Advanced Telecommunications and Signal Processing Group

Academic and Research Staff
Professor Jae S. Lim

Research Affiliate
Carlos Kennedy

Visiting Scientist
Professor Chang Dong Yoo

Graduate Students
Brian Heng
Fatih Kamisli

Support Staff
Cindy LeBlanc

Introduction

The present television system was designed nearly 50 years ago. Since then, there have been significant developments in technology, which are highly relevant to the television industries. 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, the research areas of the program focused on a number of issues related to digital television design. As a result of this effort, significant advances have already been made and these advances have been included in the U.S. digital television standard. Specifically, the ATSP group represented MIT in MIT's participation in the Grand Alliance, which consisted of MIT, AT&T, Zenith Electronics Corporation, General Instrument Corporation, David Sarnoff Research Center, Philips Laboratories, and Thomson Consumer Electronics. The Grand Alliance digital television system served as the basis for the U.S. Digital Television (DTV) standard, which was formally adopted by the U.S. Federal Communications Commission in December 1996. The standard imposes substantial constraints on the way the digital television signal is transmitted and received. The standard also leaves considerable room for future improvements through technological advances. Current research focuses on making these future improvements.

In addition to research on issues related to the design of digital television system, the research program also includes research on signal processing for telecommunications applications.

1. Adaptive Multiple Description Mode Selection for Error Resilient Video Communications

Sponsor: Advanced Telecommunications Research Program

Project Staff: Brian Heng

The transmission of video information over error prone channels poses a number of interesting challenges. One would like to compress video data as much as possible in order to transmit it in a timely manner and/or store it within a limited amount of space, yet compressing a video sequence tends to make it more susceptible to errors and transmission losses. Streaming video applications must be able to withstand the potentially harsh conditions present on best-effort networks like the Internet, including variations in available bandwidth, packet losses, and delay. Those which are unable to adapt to these conditions can suffer severe performance degradations each time the network becomes congested.

Multiple description (MD) video coding is one approach that can be used to reduce the detrimental effects caused by packet loss on best-effort networks. In a multiple description system, a video sequence is coded into two or more complementary streams in such a way that each stream is independently decodable. The quality of the received video improves with each received description, but the loss of any one of these descriptions does not cause complete failure. If one of the streams is lost or delivered late, the video playback can continue with only a slight reduction in overall quality.

There have been a number of proposals for MD video coding, each providing their own tradeoff between compression efficiency and error resilience. Previous MD coding approaches applied a single MD technique to an entire sequence. However, the optimal MD coding method will depend on many factors including the amount of motion in the scene, the amount of spatial detail, desired bit rates, error recovery capabilities of each technique, current network conditions, etc. This research examines the adaptive use of multiple MD coding modes within a single sequence. Specifically, this work proposes an adaptive MD coder which selects among MD coding modes in an end-to-end rate-distortion (R-D) optimized manner as a function of local video characteristics and network conditions. This approach makes optimized decisions using a model of expected end-to-end distortion allowing the encoder to minimize the expected end-to-end distortion of the system. Simulation results have shown how one such system based on H.264 is able to adapt to local characteristics of the video and to network conditions on multiple paths and have shown the potential for this adaptive approach to significantly improve video quality. These simulations demonstrate how this system accounts for the characteristics of the video source, e.g. using more redundant modes in regions particularly susceptible to losses, and how it adapts to conditions on the network, e.g. switching from more efficient methods to more resilient methods as the loss rate increases. The results with this approach appear quite promising, and we believe that adaptive MD mode selection can be a useful tool for reliable delivery of video streams over lossy packet networks.

2. Reduction of Blocking Artifacts

Sponsor: Advanced Telecommunications Research Program, Higher Education Council of Turkey

Project Staff: Fatih Kamisli

Block-based video compression methods are used extensively in practice. For example, digital television standards in the United States utilize block-based video compression methods. When used in low bit-rate applications, however, they suffer from annoying artifacts known as blocking artifacts. A number of different approaches were proposed in the past to reduce the blocking artifacts. In this research, we have reviewed the advantages and disadvantages of these existing methods. We have developed a new approach where blocking artifacts are reduced using side information transmitted from the encoder to the decoder. Using side information enables the use of the original image in deblocking, which improves performance. A key question that arises is if the gain in performance can compensate for the increase in the bit rate due to transmission of side information. Preliminary results indicate that this approach is useful for some range of coding bit rates.

3. Video Fingerprinting based on Centroids of Gradient Orientations

Sponsor: Advanced Telecommunications Research Program

Project Staff: Professor Chang Dong Yoo

A novel video fingerprinting method based on centroids of gradient orientations is proposed. A fingerprint is a short feature vector that can uniquely characterize a video content. In order to fulfill its intended purpose, it needs to be robust against various distortions, it must be pair-wise independent, and it must be conducive for conducting database (DB) search.

The goal of a video fingerprinting system is to judge whether two videos have the same contents by measuring the distance between the fingerprints extracted from them. A video fingerprinting system consists of three essential parts: fingerprint extraction, the DB search, and the fingerprint matching. First,

a fingerprint is extracted from an input video. Then, the fingerprints similar to the extracted fingerprints are searched in the DB. Finally, the searched fingerprints are matched with the extracted fingerprints and the fingerprint with minimum distance is chosen as the fingerprinting result. Among three parts, this research focuses on the fingerprint extraction and matching. The promising applications of the video fingerprinting are the filtering for file-sharing service, broadcast monitoring, connected multimedia, and automated indexing of large-scale video archives.

The fingerprint is extracted as follows. First, an input video is re-sampled at 10 frames per second. This makes the fingerprint robust against frame-rate change. Next, re-sampled frames are converted to grayscale to cope with variation in color. Then, the frames are resized to a certain dimension. This makes the fingerprint robust against the resizing. Finally, frames are partitioned into 8 blocks and the centroids of gradient orientations (CGO) are calculated for each block. As a result, an 8-dimensional fingerprint is obtained for each frame.

The gradient orientations and its histogram have been widely used as image features. However, their high dimensionality has rendered them unsuitable for the video fingerprinting. In this work, we propose the CGO as a video fingerprint. For each pixel, the gradient magnitude and orientation are calculated. The direction of the gradient is the orientation in which the directional derivative has the largest value. Then, the CGO of the block is calculated. The value of the centroids ranges from -0.5 PI to 0.5 PI.

An 8-dimensional fingerprint from a single frame was considered too short for reliable matching. For this reason, fingerprints from video of 10 seconds are concatenated to make an 800-dimensional fingerprint sequence, and this is used in the matching. This restricts the minimum length of a query video to 10 seconds. In the fingerprint matching, a certain threshold is defined, and two videos are declared similar if the distance between the fingerprint sequences from them is below a certain threshold. Otherwise, two videos are declared dissimilar.

To determine the threshold, the false alarm rate and the false rejection rate are considered. But, in practice, it is common to deal with only false alarm rate for choosing the threshold, since false rejection rate is difficult to analyze.

To analyze the false alarm rate, the proposed fingerprint is assumed as a realization of a stationary process. A fingerprint sequence is further normalized by its mean and variance. The stochastic model of the proposed fingerprint is simplified as the first-order auto-correlation, and the model parameters are estimated from actual video data.

As the distance measure, squared Euclidean distance is used. Then, by using the first-order model described above, the mean and standard deviation of the distance was determined. Assuming that the length of the fingerprint sequence is sufficiently long, the distance is assumed to be Gaussian distributed following the central limit theorem. Then, based on this approximation, the probability of false alarm can be obtained. Now we can choose the threshold given a certain value of the false alarm rate.

For the performance evaluation, a fingerprint DB generated from 60 videos belonging to various genres was generated. The resolution ranges from 320x240 to 720x400, and the frame rate is 29.97 Hz. To evaluate the pair-wise independence of the proposed fingerprint, 101,768 pairs of fingerprint sequences are randomly selected from the database, and the distance between the sequences in each pair is calculated. The mean and standard deviation of the measured distance were 1.9663 and 0.2936. They were very close to the theoretically derived mean and standard deviation. The theoretically derived distribution of the distance and the histogram of the measured distance were also very similar. This result shows that the proposed fingerprint follows the stochastic model reasonably well and is pair-wise independent. To evaluate the robustness of the proposed fingerprint, the video clips in the test database are subjected to various video processing steps. In the experiments, the measured distance between the original and distorted fingerprints was below the set threshold for all distortions except the histogram equalization. This result is not surprising since the histogram equalization sometimes severely degrades the perceptual quality of the video, and the falsely rejected fingerprints were those extracted from the severely distorted portion of the video. The robustness of the proposed fingerprint is also compared with

that of the luminance histogram, block mean luminance, and difference of block mean luminance. The experimental results show that the proposed fingerprint outperforms other features.

In conclusion, the experimental results show that the proposed fingerprint is not only pair-wise independent but also robust to common video processing steps. And the proposed fingerprint outperforms other features in the context of video fingerprinting.

Publications

Journal Articles, Submitted for Publication

Brian A. Heng, John G. Apostolopoulos, and Jae S. Lim, "End-to-End Rate-Distortion Optimized MD Mode Selection for Multiple Description Video Coding," EURASIP Journal on Applied Signal Processing **2006** (2006), Article ID 32592, 12 pages.

Patents

Lim, J.S., "Method and Apparatus for Encoding Decoding and Compression of Audio-Type Data Using Reference Coefficients Located within a Band of Coefficients," Canadian Patent No.2,128,216 issued October, 2005.

Theses

Heng, Brian A., "Adaptive Multiple Description Mode Selection for Error Resilient Vide Communications," Ph.D. Thesis, Department of Electrical Engineering and Computer Science, June, 2005.

Kamisli, Fatih, "Reduction of Blocking Artifacts Using Side Information," MS Thesis, Department of Electrical Engineering and Computer Science, February, 2006.