Multiscale Coding of Images: Approaching the Problem from a Filmmaker's Perspective

by

Alan Lasky

Abstract

The use of the of the video coding system proposed by William J. Butera in the thesis "Multiscale Coding of Images" is explored in terms of its relevance to current and proposed work in the Film/Video group. Also, utilizing knowledge of motion picture techniques, a strategy for the possible optimization of the encoding process is explored.

Advisor: *Glorianna Davenport*

Acknowledgments:
Pasquel Romano, Alex Benenson, Che Kong Mok, Chen, Henry Holtzman, Glorianna Davenport, Uri Feldman, Andy Lippman

Introduction

In July of 1988 William J. Butera, a student in the Movies of the Future program, submitted a thesis titled "Multiscale Coding of Images" to the department of Media Arts and Sciences. In his thesis, Mr. Butera outlined a video coding system that substantially lowered the bandwidth necessary for the storage and transmission of near NTSC resolution image sequences. It is not in the scope or intent of this paper to describe the technical details of the multiscale coding system, as Mr. Butera provides a coherent and complete coverage of this topic in his work. This paper is meant to evaluate techniques that might be used to optimize the coding process, and to research ways in which the multiscale coding system might be utilized in current and proposed research in the Film/Video group. Although the paper is targeted at the laymen, some familiarity with the thesis "Multiscale Coding of Images" is encouraged.

The first section of the paper will deal with suggestions for the optimization of the encoding process. As my background is in film and video technology, and not electrical engineering, the suggestions are drawn from that knowledge base. Some of the suggestions are elaborations on conclusions made by Mr. Butera, and others are products of outside research.

The second section of the paper deals with the effects of the coding scheme on the work at Film/Video. How multiscale coding will help current work on "knowledgeable" editing systems, interactive video, and other projects is explored. Some suggestions for future research and comparisons with other available systems, such as DV-I, are made.

Optimization of Video Coding

Although the multiscale coding system is a relatively straight-forward process, it is still quite computationally expensive and time consuming. The question was raised as to how we might be able to scan a film (a series of image sequences) in order to look for ways to optimize the coding process. Also, the idea of making a global "energy map" representing the distribution of subband energy across the entire temporal range of the film needs to be researched. In order to fully explore these topics we must look more closely at both the technique of subband quantization and the nature and structure of motion pictures.

When studying the multiscale coding process the concept of codebooks becomes crucial. In the search for a simple definition of codebooks analogies are very important. Alex Benenson defines a codebook as "the color look-up table of vector quantization." Uri Feldman has another interpretation: "codebooks are the zip-codes of vectors." When coding an image (or image sequence) an arbitrary choice of block size is made, usually a 3X3 or 2X2 matrix, then these matrices are used to group the image into blocks of vectors. A tabulation of the kinds of vectors in an image is created, giving what is known as the histogram. The address log of these vectors is known as the codebook.

In chapter 4 of "Multiscale Coding of Images," Vector Quantization of Subbands, Mr. Butera "proposes four techniques for more efficiently exploiting the characteristics of the subband representation by using the VQ (vector quantization) to eliminate statistical redundancy across scale, orientation, and time." The four techniques suggested 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.

These techniques for reducing the codebook size provoked me to research the nature of film sequences and how that nature might provide ways to optimize the coding process. Mr. Butera’s fourth technique: "Comparing codebooks to eliminate redundancies between temporally neighboring images" is a technique that led me to think about the way films are structured through editing and how this structure might lead to a more global application of Mr. Butera’s idea.

In film, or more specifically in film sequences, narrative structure is built through the use of a series of individual shots. Many times similar or exact duplicate shots are used throughout a film. Consequently these similar shots would have very similar, if not matching codebooks associated with their frames and temporal regions. It appears that it would be beneficial to scan a film once, taking ordered, temporally spaced codebook samples in order to search for and eliminate codebook redundancy, and hence optimize the coding process.

Perhaps a more thorough investigation of the cinematic principles that make this "look-ahead" scheme possible is necessary. Many of the basic principles of film grammar make the use of similar or exact duplicate shots allowable and sometimes necessary. Typically sequences are blocked out in terms of master shots, medium shots, and close-ups. These shots are often intercut within a scene, and returned to throughout the entire length of the film. Parallel action and other narrative devices promote the use of duplicate shots temporally spaced over large tracts of cinematic "real estate." All of these cases make a codebook sampling "look-ahead" scheme viable for video coding optimization.

Work on a sampling "look ahead" scheme has already been proposed by a former Media Lab Student, although for a very different purpose. Russell Mayo Sasnett, in his 1986 thesis: "Reconfigurable Video," fleshed out a technique for detecting scene changes in video sequences. In section III.C. "Obtaining the Data" (pp. 64-72), he writes:

> In an attempt to derive direct visual information from movie materials by computer, a simple scene detector was created. This device continuously monitors a video channel, watching for the abrupt changes in picture content that usually signal a cut or edit. These events are noted in a master log, enabling the computer to separate video materials into semantic 'chunks' without user intervention or interpretation. For simplicity and speed, only luminance and chrominance are sampled at a small number of screen locations; statistical methods are used to infer the answer.

This technique for sampling video in order to detect scene changes captures the essence of what the "look ahead" system for video coding would do. If a film were to be scanned and every 10 seconds (300 video frames) one frame was subjected to Mr. Butera's coding system, a "codebook calendar" could be derived. The codebooks of all sampled shots could then be compared to find similar and matching codebooks from the entire global range of the film. Perhaps having pregenerated codebooks could prevent computational redundancy and in turn, optimize the coding process.

After looking at this rather specific scheme it would be beneficial to look at other methods suggested for optimization. Several other factors could provide information for the coding optimizer. Glorianna Davenport has suggested that we look to the soundtrack for this information. An analysis of the sound track could provide much in the way of knowledge about the dynamics of a scene. The history of film itself could give much information about what we could expect in terms of energy. As the film industry progressed both aesthetically and technologically the creative energy in film changed as well. In films prior to about 1960 there is a tendency toward a very regular shooting style, brought on by both the aesthetic conservatism of the times and the sheer trauma of doing anything very specialized with the equipment. If we could feed our optimizer information about the time the film was made, it could make decisions based on the amount of energy it could expect in the scenes. This is also true of genre, for we could expect a much less energetic global histogram from a Romance film than the one that would surely be generated by a Western or Action picture.

The use of Multiscale Coding in Work at Film/Video

After looking at optimization schemes for the video coding process the next logical area of research becomes the use of the process in work, both current and proposed, at the Film/Video group. Also, how might other research being done to enhance Mr. Butera's work fit in with Film/Video's work? And finally, what other systems for digital video are available and how do they rate?

In its saving of bandwidth space, the most important contribution the multiscale image coding process can make is with our work on interactive video. At the present time our interactive movies are using the analog Laser Disc format, a medium which is constrained to storage of 30 minutes of full motion video. With the multiscale coding system 72 minutes of full motion video can be stored on a compact disc, a substantial increase in the amount of information available to the interactive viewer. Many of the limitations on our current interactive work will be lifted when the scheme is put to use for motion video storage.

One of the major projects in progress at Film/Video is research in knowledge based editing. One of the main components of a knowledge based editing system is the database describing the available shots. At the lowest level the information needed by this database is the same information that an editor needs to know about a shot: what sort of shot is it? Is there camera motion? Object motion? If so, which direction? When looking at this low level shot information the question becomes could we derive any of this information from looking at the energy subbands? I believe that we could. If a subband interpreter could be made to look at all the spatio-temporal subbands and process this data into information about an image (and hence a shot), this image could then be fed to a database and low level editing decisions could be made with it. Of course much of this depends on the amount of quantization done, the efficiency of the interpreter, and our knowledge base.

Another avenue of research that could be of benefit to our interactive research is the work on "knowledgeable pictures" done by Che Kong Mok. Mok used the subband information to give the image more knowledge about the objects within it. In the demo he created, a man standing in front of a wall was separated from the background through manipulation of the quantization. The picture essentially "knew" its component parts. This sort of knowledge could help the work in interactive video in many ways. Perhaps if an image had knowledge of objects within, they could be made into clickable buttons, self reflexive databases, or simply more "knowledgeable" images. This work could also help in the knowledge based editing research by passing internal knowledge to a database.

Alex Benenson, another researcher in the Movies of the Future program, used to work at RCA on the Digital Video Interactive (DV-I) system. This system uses a series of hardware based chips to pull 60 minutes of full motion video from a compact disc. RCA’s compression scheme is very different from that of Mr. Butera, although the end result is very similar. Many of the effects brought into DV-I bode well for the future of the multi scale process.

CONCLUSION

As we have seen, much more work needs to be done in order to bring the multiscale coding scheme to its full potential. Many of the techniques described need to be tested, and more need to be developed, but it is clear that the compression of video for transmission and storage will radically change much of current media production. Experimental footage should be shot with subband coding in mind in order to test some of our theories about optimization. Looking to the future, we can dream of a perfect encoding "optimization box." If we were to have this perfect "optimization box" it would :

* Be able to sample temporally spaced frames for codebook redundancy.

* accept data about genre, time of production, and perhaps author. *scan sound tracks for clues about scene dynamics.

* accept pregenerated edit lists and "smart camera" data.

If we were able to put this information into a simple interface it could radically alter the methods for data compression, and perhaps optimize a very expensive process.