Chapter 2. Advanced Television and Signal Processing Program

Academic and Research Staff

Professor Jae S. Lim, Giampiero Sciutto

Graduate Students

John G. Apostolopoulos, Babak Ayazifar, Matthew M. Bace, David M. Baylon, Michael S. Brandstein, Shiufun Cheung, Ibrahim A. Hajjahmad, John C. Hardwick, Eddie F. Lee, Peter A. Monta, Aradhana Narula, Julien J. Nicolas, Lon E. Sunshine, Chang Dong Yoo

Technical and Support Staff

Debra L. Harring, Cindy LeBlanc

2.1 Introduction

The present television system was designed nearly 35 years ago. Since then, there have been significant developments in technology which are highly relevant to the television industry. For example, advances in the very large scale integration (VLSI) technology and signal processing theories make it feasible to incorporate frame-store memory and sophisticated signal processing capabilities in a television receiver at a reasonable cost. To exploit this new technology in developing future television systems, Japan and Europe established large laboratories which are funded by government or industry-wide consortia. The lack of this type of organization in the United States was considered detrimental to the broadcasting and equipment manufacturing industries, and in 1983 the Advanced Television Research Program (ATRP) was established at MIT by a consortium of American companies.

Currently, the consortium members include ABC, Ampex, General Instrument, Kodak, Motorola, NBC, NBC Affiliates, PBS, Tektronix, and Zenith. The major objectives of the ATRP are:

1. To develop the theoretical and empirical basis for the improvement of existing television systems, as well as the design of future television systems.
2. To educate students through television-related research and development and to motivate them to undertake careers in television-related industries.
3. To facilitate continuing education of scientists and engineers already working in the industry.
4. To establish a resource center to which problems and proposals can be brought for discussion and detailed study.
5. To transfer the technology developed from this program to the industries.

The research areas of the program include the design of a channel-compatible advanced television (ATV) system, a receiver-compatible ATV system and digital ATV system, and development of transcoding methods. Significant advances have already been made in some of these research areas. A digital ATV system has been designed and is scheduled to be tested in 1992 by the Federal Communications Commission (FCC) for its possible adoption as the U.S. HDTV standard for terrestrial broadcasting.

In addition to research on advanced television systems, the research program also includes research on speech processing. Current research topics include development of a new speech model and algorithms to enhance speech degraded by background noise. We are also investigating methods to display spectrograms more efficiently.

2.2 ATRP Facilities

The ATRP facilities are currently based on a network of eight Sun-4 workstations. There is approximately 14.0 GB of disk space distributed among the various machines. Attached to one of the Sun-4s is a VTE display system with 256 MB of RAM. This display system is capable of driving the Sun-4 monitors or a 29-inch Conrac monitor in the lab at rates up to 60 frames/sec. In addition to displaying high-resolution, real-time sequences, ATRP facilities include a Metheus frame buffer which drives a Sony 2k×2k monitor. For hard copy output, the lab uses a Kodak XL7700 thermal imaging printer which can produce 2k×2k color or black and white images on 11-inch × 11-inch photographic paper.

Other peripherals include an Exabyte 8 mm tape drive, a 16-bit digital audio interface with two channels and sampling rates up to 48 kHz per channel, and an "audio workstation" with power amplifier, speakers, CD player, tape deck, etc. Additionally, the lab has a 650 MB optical disk drive, a CD-ROM drive, and two laser printers. For preparing presentations, ATRP facilities also include a MacIntosh SE30 microcomputer, a Mac IIx, and an Apple LaserWriter.

To support the growing computation needs of the group, several additional Sun-4 workstations will be installed in the near future. They will have 24-bit color displays, local disk storage, and digital signal processing boards to assist with computation-intensive image processing. Some of the existing machines may also be supplemented with these DSP boards. In addition, the group will be purchasing image scanning equipment.

We are considering installing a fast network (FDDI) to augment the current 10 Mbps Ethernet. The new network would enable much faster data transfer to display devices, and it would support large NFS transfers more easily.

2.3 Coding of the Motion-Compensated Residual for an All-Digital HDTV System

Sponsor

Advanced Television Research Program

Project Staff

John G. Apostolopoulos

An All-Digital High-Definition Television System is being developed at MIT to transmit higher quality video and audio information in the same channel bandwidth as today's conventional television. To achieve this goal, the system must reduce redundant information which exists because of the high correlation inherent to video and audio.

For normal television broadcasts, the image from one frame to the next is very similar. In order to reduce this temporal redundancy, motion-estimation/motion-compensation is applied to the HDTV video to predict the current frame from the previous frame. The error in this prediction, also referred to as the motion-compensated residual (MC-residual), is then coded and transmitted over the channel. The focus of this research is to determine the optimal method to represent and encode the MC-residual so that the highest quality video can be produced within the limited available bit rate (approximately 0.2 to 0.35 bits/pixel).

To reduce the spatial redundancy of the MC-residual, a transform/subband filtering scheme must be applied. Toward this goal, the Block DCT, the Lapped Orthogonal Transform, and the Multi-scale schemes are examined. A number of important observations/considerations have arisen from this work:

1. Both the Block DCT and the Lapped Orthogonal Transform suffer from structured blocking artifacts that are visually degrading. The Multi-scale scheme, on the other hand, does not suffer from these artifacts.

2. The Multi-scale scheme can be implemented using a number of filterbanks, but the precise filterbank chosen is critical to its performance. For example, the original filterbank utilized for the Multi-scale scheme resulted in detracting ripple artifacts. A new filterbank eliminated these artifacts, but at the expense of reduced coding efficiency.

3. The MC-residual is more highpass in nature than a typical image, and, to take advantage of this, new frequency decompositions should be investigated.

4. Also, as the temporal prediction may be in error, a spatially adaptive inter/intra-coding scheme is found to be essential in order to achieve the highest quality reconstructed video.

This research was completed in August 1991.

2.4 Motion-Compensated Vertico-Temporal and Spatial Interpolation

Sponsor

Advanced Television Research Program

Project Staff

Babak Ayazifar

In this project, we examine the application of a motion-estimation algorithm to the field and line-rate conversion issues which exist in the process of converting video signals from European to American standards and vice-versa.

Topics explored are simultaneous temporal and vertical interpolation of image sequences and also strictly spatial interpolation of individual frames (e.g., line and column doubling) using a novel generalized form of the well known "spatio-temporal" constraint equation-based motion estimation algorithms.

This work was completed in January 1992.

2.5 Design of a Channel-Compatible HDTV System

Sponsors

Adams-Russell Electronics, Inc.
Advanced Television Research Program
National Science Foundation Fellowship
Grant MIP 87-14969

Project Staff

Matthew M. Bace, Lon E. Sunshine

The MIT-Hybrid High-Definition Television System is a video coding scheme designed to deliver high-quality video through a standard 6 MHz terrestrial broadcast channel. It is intended to be applied to a progressively scanned source which has a spatial resolution of 720×1280 pixels and a temporal resolution of 30 frames per second. The MIT-Hybrid System makes use of a hybrid modulation scheme which allows it to deliver 10 million bits and 10 million samples per second through a standard 6 MHz terrestrial broadcast channel. Some of the digital bandwidth is set aside for four channels of CD-quality audio.

The primary basis for the MIT-Hybrid System is spatial subband decomposition and block adaptive selection of important subband coefficients. Several additional measures are incorporated to combat channel noise, most notably adaptive amplitude modulation (pre-emphasis) of the selected high spatial frequency components and a hybrid analog/digital representation of important low-frequency components.

Energy compaction is achieved by spatially transforming 8×8 blocks into their subband representation. An 8-band, 16-tap LOT-type filter bank is applied in both the vertical and horizontal directions. In order to minimize the complexity and the storage requirements of the system, no temporal filtering is performed.

Subband coefficient selection is done on a block adaptive basis. The Y, I, and Q components are each handled separately. For each component, every 8×8 block (in its transform domain representation) is divided into several zones. The selection procedure starts with the computation of the average energy in each zone in every block of the frame. Then the most energetic zones in the entire frame are selected. Thus, the number of zones (and coefficients as well) selected may vary from one block to another. Some of the zones are always selected, such as the luminance (Y) DC component, and some are never selected, such as the high frequency I and Q coefficients.

Channel noise is combatted primarily through two techniques. A hybrid representation, which uses digital information to make each analog coefficient more robust in the presence of noise, is applied to the low frequency coefficients in the Y, I, and Q components. The number of bits used for a particular coefficient depends on the number of bits used for coding the zonal selection information, as well as the actual coefficient itself. The DC luminance coefficient, for example, is allocated the most bits. The high-frequency coefficients are protected from channel degradation by adaptive amplitude modulation. In this procedure, several of the selection zones are grouped together, and one digital value is used to scale all the coefficients in those zones by the same amount. Since one digital value is used for several coefficients, the overall digital requirement for adaptive modulation is less than that for the hybrid representation. Adaptive modulation, while being effective for combating noise in the high-frequency coefficients, is not suitable for the low-frequency components. Therefore, another scheme (such as the hybrid representation) must be employed.

2.6 Multirate Systems and Structures for Image and Video

Sponsor

Advanced Television Research Program

Project Staff

David M. Baylon

Recently, there has been a growing interest in scalable and extensible systems for image and video. Such systems are desirable in many applications because they can interface with a wide variety of source material and output devices without requiring significant modification to the system. For example, a high resolution input source can be displayed on either a high-resolution or lower resolution monitor with only slight modifications to the receiver.

This research focuses on determining how to optimally process image and video for such applications. One aspect under study involves signal representation. Because data rate is constrained in many applications, we are focusing on efficient representations that allow for data rate reduction without sacrificing quality of the resynthesized image and video. However, the algorithms for data rate reduction will allow for variable data rate reduction, so that resolution and quality can be

traded-off for bandwidth in a particular application. Both intraframe and interframe algorithms are being studied.

Another aspect of this research involves multirate/multistage structures. Multirate schemes allow resynthesis at different scales, so that image and video can be displayed on devices that may differ in size. We are investigating structures for implementing such schemes in a multistage fashion. Conditions under which single-stage structures can be implemented in a multistage fashion are also being studied. Particular attention is given to reconstructed image and video quality and computational complexity.

2.7 Development of a 1.5 Kbps Speech Vocoder

Sponsors

National Science Foundation Fellowship
U.S. Navy - Office of Naval Research
Grant N00014-89-J-1489

Project Staff

Michael S. Brandstein

The recently developed Multi-Band Excitation Speech Model has been shown to accurately reproduce a wide range of speech signals without many of the limitations inherent in existing speech model based systems.[1] The robustness of this model makes it particularly applicable to low bit rate, high quality speech vocoders. In Griffith and Lim,[2] a 9.6 Kbps speech coder based on this model was first described. Later work resulted in a 4.8 Kbps speech coding system.[3] Both of these systems have been shown to be capable of high quality speech reproduction in both low and high SNR conditions.

The purpose of this research is to explore methods of using the new speech model at the 1.5 Kbps rate. Results indicate that a substantial amount of redundancy exists between the model parameters. Current research is focused on exploiting these redundancies to quantize these parameters more efficiently. Attempts are also underway to simplify the existing model without significant reduction in speech quality.

This research was completed in June 1990.

2.8 A New Method for Representing Speech Spectrograms

Sponsors

National Science Foundation
Grant MIP 87-14969
U.S. Navy - Office of Naval Research
Grant N00014-89-J-1489

Project Staff

Shiufun Cheung

The spectrogram, a two-dimensional time-frequency display of a one-dimensional signal, is used extensively in speech research. Existing spectrograms are generally divided into two types, wideband spectrograms and narrowband spectrograms, according to the bandwidth of the analysis filters used to generate them. Due to the different characteristics of the two types of spectrograms, they are employed for different purposes. The wideband spectrogram is valued for its quick temporal response and is used for word boundary location and formant tracking. On the other hand, the narrowband spectrogram, with its high frequency resolution, is primarily used for measuring the pitch frequency.

Various attempts have been made to improve the spectrographic display. Past efforts include development of (1) neural spectrograms which use critical bandwidth analysis filters in imitation of the human auditory system and (2) better time-frequency distributions such as the Wigner distribution.

In this research, we propose a different approach. The spectrogram is viewed as a two-dimensional digital image instead of a transformed one-dimensional speech signal. Image processing techniques are used to create an improved spectrogram which preserves the desirable visual

---

1 D.W. Griffin and J.S. Lim, "A New Model-Based Speech Analysis/Synthesis System," *IEEE International Conference on Acoustic, Speech and Signal Processing,* Tampa, Florida, March 26-29, 1985, pp. 513-516.

2 D.W. Griffin and J.S. Lim, "A High Quality 9.6 kbps Speech Coding System," *IEEE International Conference on Acoustic, Speech and Signal Processing,* Tokyo, Japan, April 8-11, 1986.

3 J.C. Hardwick, *A 4.8 Kbps Multi-Band Excitation Speech Coder,* S.M. thesis, Dept. of Electr. Eng. and Comput. Sci., MIT, 1988.

features of the wideband and narrowband spectrograms. This transforms a speech processing problem into an image processing problem.

Among the techniques investigated are geometric-mean merge, smaller-value merge, and use of pseudocolor. In geometric-mean merge, spectrograms are combined by evaluating the geometric mean of corresponding short-time Fourier transform magnitudes. In smaller-value merge, the combined spectrogram displays only the smaller value of the corresponding pixels in wideband and narrowband spectrograms. With pseudocolor, the spectrograms are combined by mapping the wideband spectrogram to one color and the narrowband spectrogram to another color.

These techniques are found to be simple and effective. However, since a standard for objective measure in the form of an "ideal spectrogram" does not exist, evaluation of the above schemes is difficult.

This research was completed in June 1991.

2.9 Transform Coding for High-Definition Television

Sponsor

Advanced Television Research Program

Project Staff

Ibrahim A. Hajjahmad

The field of image coding is useful for many areas. Foremost of these areas is the reduction of channel bandwidth needed for image transmission systems, such as HDTV, video conferencing, and facsimile. Another area is reduction of storage requirements. One class of image coders is known as transform image coder.[4] In transform image coding, an image is transformed to another domain more suitable for coding than the spatial domain. The transform coefficients obtained are quantized and then coded. At the receiver, the coded coefficients are decoded and then inverse transformed to obtain the reconstructed image.

One transform which has shown promising results is the Discrete Cosine Transform (DCT).[5] The DCT is a real transform with two important properties that make it very useful in image coding. One is the energy compaction property, in which a large amount of energy is concentrated in a small fraction of the transform coefficients (typically low frequency components). This property allows us to code a small fraction of the transform coefficients with a small sacrifice in quality and intelligibility of the coded images. Second is the correlation reduction property. In the spatial domain there is a high correlation among image pixel intensities. The DCT reduces this correlation, and redundant information does not have to be coded.

Currently, we are investigating the use of the DCT for bandwidth compression. New adaptive techniques are also being studied for quantization and bit allocation that can further reduce the bit rate without reducing image quality and intelligibility.

2.10 A Dual Excitation Speech Model

Sponsors

U.S. Air Force - Electronic Systems Division
Contract F19628-89-K-0041
U.S. Navy - Office of Naval Research
Grant N00014-89-J-1489

Project Staff

John C. Hardwick

One class of speech analysis/synthesis systems (vocoders) which have been extensively studied and used in practice are based on an underlying model of speech. Even though traditional vocoders have been quite successful in synthesizing intelligible speech, they have not been successful in synthesizing high quality speech. The Multi-Band Excitation (MBE) speech model, introduced by Griffin, improves the quality of vocoder speech through the use of a series of frequency dependent voiced/unvoiced decisions. The MBE speech model, however, still results in a loss of quality as compared to the original speech. This degradation is caused in part by the voiced/unvoiced decision process. A large number of frequency regions contain a substantial amount of both voiced and unvoiced energy. If a region of this type is declared voiced, then a tonal or hollow quality is added to the synthesized speech. Similarly, if the region is declared unvoiced, then addi-

---

[4] J.S. Lim, *Two-Dimensional Signal and Image Processing,* (Englewood Cliffs, New Jersey: Prentice Hall, 1990); R.J. Clarke, *Transform Coding of Images,* (London: Academic Press, 1985).

[5] N. Ahmed, T. Natarajan, and K.R. Rao, "Discrete Cosine Transform," *IEEE Trans. Comput.* C-23: 90-93 (1974).

tional noise occurs in the synthesized speech. As the signal-to-noise ratio decreases, classification of speech as either voiced or unvoiced becomes more difficult, and, consequently, degradation is increased.

A new speech model has been proposed in response to the aforementioned problems. This model is referred to as the Dual Excitation (DE) speech model due to its dual excitation and filter structure. The DE speech model is a generalization of most previously developed speech models, and, with the proper selection of the model parameters, it reduces to either the MBE speech model or to a variety of more traditional speech models.

We are currently examining use of this speech model for speech enhancement, time scale modification, and bandwidth compression. Additional areas of study include further refinements of the model and improvements of the estimation algorithms.

2.11 Design of an HDTV Display System

Sponsor

Advanced Television Research Program

Project Staff

Eddie F. Lee

Several years ago, a video filter and display unit were built by graduate students to aid in the development of an HDTV system. This unit could read large amounts of digital video data from memory, filter the data, and display the data at a high rate on a large, high-resolution monitor. Unfortunately, this display system is very complex and highly unreliable; and there is very little documentation to help diagnose any problems with the system.

This project involves the design of a new, simpler, and more reliable display system. Since little documentation is available, the older system is being analyzed to determine input specifications.

2.12 Signal Processing for Advanced Television Systems

Sponsor

Advanced Television Research Program

Project Staff

Peter A. Monta

Digital signal processing will play a large role in future advanced television systems. Major applications are: (1) source coding to reduce the channel capacity necessary to transmit a television signal and (2) display processing such as spatial and temporal interpolation. Present-day television standards will also benefit significantly from signal processing designed to remove transmission and display artifacts. This thesis will focus on algorithms and signal models designed to enhance current standards (both compatibly and with some degree of cooperative processing at both transmitter and receiver) and improve proposed HDTV systems.

The American television standard, NTSC, could be improved in a number of ways with a receiver with a high-quality display and significant computation and memory. Interlace artifacts, such as line visibility and flicker, can be removed by converting the signal to a progressive format prior to display. Color cross-effects can be greatly reduced with accurate color demodulators implemented with digital signal processing. If the original source material is film, an advanced receiver can recover a much improved image by exploiting structure in the film-NTSC transcoding process; such an algorithm has been implemented and tested.

Similar ideas apply to HDTV systems. For example, film will be a major source material well into the next century, and HDTV source coders should recognize film as a special case, trading off inherent reduced temporal bandwidth for better spatial resolution.

2.13 Relative Importance of Encoded Data Types in an All-Digital HDTV System

Sponsor

Advanced Television Research Program

Project Staff

Aradhana Narula

The Federal Communications Commission (FCC) has regulated that transmission of HDTV signals must be restricted to occupy only the same 6 MHz channel bandwidth that has been allocated for the current television system, NTSC signals. In order to fulfill this requirement and still provide the high quality images that HDTV is designed for, digital image processing techniques must be applied. An efficient coding technique is needed to compress the enormous amount of information into as few bits as possible while maintaining a high degree of

picture resolution and quality. The amount of information which must be transmitted is compressed by taking advantage of the inherent redundancy both within a single image and between subsequent picture frames, taking into account the capabilities and limitations of the human visual system.

Video compression techniques essentially form a new representation of image sequences. In the original representation, each data value corresponds to only one pixel. The data in the new representation, however, affects blocks of pixels, the size of the block varying with different types of data. For example, control information includes frame synchronization data, type of video signal, and level of quantization; in general, information that affects the entire frame. Other values of data may only affect blocks of 32×32 pixels or 8×8 size blocks. As a result of using an efficient coding scheme, the effect of transmission errors becomes more serious. A bit error in the original representation can only affect one pixel, but, in the compressed representation, the error could destroy blocks of pixels or even the entire frame. To reduce the effect of the channel errors, it is necessary to add redundancy back into the data through the use of channel coding. Some bits are more vital to the image than others, and thus it may be useful to protect the more important bits to a greater extent. In this research, the relative importance of the different types of data in video compressed representation will be determined.

2.14 Transmission of HDTV Signals in a Terrestrial Broadcast Environment

Sponsor

Advanced Television Research Program

Project Staff

Julien J. Nicolas

High-Definition Television Systems currently being developed for broadcast applications require 15-20 Mbps to yield good quality images for approximately twice the horizontal and vertical resolutions of the current NTSC standard. Efficient transmission techniques must be found in order to deliver this signal to a maximum number of receivers while respecting the limitations stipulated by the FCC for over-the-air transmission. This research focuses on the principles that should guide the design of these transmission systems.

The major constraints related to transmission of broadcast HDTV include (1) a bandwidth limitation (6 MHz, identical to NTSC); (2) a requirement for simultaneous transmission of both NTSC and HDTV signals on two different channels (Simulcast approach); and (3) a tight control of the interference effects between NTSC and HDTV, particularly when the signals are sharing the same frequency bands. Other considerations include complexity and cost issues of the receivers, degradation of the signal as a function of range, etc.

A number of ideas are currently under study. Most systems proposed to date use some form of forward error-correction in order to combat channel noise and interference from other signals. The overhead data reserved for the error-correction schemes represent up to 30 percent of the total data, and it is therefore worthwhile trying to optimize these schemes. Current work is focusing on the use of combined modulation/coding schemes capable of exploiting specific features of the broadcast channel and the interference signals. Other areas of interest include the use of combined source/channel coding schemes for HDTV applications and multiresolution coded modulation schemes.

2.15 Hybrid Analog/Digital Representation of Analog Signals

Sponsor

Advanced Television Research Program

Project Staff

Lon E. Sunshine

Transform coding has been shown to be an effective way to represent images, allowing for significant amount of data compression while still enabling high-quality reproduction of the original picture. One result of transform coding of images is that, at a given signal-to-noise ratio (spatially), low-frequency components are much more sensitive to additive noise that high-frequency components.

In the MIT-CC television system, we must employ a noise reduction technique that is able to eliminate the effect of additive noise at low frequencies. We represent these analog (continuous-amplitude) coefficients by a hybrid analog/digital signal. This representation consists of a new analog value plus a discrete-valued piece of side information. The advantage of using this hybrid format is that we can reduce the noise added to a particular coefficient by at least 6 dB for each bit used in the side information.

In our research we consider the task of determining the "best" hybrid representation for an image. Here, "best" is characterized by a tradeoff between sufficient noise reduction, simplicity in implementation, and minimization of necessary side information.

This work was completed in February 1992.

2.16 An Iterative Method for Designing Separable Wiener Filter

Sponsor

Advanced Television Research Program

Project Staff

Chang Dong Yoo

One of the most common problems in signal processing is noise cancellation or reduction. If the statistics of the undegraded signal and the noise are given, the optimal solution that minimizes the mean square error between recovered signal and undegraded original signal is the Wiener filter. In this work, we impose an additional constraint of separability on the two-dimensional filter design problem. The motivation for developing these algorithms is clearly computational efficiency. This design problem is not the same as trying to fit a separable filter to an ideal two-dimensional filter. It has been shown that the solution to this problem can be obtained by computing the SVD of the ideal two-dimensional filter. In this work, a method is being studied to go directly from the statistics of the signals to the coefficient of the separable filter without the intermediate step of computing the ideal unconstrained two-dimensional filter.

A set of nonlinear equations in the design parameters is derived and an iterative algorithm to solve them is presented. Application of this filter towards images is being explored.