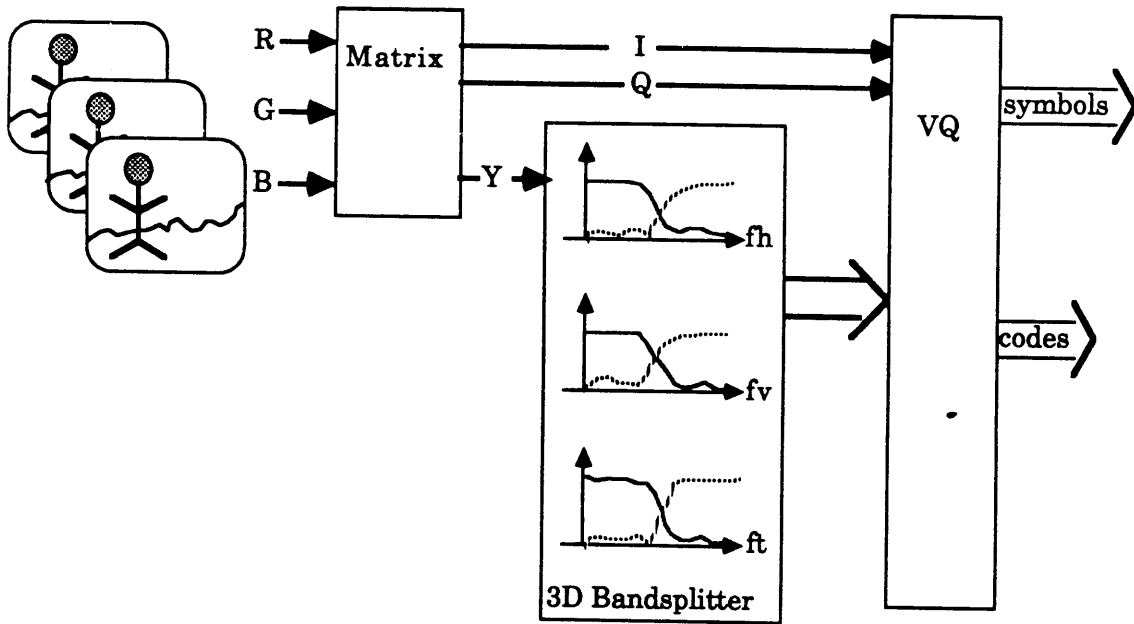


Figure 1.1: *Block diagram of coding process*

The remainder of this paper is organized in five chapters. Chapter two reviews some of the previous work in the fields of human visual acuity, pyramid representation of images, coding of spectral subbands, and vector quantization. Chapter three discusses the decomposition of the sequences into various bands according to color and oriented spectral content. The treatment of color is quickly reviewed. Subband coding techniques and the pyramid representation are both explained. Previous results from related work in image coding are interpreted in terms of the subband representation in order to infer the the relative perceptibility of quantization distortion introduced into the various subimages. Finally, tests are carried out to support

the implied sensitivity hierarchy.

Chapter four describes the application of vector quantization to the task of efficiently coding selected spatio-temporal components. Four techniques are proposed for limiting the size of the VQ codebook by exploiting the statistical similarities between the bands. Chapter five presents coding examples and sample bandwidth calculations. Chapter six contains suggestions for future work and presents the conclusions.

Chapter 2

Visual & Coding Principles

This chapter reviews previous work in image compression techniques. The work selected for review is chosen because it relates to either reduction of signal redundancy or to the nonuniform sensitivity of the HVS. The first section reviews efforts to model the response of the HVS to different constituents of the visual world such as color, detail, contrast and motion. Work from the fields of neurophysiology, psychophysics, and image coding is examined in order to outline our current under-

standing of how the HVS segments and processes the visual input. The next section reviews the use of subband transforms in the coding of images. The recursive application of the subband transformations to form image pyramids is also discussed. The last section outlines the use of block coding or vector quantization (VQ) to exploit signal redundancy.

2.1 Nonuniform Response of the HVS

Any attempt to match a coding scheme to the sensitivity of the human observer must take into account the varying response of the HVS. It has long been understood that the visual world is made up of components for which humans have varying degrees of acuity. Examples of this include the blurring of detail on the surface of moving objects and the waning of the color percept in very dark environments. In the fields of image coding, psychophysics and physiology, a large body of work has addressed the division of the visual input into its component parts and the measurement of human acuity to energy in these components. Psychophysics and physiology examine human acuity in order to gain insight into the makeup and the functioning of the human visual system. Work in image coding is motivated by the

desire to understand which information in the visual input is necessary in order to code, transmit and reconstruct the sequence.

2.1.1 Results from Physiology and Psychophysics

Work in neurophysiology has demonstrated the segmentation of the visual stimuli at a number of levels in the early visual pathway¹. It is widely acknowledged that the ganglion cells which populate the retina fall into at least two categories, referred to by physiologists as X and Y type cells². The X type cells have smaller receptive fields, higher spatial selectivity, are color selective but have a slower time constant. The Y type cells exhibit less spatial and color selectivity, but have a more transient response to temporal variation. Both of these cell types have, on average, isotropic, center-surround receptive fields, the bandpass response of which has been closely modeled as the difference of two Gaussian distributions [Hildreth 1985].

One of the next processing centers in the early visual pathway is the striate cortex. Studies that have measured the response of cortical cells to movement of oriented line pairs or sinusoidal patterns, have indicated that the cortical cells are

¹for an excellent overview of early visual processing, see Marr's *Vision* [Marr 1982]

²most texts refer to these as cones and rods, respectively

locally selective for orientation, spatial frequency, and direction of motion [Hubel 1986] [Schiller 1976] [Maunsell 1983].

Work in psychophysics has focused both on measuring the performance of the HVS and on constructing models to explain its behavior. In a recent set of experiments, Kelly [Kelly 1983] measured the HVS's sensitivity thresholds to moving sinusoidal gratings. By measuring the contrast threshold for gratings of different spatio-temporal frequencies, he plotted the HVS's contrast-threshold response for both chromatic and achromatic stimuli. Comparison of the two contrast-threshold surfaces illustrates that for temporal rates above 5 Hz., the sensitivity to chromatic flicker is significantly below that of luminance flicker [Kelly 1979] [Kelly 1983]

While most experiments indicate that the HVS's response is highly non-linear, a number of models have been advanced that are based on linear processes and that successfully predict the response of the HVS to simple stimuli. Wilson [Wilson 1979] proposed a model which splits early visual processing into four channels. Each of these channels is tuned to a particular spatial frequency. The two wider bandwidth channels exhibit transient temporal response while the channels with the smaller bandwidths have more sustained responses. This model correctly predicts the threshold contrast levels for simple cosine gratings. In the case of motion

perception, data have been presented that indicates the mechanisms responsible for the coherence of simple moving patterns also pass their input through a series of orientation tuned spatial filters [Movshon 1986]. Models which use localized spatio-temporally oriented filters for the calculation of optical flow have been shown to be consonant with the HVS's performance on simple inputs [Heeger 1987].

The above work was selected for review in order to make the following points :

- The HVS records chromatic stimuli from receptors with an inherently longer time constant and therefore at reduced temporal acuity relative to luminance.
- The retinal signals on which the HVS builds its visual representations appear to be localized in both space and spatial frequency.
- Early visual processing appears to take place in a number of separate channels which differ in their selectivity to spatial and temporal frequency components.
- The information in these channels is further processed, in part, by simple cortical cells which also appear to be orientation selective.

Chapter three draws on these points as a guide for matching the specifics of the subband coder to the performance of the HVS.

2.1.2 Results from Image Coding

Work in image coding has focused on using the considerable power of linear systems theory³ to exploit the nonuniform sensitivity of the HVS. The theme for much of this work is the optimal decomposition of images under the restriction that the decomposition be a linear process. Examples of such work include the treatment of color in NTSC, multi-channel coding schemes for still images, variable rate multi-channel coding techniques for HDTV transmission, temporally modulating luminance enhancement signals onto subsampled chrominance carriers, and subband coding.

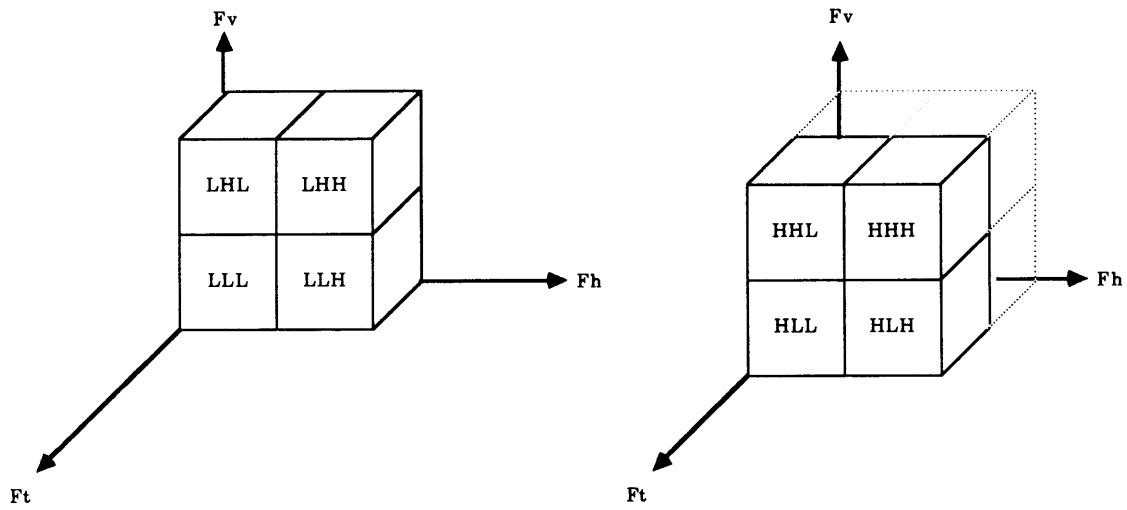
In the case of color in the NTSC system, the significantly reduced human acuity to spatial variations in chromaticity (compared to that of luminance) was exploited to significantly reduce the bandwidth of the color difference signals. In another popular coding scheme, the spatial highs and lows are separated and transmitted in two distinct channels [Troxel 1980]. The reduced sensitivity to quantization distortion in the isotropic spatial highs signal is then used to reduce the bit rate in the highs channel. In a variable rate two channel system proposed by Glenn [Glenn 1983] [Glenn 1984], the high luminance detail component of image sequences is separately sampled at a reduced temporal rate. This scheme is based on the the

³basic references for linear systems theory include [Oppenheim 1975] and [Dudgeon 1984]

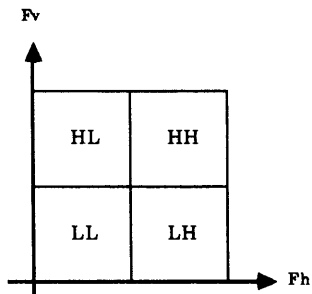
HVS's varying temporal response for the motion of the different spatial components. Finally, Schreiber and Lippman [Schreiber 1988] proposed the decomposition of a high resolution sequence into spectral blocks uniformly spaced along the spatio-temporal frequency axes. Visually important blocks from this ensemble are then selected for transmission on a per scene basis.

2.2 Subband Coding

Subband coders operate by breaking up a signal into a set of component spectral blocks which can then be used for either analysis or coding. In image processing, subband coders filter along the space-time axes to produce a set of spatio-temporal blocks. These component blocks may then be subsampled by a factor consistent with the passband of the filter. Figure 2.1 illustrates the spectral segmentation of both 2D and 3D subband coders which partition the spectrum using separable filters. There are a variety of techniques available to perform the bandsplitting. These include convolution with sharp cutoff FIR filters which minimize aliasing but introduce ringing, filtering with matched quadrature mirror filter (QMF) pairs, and orthogonal transformations such as blocked DCT's with small regions of support.



(a) 3 Dimensional Subbands



(b) 2 Dimensional Subbands

Figure 2.1: *Spectral Block of Subband Coders*

Subband transforms enjoy a number of advantages. They can isolate localized, oriented image features at a particular scale in a manner similar to that previously described for early visual processing in the HVS. Techniques have been developed which allow complete representation of the image and nearly exact reconstruction [Adelson 1987] [Simoncelli 1988] Subband transforms can be recursively applied to form image pyramids. The pyramid representation provide a uniform representation of information appearing at different scales [Burt 1984]. The spectral components can be coded individually which allows the quantization distortion in a given component to be optimally matched to the sensitivity of the corresponding channel in the HVS.

Subband transforms were initially applied to speech coding where the noise due to quantization was 'shaped' to match the response of the human auditory system. The use of subband coders has recently been extended to the compression and analysis of images. Spatially compact orthogonal transforms for pyramid coding of images were developed and shown to give good results when applied in two dimensions [Adelson 1987]. Flexible techniques were developed for the design of quadrature mirror filter sets which allow trade offs between filter complexity and image quality [Simoncelli 1988] Investigators have looked at different methods of coding the subbands including DPCM [Woods 1986], adaptive bit allocation [West-

erink 1988], companded PCM [Gharavi 1987], and rate adaptive quantization [Chen 1984].

2.3 Vector Quantization

Vector quantization (VQ) is a redundancy reduction technique which breaks a signal into a group of blocks and assigns to all similar blocks a single compact channel symbol. The individual values in the blocks are used to form vectors. The code-words assigned to the channel symbols are calculated by partitioning a vector space populated by vectors from a training set. Identical codebooks are present at both the decoder and the encoder so that only the channel symbols need be transmitted.

VQ, which was initially used to compress speech signals, is a relatively new entry into the field of image compression. Application of VQ to 2D images and sequences first became practical as algorithms for optimally subdividing a populated vector space were developed [Friedman 1977] [Equitz 1987]. In an early application, block coding has been utilized to adaptively code the color space of 24-bit still images with 8-bit symbols [Heckbert 1982]. VQ encoding of color images has also been

used to encode stereo pair images for fast display on framestores with color maps [Chesnais 1988].

Most of the recent work on VQ for image coding has focused on the coding of monochrome still images. This work has served to illustrate two fundamental problems with the use of VQ for encoding images or entire sequences. These problems are the distribution of the resulting quantization error and the size of the codebook necessary to represent a sequence. Early work relied on the mean square error (MSE) as a distortion measure [Gray 1984]. Acknowledging that this measure is not strongly correlated to subjective image quality, Ramamurthi and Gersho [Ramamurthi 1986] proposed a classification scheme which specifically favored oriented edge detail. While this technique renders most edges with more fidelity, it is limited to favoring one orientation per local neighborhood and is ill equipped to adapt to concurrent activity in multiple image components.

The second limitation to the use of VQ to encode image sequences is the size of the codebook. The approach taken in many recent papers has been to design a single codebook which would be applicable to all natural images. However, because such a codebook must be designed using a finite training set, there are no assurances that the statistics of all images can be sufficiently represented. Further, such

global codebooks allow little flexibility in the distributing the quantization distortion. Other recent papers have acknowledged the advantages of codebooks which are tuned specifically for individual images or specific temporal neighborhoods. [Aravind 1987] [Goldberg 1986] [Heckbert 1982]. In this case, in order for straight VQ to be useful for low bit rate coding of images, the codebooks must either be compressed and pre-transmitted or incrementally updated. Both approaches require the application of some distortion measure, preferably one that has some subjective meaning.

Chapter 3

Preprocessing for Perceptual Sensitivity

In order to match the performance of an image coding algorithm to the sensitivity of the human observer, it is necessary to account for the variable acuity of the HVS. This chapter describes how some of this variable acuity may be accounted for by splitting the image sequence into bands along chromatic and spectral boundaries and processing the bands individually. The encoding of the chrominance components is briefly explained. The issues involved in the selection of the subband boundaries are reviewed. The bandsplitting technique is outlined. Selected recent works in image processing are shown to provide insights into how the subbands can be grouped based on visual sensitivity to distortion. Finally, simple tests are performed to verify this proposed grouping.

3.1 Color

The decomposition begins with the image sequence being divided into luminance and chrominance components. Color is encoded using well known characteristics of the HVS's spatial sensitivity to stationary chromatic detail and less well explored characteristics of the visual system's sensitivity to temporal chromatic detail.

3.1.1 Spatial Chromatic Resolution

As the goal of this work is near-broadcast quality resolution of moving imagery, I have adopted the spatial chromatic resolution specified by the NTSC color television standard. This standard calls for matrixing the red, green, and blue signals into a luminance (Y) and two chrominance signals, I and Q. The I and Q signals have a horizontal resolution of 120 pixels and 50 pixels per line respectively. It is widely recognized that human sensitivity to stationary detail is nearly equal in the horizontal and vertical direction. This implies that vertical chrominance resolution could be reduced to be comparable to the horizontal resolution at little cost in perceived image quality. While the resolution of the I signal was initially set at 120

pixels per line, it is common practice today by television manufactures to also limit the horizontal resolution of the I signal to that of the Q signal. Consequently, a resolution of 64 pixels across by 48 pixels high was chosen for the I and the Q channels. The exact dimensions were motivated in part by the dimensions of the framestore used for this work; the 640 by 480 framestore of a Symbolics Lisp machine.

3.1.2 Temporal Chromatic Resolution

There is much less agreement about the HVS's response to temporally varying chroma. However, recent work in the fields of neurophysiology and psychophysics has established that the critical fusion frequency¹ for chrominance is significantly below that of luminance [Glenn 1983] [Kelly 1979] [Kelly 1983]. In studies of apparent motion², rapid changes in chromaticity have been shown to be insufficient to produce the impression of smooth motion [Anstis 1980] [Ramachandran 1978].

As part of this project, tests were carried out to measure the perceptibility of temporally varying chrominance noise in the presence of different luminance signals. The visibility of the chrominance noise was measured as a function of both

¹frequency at which the HVS begins to function as an integrator

²the percept of smooth motion generated by displaying similar images in quick succession

the power of the noise and the correlation to the luminance signal. It was found that when chrominance noise was correlated with luminance detail, a significant amount of noise was undetectable. This closely resembles the kind of error that occurs if the chrominance components are updated at a slower rate than the luminance component. Pursuing this reasoning, a number of short sequences were animated with the chrominance update rate reduced by a factor of two (to 15 frames per second). In these simple test cases, the effect varied from imperceptible to marginally perceptible.

As a final result of these considerations, the color information in our system is extracted from the red, green and blue signals using a standard NTSC matrix transformation to obtain an I and Q channel. These channels are coded as 64 by 48 arrays per chrominance frame at a rate of 12 frames per second.

3.2 Partitioning of the Luminance Spectrum

Once the chrominance components are extracted for separate processing, the remaining luminance signal must be encoded. As illustrated in figure 1.1 , the sub-

band coder is responsible for transforming the luminance sequence into a series of component spectral channels, each of which is then fed separately to a block coder. This section uses the architecture diagrammed in figure 1.1 to motivate the segmentation of the luminance spectrum.

Because the subband transforms are VQ'ed, the representation of the individual subbands should in some way be optimized for vector coding. One way to achieve this is to segment the luminance spectrum in such a way that energy in many of the subbands have a uniform representation. This is realized by the representation of high spatial frequency as a series of oriented gradients. In components which are composed of gradients with identical orientation, the statistical variation is confined to variations in the slope and the magnitudes of these gradients. This reduces the statistical space that the VQ must quantize.

One way to distribute the quantization distortion due to VQ according to some perceptual metric is to subdivide the luminance spectrum in such a way as to approximate the spatio-temporal bandlimiting of the early visual pathway. A short, critical review of this approach is presented at the end of this section. However, given that the bandsplitter should be a realizable, linear operation which models the observed behavior of the early visual pathway, it should filter separably along

the axes of space and time, the filters should have local regions of support, and the bandsplitting should take place on octave boundaries.

The justification for bandsplitting on octave boundaries is different for the temporal and spatial cases. The HVS's response to even simple temporally varying stimuli is demonstrably non-separable. Accordingly, to model even the contrast sensitivity threshold function of the HVS would require a complex and extremely non-linear operator [Kelly, 1979]. By comparison, a simple halfband filter separably applied along temporal axis, captures the essence of the bimodal³ nature of the HVS's temporal response. Spatially, available evidence suggests that the subbands should be arranged as a series of oriented, bandpass channels [Marr 1982] [Hildreth 1985] [Hubel 1979] [Schiller 1976] The particular bandwidths of the individual channels vary as a function of the retinal eccentricity of the receptive field. However, measurements indicate that for a fixed retinal position, the bandwidths vary in increments of roughly single octaves [Wilson 1979]

Finally, the partitioning of the luminance spectrum should be influenced by the particulars of the source image sequences. In this report, the coding performance is tested using natural scene data which is digitized from a 24 frames per second

³temporally, the HVS functions as either an integrator or differentiator

(fps) film recording of a popular TV series. The spatial and chromatic detail of this data is quite high and allows the selection of spatial filters from a wide range of possibilities.⁴ However the temporal resolution of this test data is limited because natural scene data contains significant temporal aliasing when recorded at 24 fps.

For the work presented here, the luminance spectrum is divided by separably filtering along the axes of space and time using halfband filters with small regions of support. The relative positions of the resulting spectral components are as illustrated in figure 2.1(a) In this figure, the labeling of the components reflects the filters used to isolate the component energy. The characters are either L or H representing low pass or high pass. The three characters in each label refer to, from left to right, the temporal, vertical and horizontal filters.

It is interesting to note that these components can also be given physically meaningful interpretations. The four temporally low passed channels can be thought of as isolating energy from stationary objects. The four temporally high pass channels isolate energy due to oriented movement in the scene. These interpretations can be used to give the following intuitive meanings to the 3D subbands :

⁴note that the data was progressively scanned eliminating interlace effects

LLL	stationary 'blurs'
LLH	stationary vertical edges ^a
LHL	stationary horizontal edges
LHH	stationary diagonal edges
HLL	moving 'blurs'
HLH	moving vertical edges
HHL	moving horizontal edges
HHH	moving diagonal edges

^anote that high pass filtering horizontally accents vertical edge detail

Because the poor temporal resolution of the test sequences makes a further level of temporal filtering somewhat meaningless, further levels of the pyramids are built by spatial filtering in two dimensions. The work in this report uses a two level pyramid representation. The second level of the pyramid is generated by separable filtering of the LLL channel from the first pyramid level as illustrated in figure 3.1

Earlier in this section (on page 25) the claim was made that spatio-temporal bandsplitting could be used to model some of behavior of the early visual pathway for purposes of evaluating subjective image quality. It should be noted that the use of linear bandsplitters for perceptually optimal allocation of visual distortion is currently *not* a defensible position from a purely signal coding point of view. When

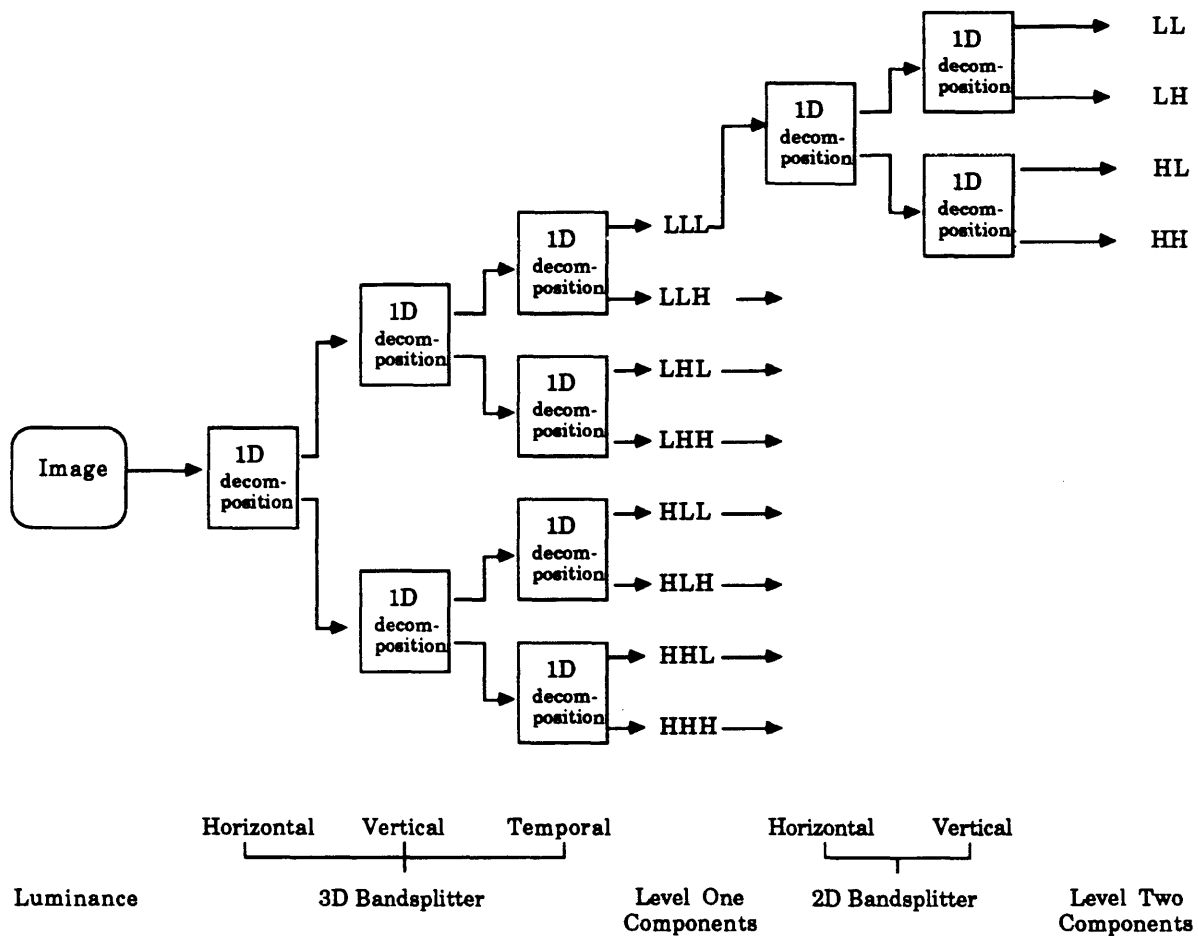


Figure 3.1: Two level pyramid decomposition

distortion is measured against one of the more common objective metrics (MSE for example), spectral partitioning based on perceptual models has no inherent claim to optimality. To complicate the issue, there is, to date, no simple linear process which is widely regarded as a complete model of early visual behavior on anything but the simplest input. Further, of the visual models that have demonstrated some utility, simple linear bandsplitting is by no means the most effective or complete.

There are, however, a number of points which address these criticisms. The first is that it is widely acknowledged in the image coding community that objective metrics do an incomplete job (at best) of predicting the subjective quality of an image. Consequently, it is now common practice in more classical coding techniques to redistribute error to areas of reduced visual sensitivity, even when such processing does not substantially alter the MSE of the decoded image. A second point in favor of using visual models, if still incomplete, is that a growing base of knowledge from neurophysiology, psychophysics and computational studies of early visual processes has increased our ability to explain selected responses of the HVS. While there is currently no unified model based on bandpass oriented channels which explains all observed behavior of the HVS, recent work continues to explain an expanding range of visual behavior in terms of models based on these channels. This work has also illustrated that, independent of the neurophysiological role that they may play in

mammalian vision, spatio-temporal subbands form a rich set of primitives on which visual percepts can be built.

3.3 Bandsplitting Techniques

The bandsplitting technique used is filtering followed by decimation. The filtering is performed by convolution with separable, one dimensional, quadrature mirror filters (QMF's). The advantages of this technique are that the separable filtering makes the reconstruction computationally tractable and that the transform can be recursively applied to create multiscale representations. Transforms using QMF's are localized in both space and frequency. They are also complete, which means that the number of transform coefficients is equal to the number of input samples. They are orthogonal which enables the same kernels to be used for both the transformation and the inverse transformation. Their principle drawback is the amount of aliasing between the component subbands. However, on reconstruction, the aliasing from the different bands cancels resulting in exact or nearly exact reconstruction. Different QMF's are used for the temporal and spatial filtering. The spatial filtering is done using a set of 9-tap QMF's which were optimized for this

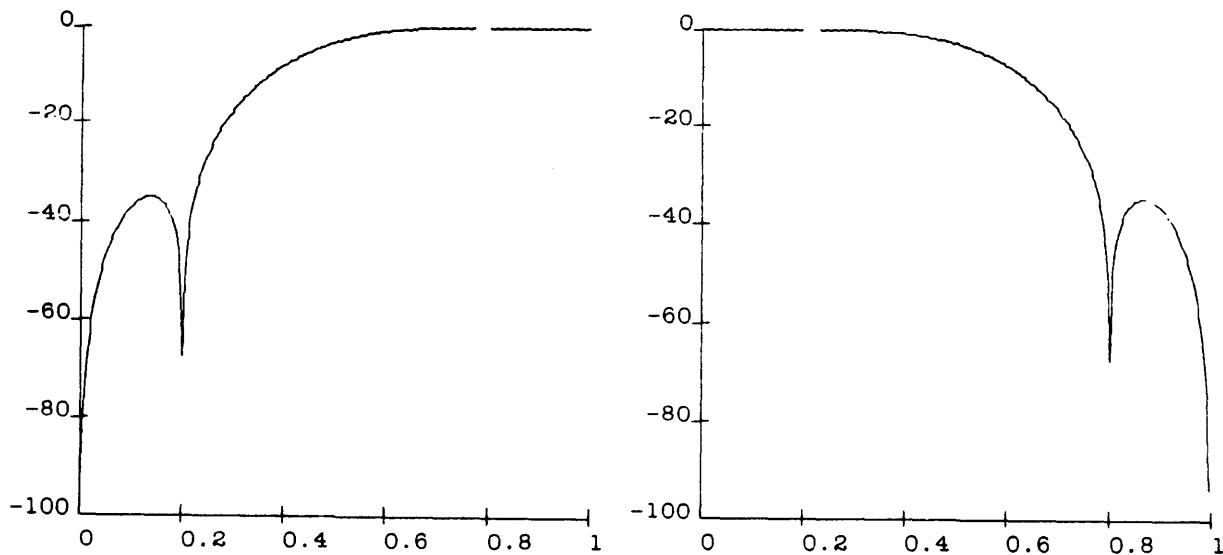


Figure 3.2: *Magnitude Response of 9-tap QMF Filter Pair*

purpose [Adelson 1987]. Figure 3.2 shows the magnitude response of these low pass and high pass filters. The filter pair used for the temporal bandsplitting is a two point Haar transform⁵⁶. Figure 3.3 shows the response of these filters.

These filters are separably combined in manner illustrated in figure 3.1 to produce the subbands which are later passed onto the VQ. Figures 3.4 through 3.9 show the subbands which are produced from two different frame pairs from a single test sequence. These figures use the labeling of figure 3.1 for the subimages. In the first frame pair, a woman is walking down an alley which is motionless. The

⁵the low pass is simple averaging and the high pass is simple differencing

⁶can also be thought of as a two point DCT

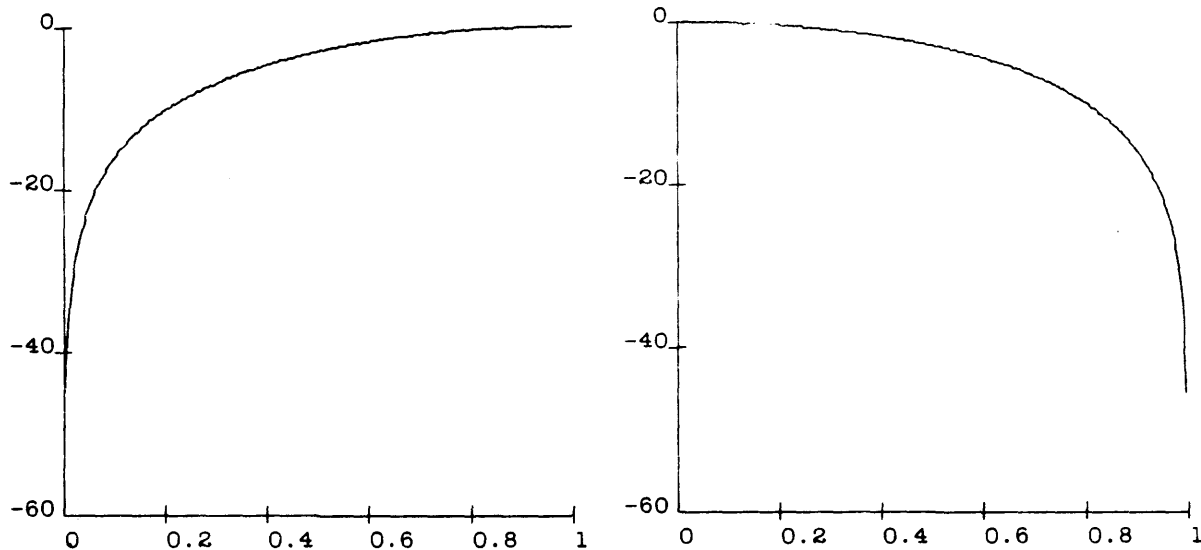
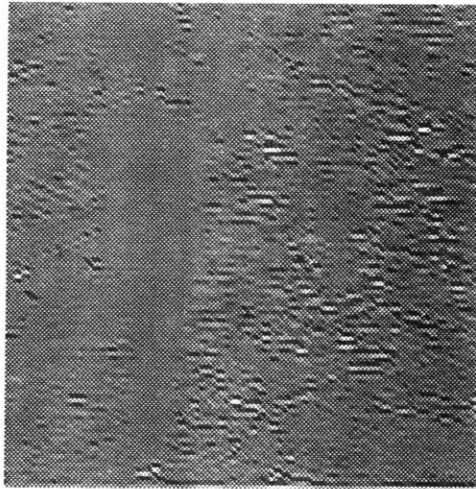


Figure 3.3: *Magnitude Response of Haar Transform Pair*

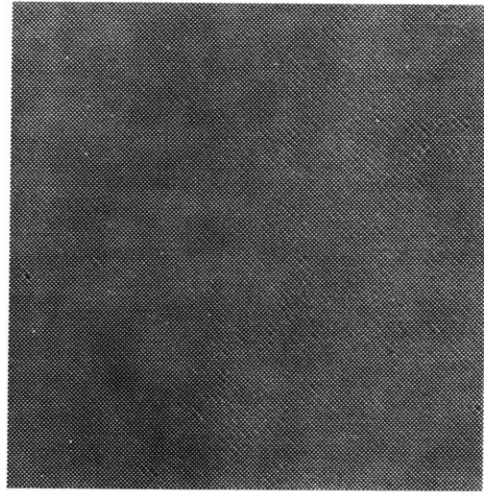
subbands of figures 3.4 through 3.6 reflect this by displaying strong stationary detail in channels LLH, LHL, LH and HL. The stationary background deprives the temporal high pass channels of energy everywhere except in the area of the woman in motion. In the second frame pair, this situation changes as the camera pans the alley in order to track the walker. The subimages of figures 3.7 through 3.9 reflect this by redistributing a portion of the energy to the moving components⁷.

These figures serve to point out a number of characteristics of the subband representation which will later prove useful. At both pyramid levels, the high frequency spatial detail always appears as oriented gradients. For all natural sequences that

⁷note that the subbands are displayed with a gain of 4.0 relative to the LL component



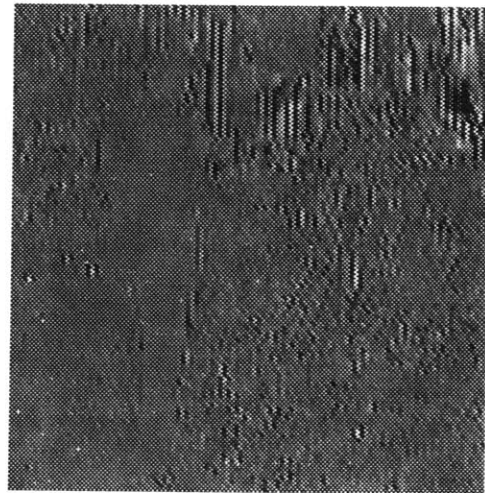
LHL



LHH

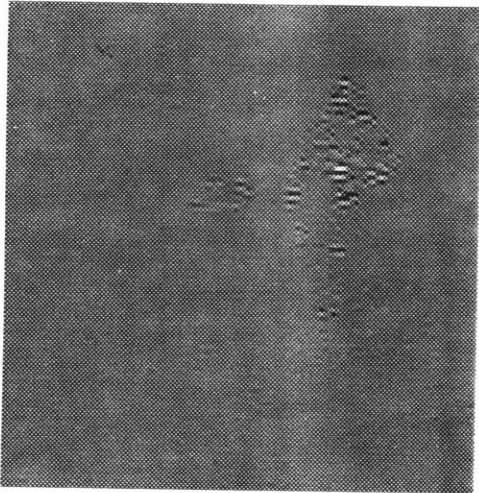


LLL

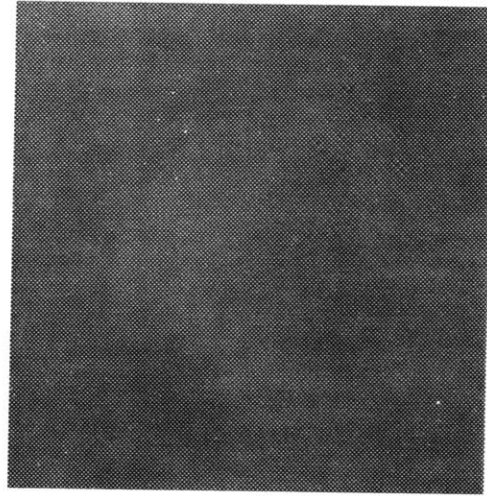


LLH

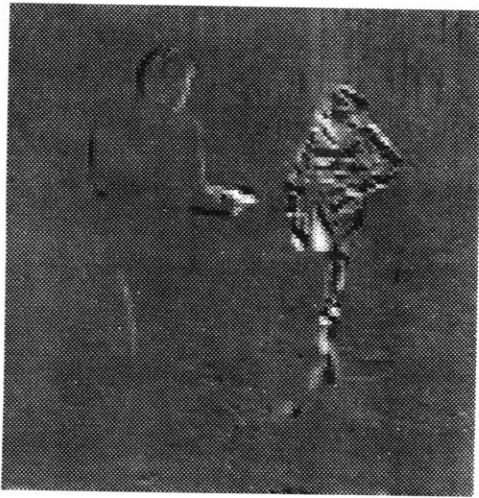
Figure 3.4: *First Image Pair: Level One Temporal Lows*



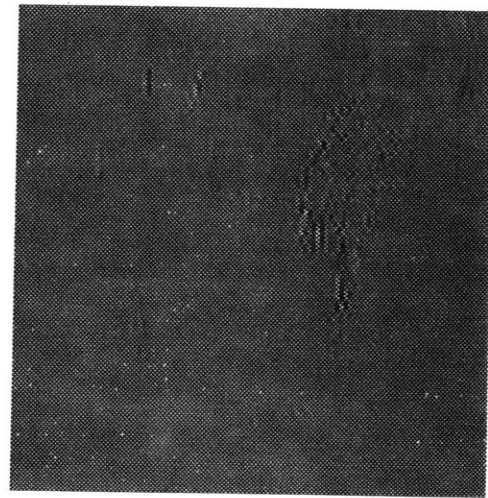
HHL



HHH



HLL



HLH

Figure 3.5: *First Image Pair: Level One Temporal Highs*

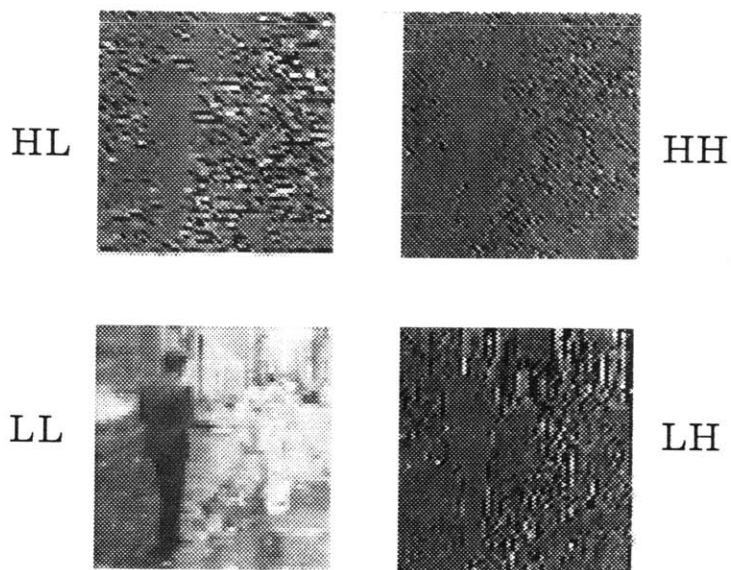
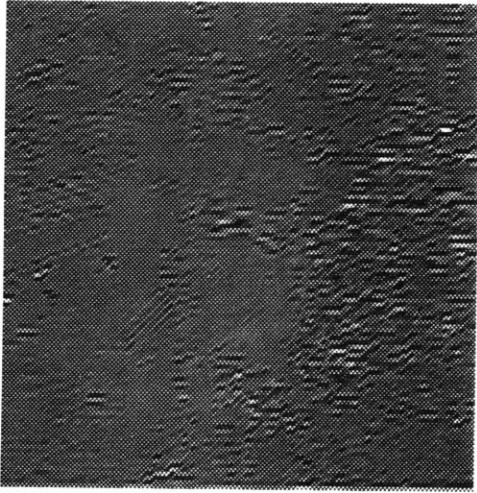
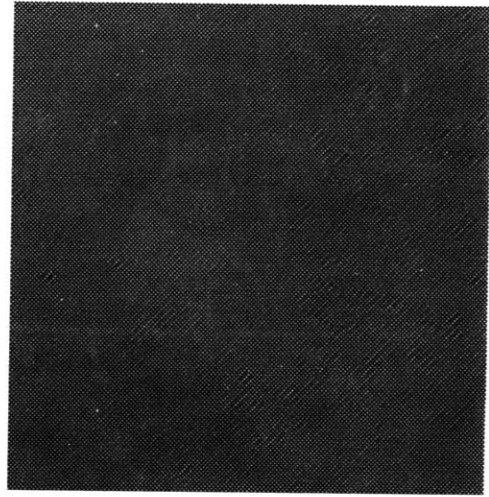


Figure 3.6: *First Image Pair: Level Two*

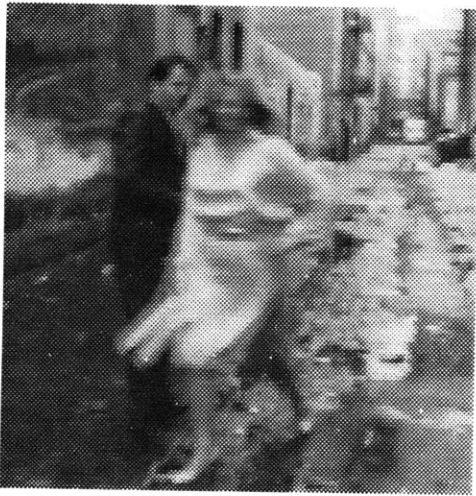
were tested, the diagonal components consistently displayed minimal energy, suggesting that they could be heavily quantized, if not altogether eliminated. Finally, because the 3D subband representation is effective at isolating both motion and detail in local areas of the image, it offers the potential for significant entropy savings in sequences where large portions are either stationary or somewhat out of focus.



LHL



LHH

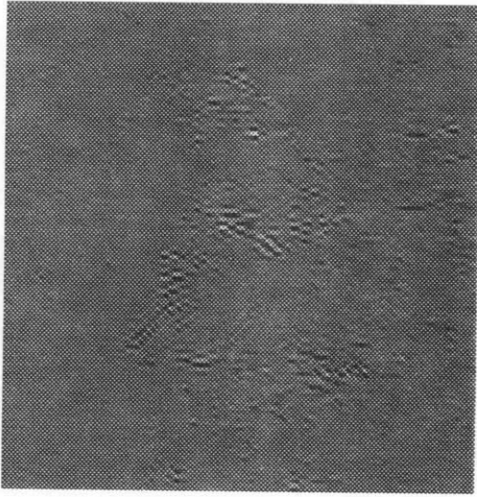


LLL

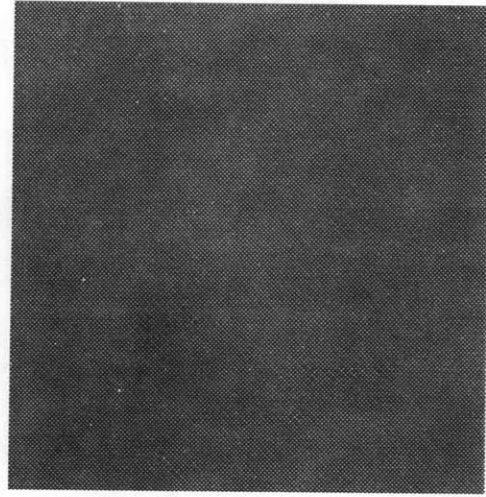


LLH

Figure 3.7: *Second Image Pair: Level One Temporal Lows*



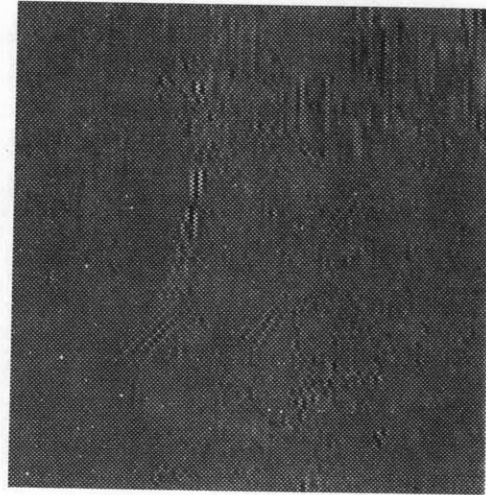
HHL



HHH



HLL



HLH

Figure 3.8: *Second Image Pair: Level One Temporal Highs*



Figure 3.9: *Second Image Pair: Level Two*

3.4 Visual Sensitivity to Energy Distribution in the Subbands

In order to distribute the quantization distortion due to VQ according to some perceptual criteria, it is necessary to have at least a qualitative understanding of the visibility of distortion in the individual subbands. The purpose of this section is to assemble related work on the coding of images based on visual sensitivity and to interpret the results in terms of the spatio-temporal subbands. From this interpretation, inferences are drawn to qualitatively group and order the various subbands based on the visibility of distortion. These proposed groupings are then tested using a number simple test images. While the results reported here are not yet absolute or conclusive, they later serve as perceptual guidelines for the distribution of the quantization distortion among the components.

An exhaustive set of tests for the threshold sensitivity to quantization distortion in the spatio-temporal subbands is far beyond the scope of this thesis. Fortunately, a large amount of recent work in image processing has focused on exploiting the varying acuity of the HVS to distortion in different spectral components. The representation of a sequence as a collection of 3D subbands offers a common framework into which much of these results can be incorporated. Work which can be used

to infer the distortion sensitivity of the spatial subbands includes work on spatial masking, two channel coding, and spatial filtering for HDTV. Results from multi-channel coding for HDTV and contrast threshold measurements can be used to postulate sensitivity of the temporal subbands.

Work on spatial masking in images has demonstrated that acuity for distortion is significantly reduced in the presence of sharp intensity changes. This pattern masking effect has demonstrated utility for efficient coding of images [Netravali 1977]. Distortion due to quantization of the spatial highs components would, by its nature, be localized to areas of high spatial frequency and would therefore be subject to these masking effects. Troxel and Schreiber [Troxel 1980] used this to minimize noise due to coarse quantization of the high frequency components. In their system, the highs are finely sampled but coarsely quantized and the lows are subsampled but finely quantized. By demonstrating the amount of scalar quantization that can be inserted into the highs channels and still go undetected, this work does much to suggest that the spatial high frequency subbands are far less noise sensitive than the spatial lows. Work on spatial prefiltering of HDTV signals has demonstrated that in the top octave of spatial detail, the visibility of distortion in the diagonal frequencies is greatly reduced relative to that of either the horizontal and vertical components [Wendland 1983] [Wendland 1984]. This result suggests

that the diagonal spatial component can be more coarsely quantized at little perceptual cost. Both psychophysical measurements of contrast thresholds [Kelly 1979] and work on multi-channel coding techniques for HDTV [Glenn 1983] [Glenn 1984] have presented evidence that there is a substantial decrease in the visibility of high spatial detail when the retinal position of the detailed object is varying temporally. In terms of the subband representation, this work implies that the distortion sensitivity of the components representing 'moving edges' is reduced relative to the sensitivity of the components representing the 'stationary edges'. However, such an inference must be approached cautiously because the conditions under which a human will track a moving object are currently not predictable.

Based on the work reviewed above, a relative hierarchy for the sensitivity to quantization distortion of the spatio-temporal subbands is proposed. For the subbands resulting from spatial bandsplitting, the hierarchy is, in order of most sensitive to least sensitive :

- o *LL*
- o *LH* and *HL*
- o *HH*

For the spatio-temporal subbands, the order of most sensitive to least sensitive is :

- o *LLL*
- o *LHL, LLH, and HLL*
- o *HLH and HHL*
- o *LHH and HHH*

This proposed hierarchy of distortion sensitivity was subjectively evaluated using three still test images and one sequence. Figures A.1 through A.3 in appendix A show the original stills. Figure A.4 shows three frames from the 24 frame *alley* sequence. The overall method that was used to evaluate the noise sensitivity of the subbands was to first subdivide the image into subbands, individually introduce distortion into a given band, reconstruct the image, and subjectively evaluate the effect. The quantization distortion was generated by applying progressively coarser scalar quantization. The work was divided into three tests. In the first two tests, the noise sensitivity of just the spatial subbands was evaluated at two pyramid levels. In the third test, the sensitivity of the spatio-temporal subbands was examined. Figure 3.10 illustrates the components which were used in the first two tests. The components examined in the third test are displayed in figure 2.1(a).

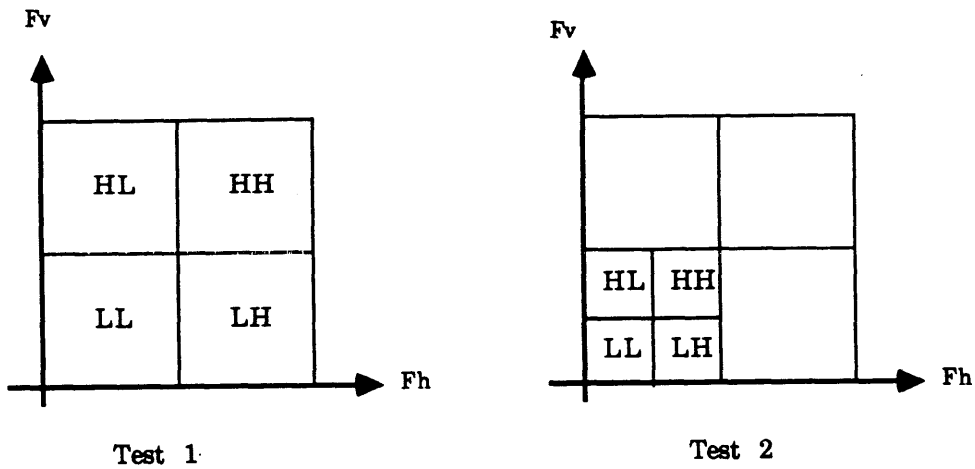


Figure 3.10: *Components for first two sensitivity tests*

In the first test, each image was transformed into four spatial subbands using the 9-tap QMF filter pair. It was quickly noticed that the diagonal component could be eliminated altogether. The horizontal and vertical detail components displayed roughly equal sensitivity⁸ The spatial lows component was most sensitive. These results are demonstrated in the figures which follow. Figures 3.11 and 3.12 compare an original with the original minus the diagonal components. Figures 3.13 through 3.15 show the same image with 24 db of distortion⁹ added to (respectively) the spatial lows component, the horizontal component, and the vertical component.

⁸note all the test images contained significant amounts of detail at both orientations

⁹calculating 6 db of distortion for each bit truncated from the word length

The second test proceeded in a like fashion as the first. In this case, the components which were quantized were the second pyramid level components which resulted from further subdividing the spatial lows component from test one into four more spatial subbands. Figures 3.16 through 3.19 show the results of individually quantizing these components to four bits per sample. Note that at this level of detail, the diagonal energy is no longer expendable. Beyond this, the grouping based on distortion sensitivity seems to be the same at level two as it is at level one. Namely, that the spatial lows are the most sensitive, the horizontal and vertical are of roughly equal sensitivity and the diagonal detail is the least sensitive.

The third test concentrated on the relative distortions between spatio-temporal bands. Here distortion due to scalar quantization was introduced into the individual subbands and the resulting degradation in the reconstructed sequence was subjectively compared. This test supported the suggestion that both the moving and the stationary diagonal components (LHH and HHH) could be discarded completely. Of the remaining components, the moving edge detail components (HLH and HHL) were found to be the least sensitive to distortion while the LLL channel was the most sensitive.

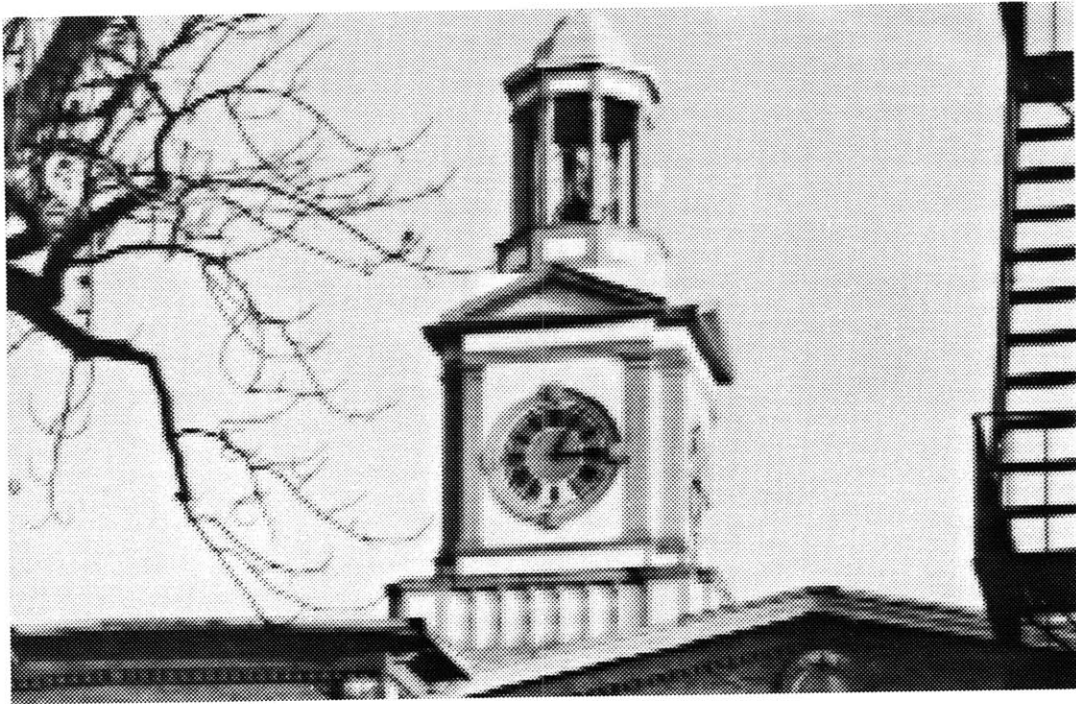


Figure 3.11: *Zoom from Original image : 'The Mill'*

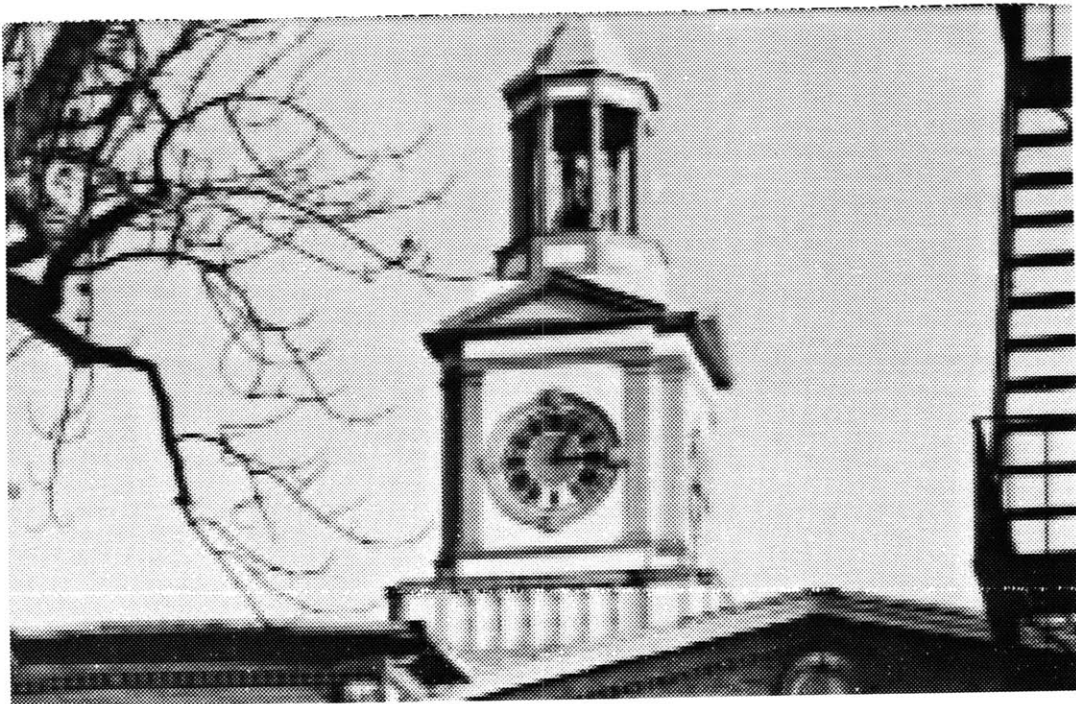


Figure 3.12: *The Diagonal Subband Removed*



Figure 3.13: *Quantization Noise in the Level One LL Channel*



Figure 3.14: *Quantization Noise in the Level One HL Channel*



Figure 3.15: *Quantization Noise in the Level One LH Channel*



Figure 3.16: *Scalar Quantization of the Level Two LL Channel*

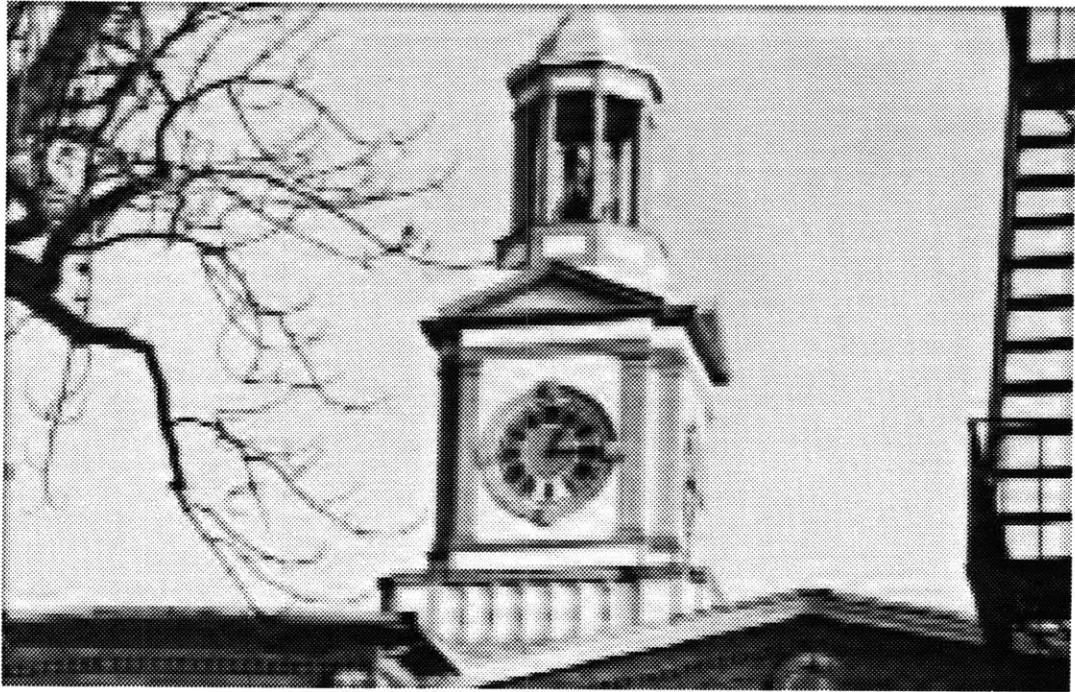


Figure 3.17: *Scalar Quantization of the Level Two HL Channel*



Figure 3.18: *Scalar Quantization of the Level Two LH Channel*

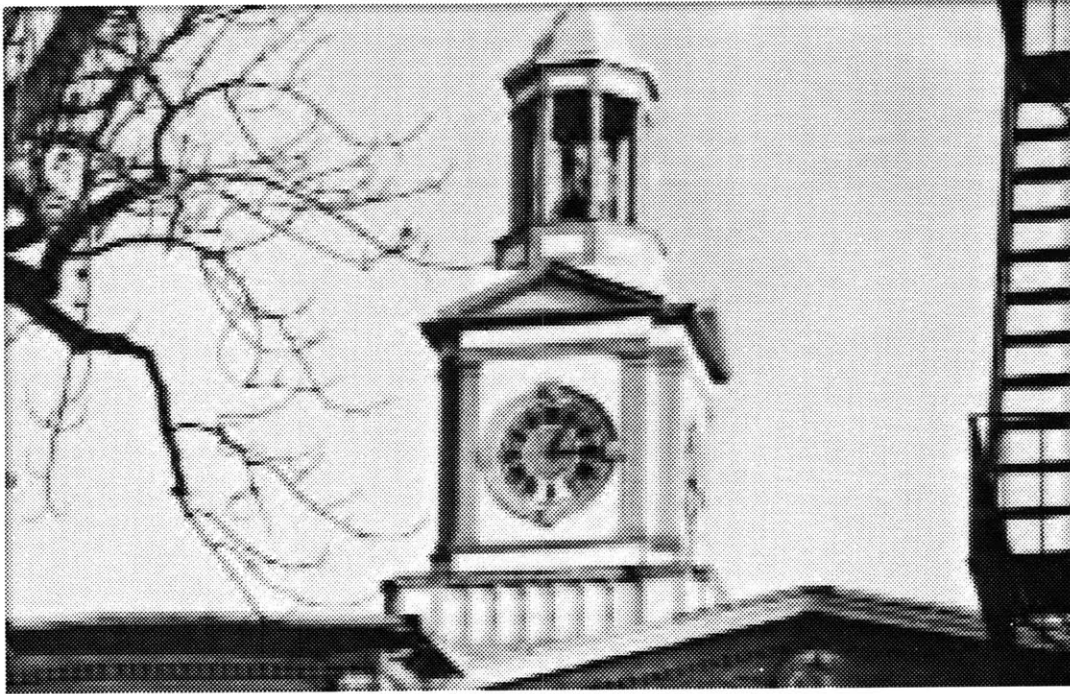


Figure 3.19: *Scalar Quantization of the Level Two HH Channel*

Based in part on these simple tests, the following relative distortion sensitivity is adopted for the remainder of this work. This list will be used in the next chapter as a guideline for the distribution of error due the VQ. From most sensitive to least sensitive :

- The level two spatial lows (LL)
- The level two detail components (LH, HL and HH)
- The level one components LLH, LHL and HLL)
- The moving highs (HLH and HHL)
- Both the stationary and the moving diagonals will be routinely discarded from the top level of detail

This ordering is admittedly dependent on the particulars of the test sequence. It is not difficult to imagine very reasonable images which would violate this proposed hierarchy. These distortion guidelines are useful because they are applicable on average and they agree with results from other previous work.

Chapter 4

Vector Quantization of Subbands

This chapter investigates the application of vector quantization (VQ) to the subbands. A straightforward application of VQ to the subband data would be to either create a separate codebook for each channel or to lump all the channels into one training set to create a global codebook. These examples represent extremes in codebook size and quantization error respectively. This chapter proposes four techniques for more efficiently exploiting the characteristics of the subband representation by

using the VQ to eliminate statistical redundancy across scale, orientation and time.

These four methods are :

- Coding the spatial highs with orientation independent codebooks.
- Cascading of codebooks through indirect addressing of codebook entries.
- Combining codebooks to eliminate similar entries across scales.
- Comparing codebooks to eliminate redundancies between temporally neighboring images.

The previous work in VQ has been divided in its approach to the generation of the codebook. One approach has been to create codebooks which are specific to an image or temporal neighborhood of a sequence and rely on dual transmission of the new codebooks and the channel symbols. In this multiple codebook approach, a primary determinant of image quality for low bandwidth image transmission is the size of codebook that must be transmitted and updated. The four techniques developed here enable significant reduction of the codebook size by coding each of the component bands in such a way that the resulting codebooks have many entries in common which can then be eliminated.

4.1 Orientation Independent Codebooks

This section focuses on exploiting the similarities between the subbands containing the spatial high frequencies. The nature of these similarities is demonstrated by comparing the LH and HL channels of figure 3.6 In this figure, the energy in the two channels is arranged in a series of gradients with constant orientation. This ensures that, to within the constant of orientation, the statistics of the two channels will have fundamental similarities. The codebooks which are used to jointly encode two such channels can be created by arranging the channels into a single training set such that all the gradients are aligned. This can be done by simply rotating one of the subimages ninety degrees and concatenating it to the other subimage. The codebook which is generated from such a training set contains no information about the orientation of the original subimages. Hence the name orientation independent codebooks.

Only the subbands with the horizontal or vertical edge detail are combined to form the orientation independent codebooks. The bands with the diagonal edge detail are excluded both because they are difficult to decode in real time, and because they contain two orientations within each subband. Another restriction on the selection of edge components which form the individual training sets is that

they all come from the same distortion sensitivity group (as shown at the end of chapter three). This is so the perceptibility of the resulting quantization noise will be uniform. The specific distribution of distortion due to VQ varies with the particulars of the algorithm used to generate the codebooks. However, most vector quantizers attempt to distribute the noise due to quantization evenly throughout the vector space. Given this, it becomes necessary to group only those components which have similar visual sensitivity to noise.

As part of this report, tests were performed to evaluate the effects of VQ with orientation independent codebooks. These tests were conducted using only the spatial subbands of the two monochrome test images shown in A.1 and A.2 of appendix A. For each image, two processes were used to isolate the highs channel as shown in figure 4.1 In the first case, the spatial highs signal is simply the difference between the original image and a low passed version of the image. The spectrum of the resulting 'nondirectional' highs signal is shown in 4.1(a) In the second case, two separate oriented 'directional' highs signals are generated by separable filtering with QMF filters followed by decimation. Figure 4.1(b) displays the combined spectrum of these two components. The nondirectional highs component was coded using standard VQ and the two directional highs components were combined and coded using the orientationless codebooks. The reconstructed images were com-

Block Dimensions	Codebook Length	MSE	
		Nondirectional	Directional
8 x 8	256	138.4	96.6
	512	113.0	64.5
	1024	69.6	35.2
4 x 4	256	102.0	89.7
	512	88.3	75.6
	1024	71.8	56.5
2 x 2	256	61.4	35.2
	512	40.3	29.0
	1024	30.8	23.6

Table 4.1: *Mill*

pared both subjectively and objectively using MSE. The tests were conducted over a range of codebook lengths and block dimensions in order to isolate the effects of extracting the edge orientation from the errors due to the size of the codebook or the dimensions of the blocks. Tables 4.1 and 4.2 present the results of these tests. Both numerically and subjectively, the use of orientation independent codebooks offered visible improvement in the quality of the reconstructed image.

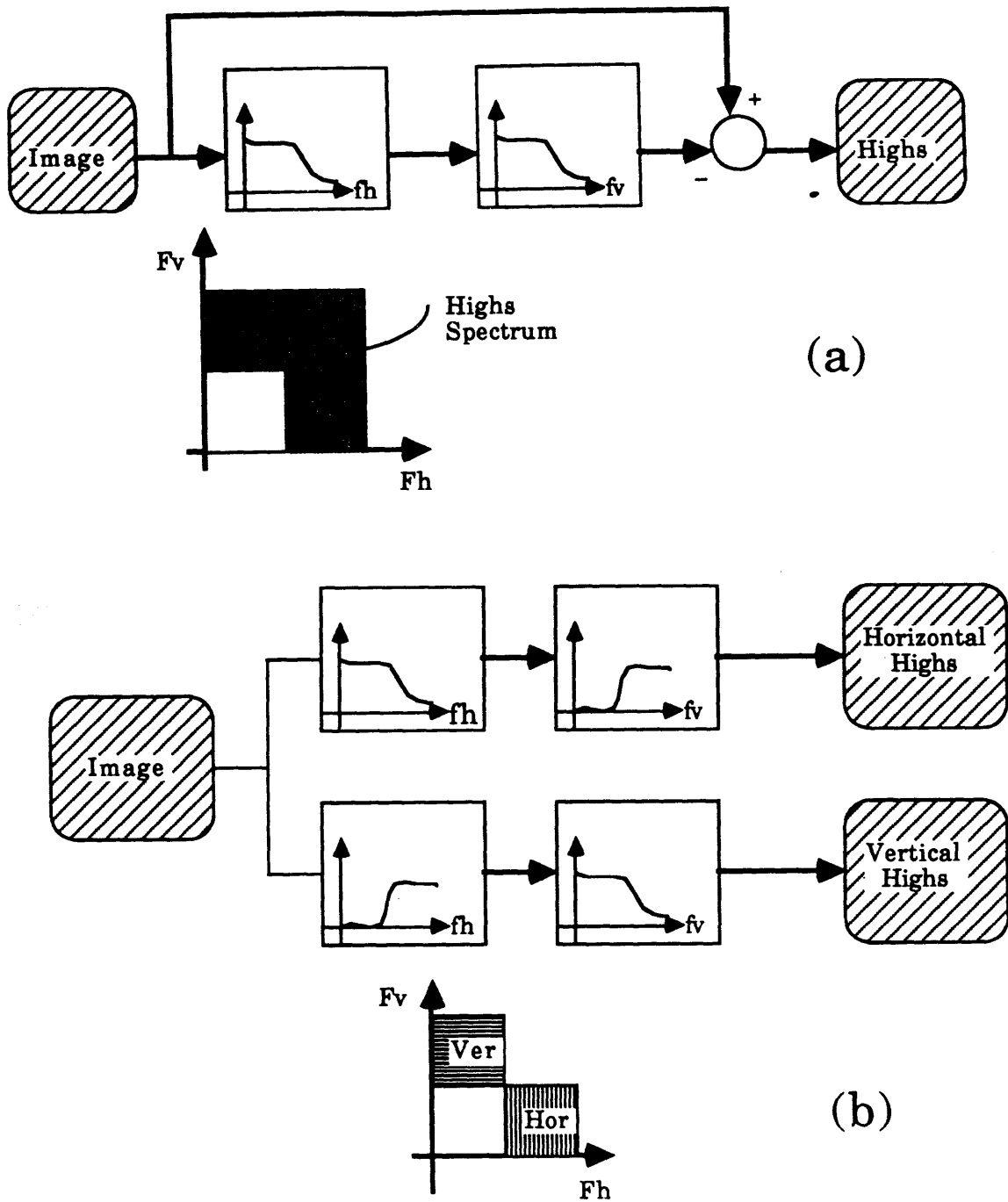


Figure 4.1: *Generating the spatial highs channels*

Block Dimensions	Codebook Length	MSE	
		Nondirectional	Directional
8 x 8	256	205.5	141.3
	512	167.2	94.4
	1024	105.2	53.1
4 x 4	256	158.3	137.8
	512	137.4	115.3
	1024	111.2	87.0
2 x 2	256	103.3	56.8
	512	81.7	47.0
	1024	69.6	38.4

Table 4.2: *Boats*

4.2 Cascading of Codebooks

This section details a method for further reducing the size of the codebook by exploiting the fact that image components have statistical redundancies at more than one block size and that these redundancies can overlap. This technique specifies that a codebook entry for each channel symbol contain not a collection of image intensities but rather a list of indexes into another codebook which contains the intensities. Figure 4.2 compares VQ using a single codebook with VQ using two codebooks which have been cascaded together. In 4.2(a), the channel symbol addresses a codebook entry which contains the image intensities for a four by four

patch of the image. In 4.2(b), the codebook entry pointed to by the channel symbol contains four pointers into the second codebook. Each entry in this second codebook contains the intensities for a two by two patch of the image. Therefore, these two code tables, when cascaded together, form the equivalent of a table whose dimensions are four by four.

There are several reasons why codebook cascading is advantageous. The first is that for a given set of block dimensions and codebook lengths, indirect codebook addressing offers upward of a fifty percent savings in codebook size at the cost of some image degradation. More importantly, cascading codebooks at the finer scale pyramid level is an effective way to share codebook entries between the pyramid levels. This is important because the self similarity of the pyramid representation gives rise to statistical redundancy across scales. In order to exploit this redundancy, it is necessary to either use identical codebook dimensions at both scales, indirect codebook addressing, or some sort of codebook interpolation.

The task of generating a pair of indexed codebooks can be thought of as identifying statistical redundancies at multiple scales of an image. There are a number of ways to do this. In this work, the VQ is run at two scales and the results are combined. This process is illustrated in figure 4.3 for the example of a two by two

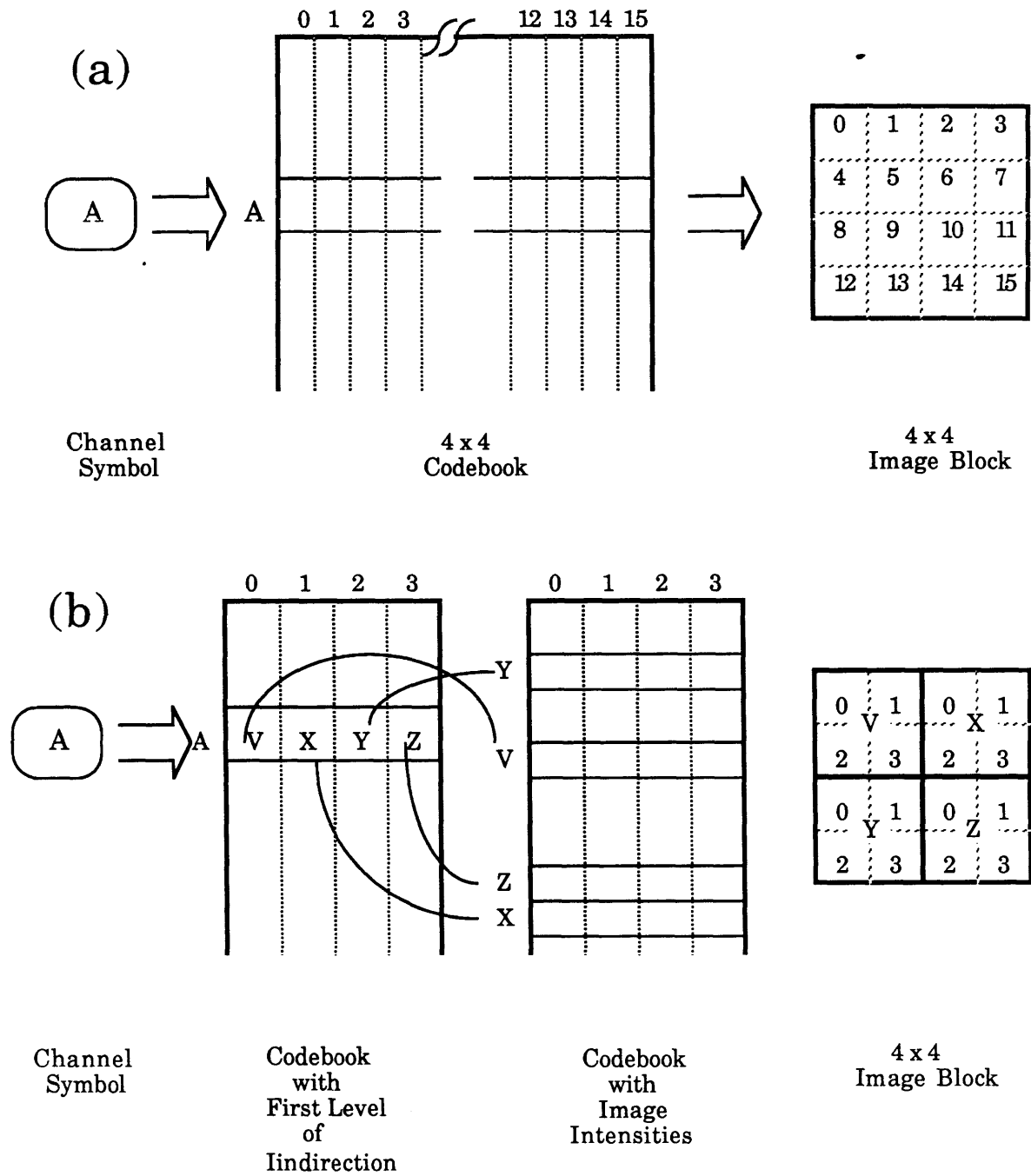


Figure 4.2: VQ Decoders, (a) Direct and (b) Indirect

index table driving a two by two codebook. In 4.3(a), an image (in this case, the subband of an image) is VQ'ed and reconstructed using a two by two block size. The statistics of the reconstructed image are therefore altered so that they cluster around the two by two vectors in the codebook. In 4.3(b), this reconstructed image is again VQ'ed using a four by four block size. Figure 4.3(c) illustrates how each two by two quadrant of the entries in the four by four codebook are then replaced by the nearest match from the two by two codebook. This produces a four by four codebook in which each entry is actually four selections from the two by two vector space. As shown in 4.3(d), the output image is reconstructed using the the altered four by four codebook along with the original four by four channel symbols from 4.3(b) In terms of the population of the vector space, this process can be thought of as overlaying a two by two vector set onto the statistical space of the image and executing a 'snap to grid'. The four by four vector set is then selected from this altered statistical space.

In order to evaluate the effect of indirect codebook addressing, the technique illustrated in figure 4.3 was performed on the oriented spatial highs channels of the test images in figures A.1 and A.2 The reason that the oriented spatial highs channels were chosen was that it is these channels that will appear at both scales in the two level pyramid decomposition described in section 3.2 Tables 4.3 and 4.4

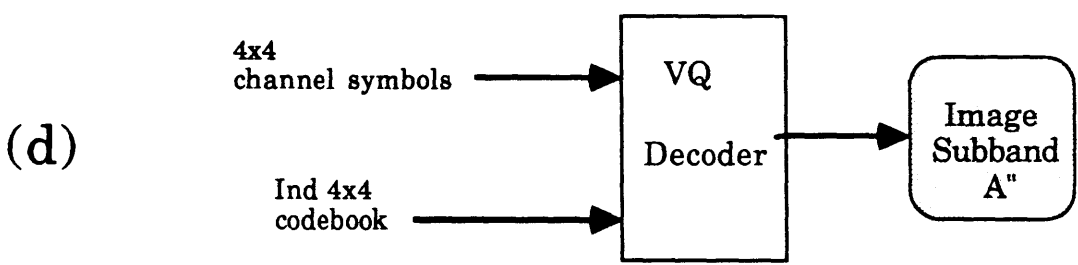
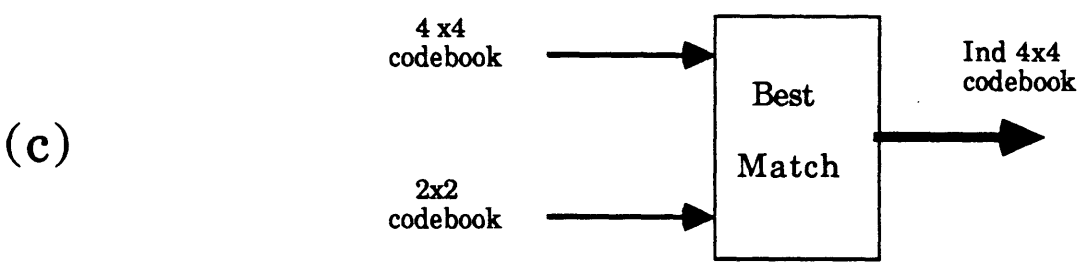
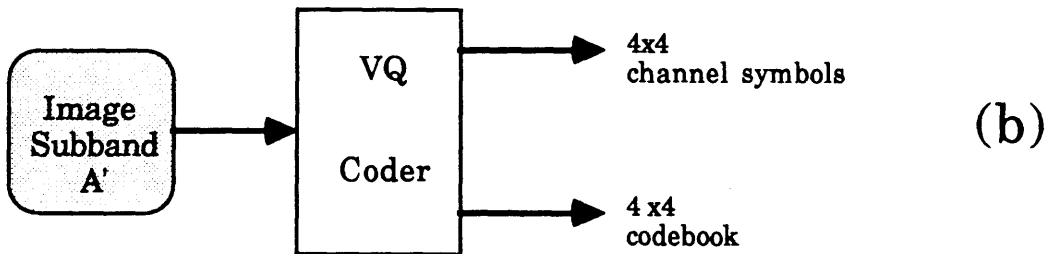
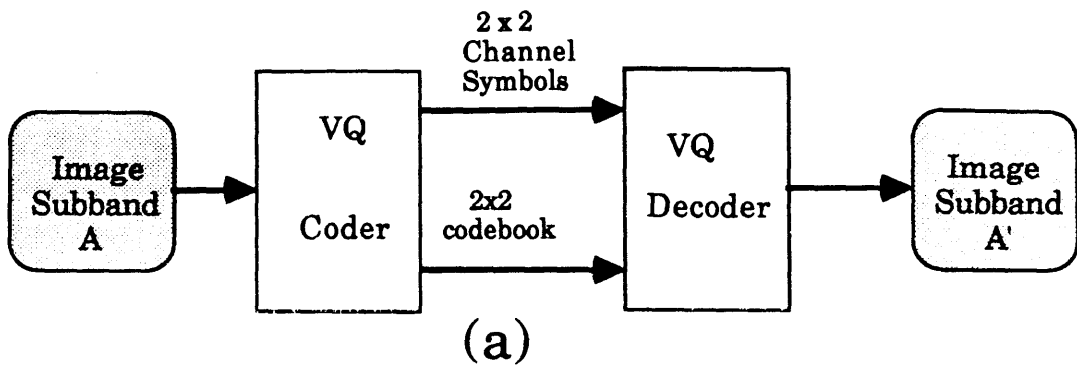


Figure 4.3: Generation of Indirectly Addressed Codebook

Block Dimensions	Codebook Length	MSE	
		Direct	Indirect
4 x 4	256	89.7	104.8
	512	75.6	96.3
	1024	56.5	82.4

Table 4.3: *Mill*

Block Dimensions	Codebook Length	MSE	
		Direct	Indirect
4 x 4	256	137.8	163.4
	512	113.3	148.7
	1024	87.0	128.3

Table 4.4: *Boat*

compare the performance of indirect codebook addressing against straight VQ using orientation independent codebooks. The data in the column marked 'Direct' is the MSE resulting from vector coding the horizontal and vertical subimages using one orientation independent codebook for both subbands. The data in the 'Indirect' column gives the MSE resulting from an additional level of indirection in the orientation independent codebook.

4.3 Multiscale Codebooks

This section investigates methods of exploiting the statistical redundancies present across the different scales of a multilevel pyramid representation of an image. The fact that a broad range of natural images exhibits detail across several scales suggests that there will be some similarity in the statistics of the subband components. This section begins by examining how image features at different scales are represented by pyramid encoding. The techniques from the previous two sections are then used to measure how the interscale redundancy might be exploited.

Most images contain information at a number of scales. In order to deal efficiently with these features, the pyramid subimages are hierarchically created by recursive application of the same separable filters. A fundamental characteristic of such pyramids is that they capture detail uniformly at all scales. This means that detail, which occurs at different scales, will have similar representation. Specifically, the vertical and horizontal subimages are, at all scales, composed of a series of oriented gradients¹ This similarity of representation admits the potential for statistical similarity.

¹see figures 3.4 through 3.9

It is beyond the scope of this work to attempt an in depth analysis of the statistical relationship between the subbands. An example of such an approach is presented by Westerink [Westerink 1988] and in chapter eleven of Jayant and Noll [Jayant 1984]. However, in order to get some feeling for the potential range of the interscale statistical overlap, a simple test was devised.

1. The fine scale (level one) highs were VQ'ed to create a two by two table.
2. The coarse scale (level two) highs were also VQ'ed to create a separate two by two table.
3. Each codeword from the fine scale was compared against all of the codewords from the coarse scale to determine a best match.
4. The square of the matching error was used to create a histogram where the number of codewords were plotted as a function of the square error .

This test was run on the still images shown in A.1 and A.2 The codebooks were 1024 entries long for both levels and were created by combining the horizontal and vertical subbands to form an orientation independent codebook. The resulting histograms are plotted in figures 4.4 and 4.5 below. In these figures, the density function of the MSE is plotted as a function of the magnitude of the MSE. We would like to use multiscale codebooks to compress the overall number of codes needed to represent an image. One way to do this is to replace all the entries from the

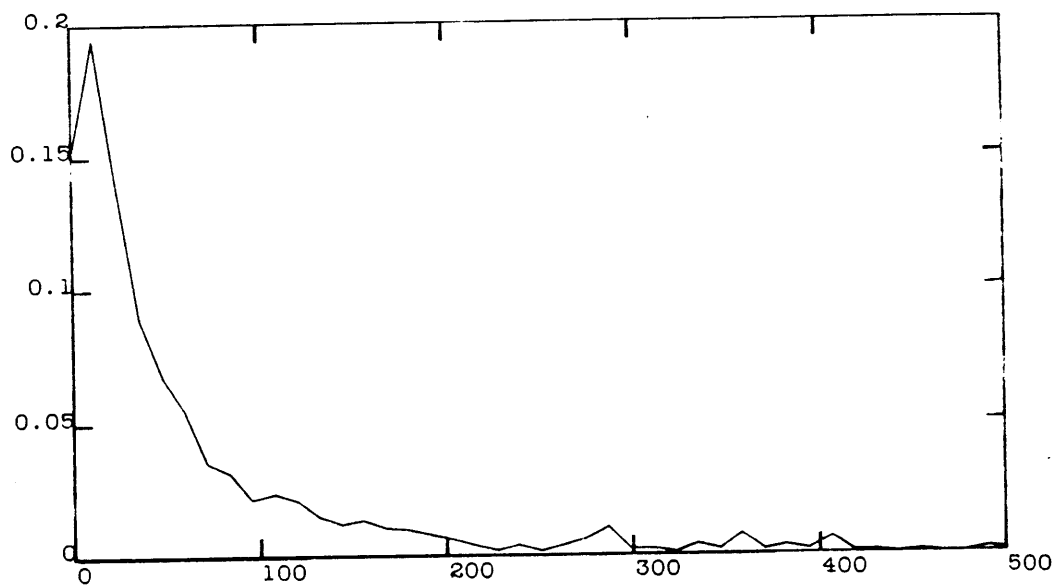


Figure 4.4: *Error Histogram for Image : Mill*

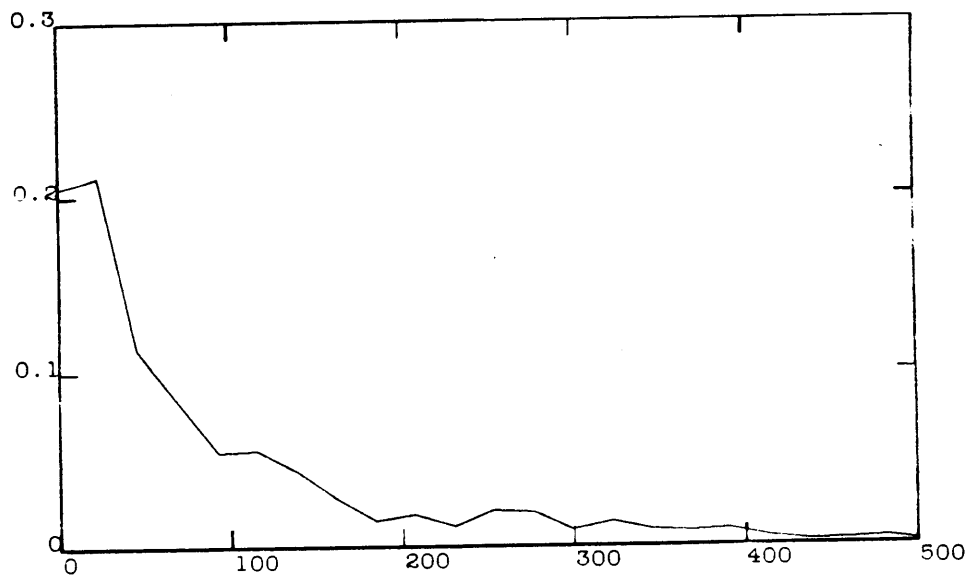


Figure 4.5: *Error Histogram for Image : Boat*

codebook at one pyramid level (target codebook) with the best matching entries from the codebook at the other pyramid level (source codebook). Such a technique confines the resulting quantization distortion to the subbands decoded from the target codebook. The distortion sensitivity hierarchy from chapter three can be used to establish that codebook from the second pyramid level should be the source codebook and that the codebook from the first pyramid level should be the target.

4.4 Temporal VQ of the Subbands

The representation of energy in the subbands is uniform in time as well as in space and scale. This implies that the same techniques described in previous sections can be extended into time. Such an extension may be straightforwardly realized by forming training sets from subimages which are concatenated over time. The issues in this case are the temporal increment and the expansion of the codebooks. In the coding examples of the next chapter, the temporal increment is arbitrarily fixed at one second.

Chapter 5

Coding Example

In this chapter, several short sequences are coded using some of the techniques previously described. The specific VQ parameters are outlined and explained. The resulting transmission bandwidth is calculated.

Three sequences have been chosen for coding. Each one consists of twenty four color images representing one second's worth of digitized film data. Each was chosen for variety of ways with which they distribute energy among the subbands.

In the first sequence (Alley), two people in the foreground are moving while the camera pans across a detailed background. In the second sequence (Leg), a woman in the foreground is walking away from the camera against a blurred, stationary background. The third sequence (Crane) shows the camera zooming and panning across a highly detailed background containing strong detail at both scales. The original resolution for all of these sequences is 240 pixels high by 320 pixels across. Frames from these three sequences are pictured in figures A.4 through A.6

Each of these sequences is coded using the scheme outlined in figure 1.1 The color components are extracted, subsampled spatially and temporally, and then transmitted. The luminance is divided into eleven subbands as illustrated in figure 3.1. The LL channel is sent explicitly with no further coding. The remaining components are grouped based on average distortion sensitivity and vector coded using either individual codebooks or orientation independent codebooks as shown below.

HH	separate 1024 entry codebook. channel coded with 2 x 2 vectors
LH & HL	grouped together to form an orientationless codebook. Each channel coded with 1024 2 x 2 vectors
HLL	separate 1024 entry codebook. channel coded with 4 x 4 vectors
LLH & LHL	grouped together to form an orientationless codebook. Each channel coded with 1024 4 x 4 vectors
HLH & HHL	grouped together to form an orientationless codebook. Each channel coded with 1024 4 x 4 vectors
LHH & HHH	not transmitted at all

The training set for each codebook is constructed by concatenating one second's worth of data for the given subband. In favor of maximizing image quality, no use is made of indirect codebook addressing or multiscale codebooks.

The bandwidth for the encoded sequences is computed by separately calculating the entropy for the channel symbols, the codebooks and the chrominance components. The entropy of the channel symbols is computed separately for each component. The entropy of the codebooks is computed for all of the codebooks combined. No entropy calculations are provided for the lows channel. Tables 5.1 through 5.3 present the bandwidth data.

Source	Channel	Dimensions	Frame Rate fps	Entropy bits/sample	Bit Rate bps
Symbols	LLH	40 x 30	12	4.17	60,048
	LHL			4.51	64,944
	HLL			4.15	59,760
	HLH			4.44	63,936
	HHL			4.27	61,488
	LH			4.18	60,192
	HL			4.47	64,368
	HH			4.45	64,080
Tables	HH	2 x 2 x 1024	1	2.52	144,507
	LH, HL	2 x 2 x 1024			
	LLH, LHL	4 x 4 x 1024			
	HLL	4 x 4 x 1024			
	HLH, HHL	4 x 4 x 1024			
Color	I	64 x 48	12	2.19	161,466
	Q				
Lows	LL	80 x 60	12	8	460,800
Total Bit Rate (bps)				1,265,589	

Table 5.1: *Bandwidth for Alley Sequence*

Source	Channel	Dimensions	Frame Rate fps	Entropy bits/sample	Bit Rate bps
Symbols	LLH	40 x 30	12	4.56	65,664
	LHL			4.45	64,080
	HLL			4.43	63,792
	HLH			4.60	66,240
	HHL			4.44	63,936
	LH			4.12	59,328
	HL			4.09	58,896
	HH			4.02	57,888
Tables	HH	2 x 2 x 1024	1	2.13	126,903
	LH, HL	2 x 2 x 1024			
	LLH, LHL	4 x 4 x 1024			
	HLL	4 x 4 x 1024			
	HLH, HHL	4 x 4 x 1024			
Color	I	64 x 48	12	2.18	160,728
	Q				
Lows	LL	80 x 60	12	8	460,800
Total Bit Rate (bps)				1,248,255	

Table 5.2: *Bandwidth for Leg Sequence*

Source	Channel	Dimensions	Frame Rate fps	Entropy bits/sample	Bit Rate bps
Symbols	LLH	40 x 30	12	4.17	60,048
	LHL			4.15	59,769
	HLL			4.27	61,488
	HLH			4.24	61,056
	HHL			4.26	61,344
	LH			4.00	57,600
	HL			4.12	59,328
	HH			3.95	56,880
Tables	HH	2 x 2 x 1024	1	2.31	132,465
	LH, HL	2 x 2 x 1024			
	LLH, LHL	4 x 4 x 1024			
	HLL	4 x 4 x 1024			
	HLH, HHL	4 x 4 x 1024			
Color	I	64 x 48	12	2.23	164,414
	Q				
Lows	LL	80 x 60	12	8	460,800
Total Bit Rate (bps)				1,235,192	

Table 5.3: *Bandwidth for Crane Sequence*

Chapter 6

Conclusions

6.1 Suggestions for Future Work

There remains much to do before the potential of this coding scheme can be realized.

Many of the parameters for the tests were fixed arbitrarily and should be examined more closely.

In the coding of color, the small number of different colors present in most images implies that the I and Q components can almost certainly be VQ'ed with no perceptible effect. A spatial VQ can be used to actually increase the chrominance resolution and still reduce the bandwidth over the current implementation. The

temporal rate of the 12 fps was a conservative choice and is completely non-adaptive

Asymmetric QMF kernels have been proposed which ease the decoding task by limiting the reconstruction to a series of shifts and adds. The hierarchy for distortion sensitivity should be tested more completely in order to fix quantitative bounds on the amount of distortion which can be introduced into each channel before the effects become visible. There is potential for great savings in the use of motion compensated predictive coding of the moving spatial lows channels (HLL in figure 3.1)

Currently, the quantized spatio-temporal components are transmitted regardless of the scene statistics. Early work on adaptively selecting components for transmission yielded poor results. This was largely due to the fact that the bandsplitting occurred on octave boundaries. A finer resolution split spatially should enable adaptive, scene-dependent selection of the components for transmission.

Finally, as better source sequences with higher temporal resolution become available, the temporal bandsplitting can also be recursively applied to build three dimensional pyramids. Use of such an extended palette of components should enable better scene-dependent adaptation.

6.2 Conclusions

A coding technique was presented here which enables low bandwidth transmission of high quality images. Source image sequences were first transformed by recursively filtering with QMF's to form a pyramid representation of the image. The subband representation was then used as a simple visual model to guide the allocation of quantization error. The individual subbands were coded such that the amount of distortion introduced into the spatio-temporal component was consistent with the HVS's sensitivity. The subband representation was shown to be well matched to the task of vector quantization.

Appendix A

Test Images

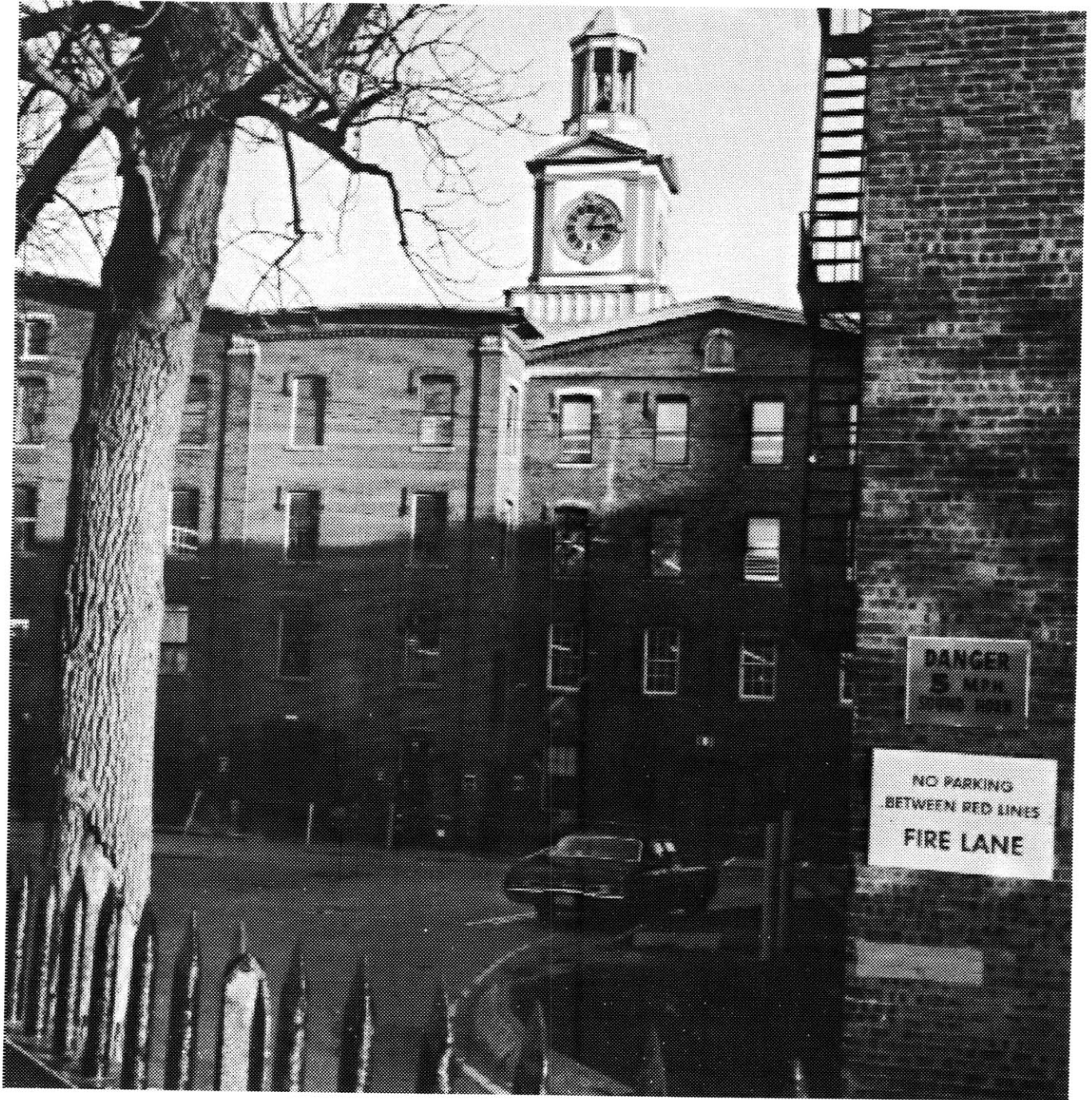


Figure A.1: *Still Image : Mill*



Figure A.2: *Still Image : Boats*



Figure A.3: *Still Image : Hat*



Figure A.4: *Test Sequence : Alley*



Figure A.5: *Test Sequence : Leg*

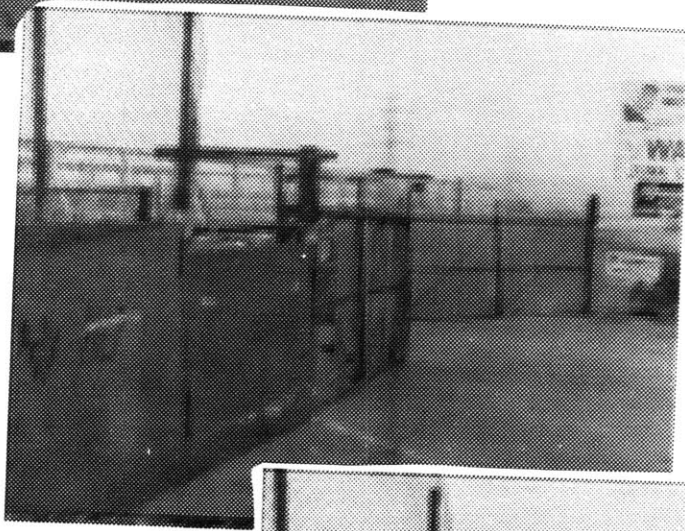
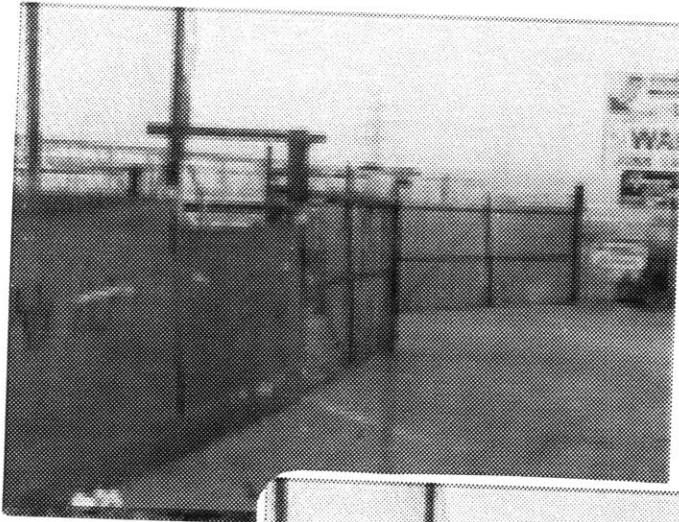


Figure A.6: *Test Sequence : Crane*

Appendix B

Acknowledgments

My heartfelt thanks to :

Professor Nicholas Negroponte and Professor Steve Benton for, respectively, founding the Media Lab and maintaining its academic standards. It has been a boundlessly exciting place to work.

To Andy Lippman for insightful discussions on VQ, for sponsoring this thesis work and for maintaining a stimulating work environment.

To Ted Adelson for suggesting the use of VQ on the subbands and for patiently instructing me in use of QMF's.

To Professor William Schreiber for thoughtful advice and frequent support.

To the members of the Advanced Television Research Program for their advice, suggestions, and generous use of their facilities and expertise.

Special thanks to Walter Bender and Eero Simoncelli for meticulously reviewing early drafts. Their thoughtful comments and corrections proved valuable and are much appreciated.

To the many members of the Garden Variety (hurrah !) and Garden Hose (booooo...) basketball teams for support, fellowship, exercise, and good memories.

To the whacked out denizens of the Terminal Garden. Working there has not always been easy, but it has always been simulating and a lot of fun. To Wad for the sanity refills, to VMB for wondering why anybody would want one, to Deirdre for boundless enthusiasm, to Pat for the DSP reviews, to Sphere for being an all around guy, to Plin and Janet for the welcome retreats back to maturity, to Foofo for the playing the CD's backwards, to JH and Mike for being first rate field marshals in the battle against demon *Unix* and *Lisp* cur, to the Chairman for his incantations, to Kenji for steadiness and Odin for his Legal(s) aid, to Cod because he could always be counted on, to Paul for philosophy made interesting, to Uri for good tunes, to Tareque for the lousy off-the-air video, to Straz for the cool competence, and to Steve and Dead for always making it worth while. Thanks are due to Pascal for always keeping the entropy up.

Finally, to my family, both present and future, for olympic servings of support and patience.

Bibliography

- [1] E. H. Adelson, C. H. Anderson, J. R. Bergen, P. J. Burt, and J. M. Ogden. Pyramid methods in image processing. *RCA Engineer*, Nov/Dec 1984.
- [2] Edward H Adelson and James R Bergen. Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America*, 2(2):284–299, February 1985.
- [3] Edward H. Adelson, Eero Simoncelli, and Rajesh Hingorani. Orthogonal pyramid transforms for image coding. *Proceedings of the SPIE*, 1987.
- [4] S. M. Anstis. The perception of apparent motion. *Phil. trans. Royal Society of London*, 150–168, 1980.
- [5] R. Aravind and Allen Gersho. Image compression based on vector quantization with finite memory. *Optical Engineering*, 26(7):570–580, July 1987.
- [6] Wen-Hsiung Chen and William K. Pratt. Scene adaptive coder. *IEEE Transactions on Communications*, COM-32(3):225–232, March 1984.
- [7] Pascal Chesnais and Wendy Plesniak. *Color Coding Stereo Pairs for Non-interlaced Displays*. Technical Report, Massachusetts Institute of Technology Media Laboratory, 1988.
- [8] Dan E. Dudgeon and Russel M. Mersereau. *Multidimensional Digital Signal Processing*. Prentice-Hall Signal Processing Series, Prentice-Hall, 1984.
- [9] William Equitz. Fast algorithms for vector quantization picture coding. *ICASSP*, 1987.

- [10] Jerome H. Friedman, Jon Louis Bentley, and Raphael Ari Finkel. An algorithm for finding best matches in logarithmic time. *ACM Transactions on Mathematical Software*, 3(3):209–226, September 1977.
- [11] H. Gharavi and A. Tabatabai. Application of quadrature mirror filtering to the coding of monochrome and color images. In *Proceedings ICASSP*, page 32.8.1 to 32.8.4, 1987.
- [12] W. E. Glenn, Karen Glenn, R. L. Dhein, and I. C. Abrahams. Compatible transmission of high definition television using bandwidth reduction. In *Proceedings of the 37th Annual Broadcast Engineering Conference*, pages 341–349, National Association of Broadcasters, 1983.
- [13] W. E. Glenn, Karen G. Glenn, Marcinka J., R. L. Dhein, and I. C. Abrahams. Reduced bandwidth requirements for compatible transmission of high definition television. In *Proceedings of the 38th Annual Broadcast Engineering Conference*, pages 297–305, National Association of Broadcasters, 1984.
- [14] Morris Goldberg and Huifang Sun. Image sequence coding using vector quantization. *IEEE transactions on Communications*, COMM-34(7):703–710, July 1986.
- [15] Robert M. Gray. Vector quantization. *IEEE ASSP Magazine*, April 1984.
- [16] Paul Heckbert. Color image quantization for frame buffer display. *Computer Graphics*, 16(3):297–307, July 1982.
- [17] Dave J. Heeger. Model for the extraction of image flow. *Journal of the Optical Society of America*, 4(8):1455–1471, August 1987.
- [18] Ellen C. Hildreth and John M. Hollerbach. *A Computational Approach to Vision and Motor Control*. A.I. Memo 864, Massachusetts Institute of Technology, August 1985.
- [19] David H. Hubel and Torsten N Wiesel. Brain mechanisms of vision. In *The Mind's Eye*, W.H. Freeman and Company, 1986. Readings from Scientific American.
- [20] N.S. Jayant and Peter Noll. *Digital Coding of Waveforms*. Prentice-Hall Signal Processing Series, Prentice-Hall, 1984.

- [21] D. H. Kelly. Motion and vision ii. stabilized spatio-temporal threshold surfaces. *Journal of the Optical Society of America*, 69(10):1340–1349, October 1979.
- [22] D. H. Kelly. Spatiotemporal variation of chromatic and achromatic contrast thresholds. *Journal of the Optical Society of America*, 73(6), June 1983.
- [23] David Marr. *Vision*. W. H. Freeman and Company, 1982.
- [24] John H. R. Maunsell and David C. Van Essen. Functional properties of neurons in middle temporal visual area of the macaque monkey. i. selectivity for stimulus direction, speed and orientation. *Journal of Neurophysiology*, 49(5):1127–1147, May 1983.
- [25] J. Anthony Movshon, Edward H. Adelson, Martin S. Gizzi, and William T. Newsome. The analysis of moving visual patterns. *Experimental Brain Research*, 1986.
- [26] Arun N. Netravali and Birendra Prasada. Adaptive quantization of picture signals using spatial masking. *Proceedings of the IEEE*, 65(4):536–548, April 1977.
- [27] Alan V. Oppenheim. *Digital Signal Processing*. Prentice-Hall, 1975.
- [28] William K. Pratt. *Digital Image Processing*. John Wiley and Sons, 1978.
- [29] V.S. Ramachandran and R. L. Gregory. Does colour provide an input to human motion perception. *Nature*, 55–56, 1978.
- [30] Bhaskar Ramamurthi and Allen Gersho. Classified vector quantization of images. *IEEE Transactions on Communications*, Comm-34(11), November 1986.
- [31] Peter H. Schiller, Barbara L. Finlay, and Susan F. Volman. Quantitative studies of single-cell properties in monkey striate cortex. i. spatiotemporal organization of receptive fields. *Journal of Neurophysiology*, 39(6):1288–1319, November 1976.
- [32] William F Schreiber. *Fundamental of Electronic Imaging Systems*. Volume 15 of *Springer Series in Information Sciences*, Springer-Verlag, 1986.
- [33] William F. Schreiber and Robert R. Buckley. A two channel picture coding system: ii-adaptive companding and color coding. *IEEE Transactions on Communications*, COM-29(12), December 1981.

- [34] William F. Schreiber and Andrew B. Lippman. *Single Channel HDTV Systems, Compatible and Noncompatible*. ATRP T-82, Massachusetts Institute of Technology, March 1988.
- [35] Eero Peter Simoncelli. *Orthogonal Sub-band Image Transforms*. Master's thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, June 1988.
- [36] D. E. Troxel, W. F. Schreiber, R. Grass, G. Hoover, and R. Sharpe. Bandwidth compression of high quality images. In *International Conference on Communications*, pages 31.9.1 – 31.9.5, June 1980.
- [37] Broder Wendland. Extended definition television with high picture quality. *SMPTE Journal*, 1028–1035, October 1983.
- [38] Peter H. Westerink, Dick E. Boekee, Jan Biemond, and John W. Woods. Sub-band coding of images using vector quantization. *IEEE transactions on Communications*, 36(6):713–719, June 1988.
- [39] Hugh R. Wilson and James R. Bergen. A four mechanism model for threshold spatial vision. *Vision Research*, 19:19–32, 1979.
- [40] John W. Woods and Sean D. O'Neil. Subband coding of images. *IEEE Transactions on Acoustic, Speech, and Signal Processing*, ASSP-34(5):1278–1288, October 1986.