An Autosophy Image Content-Based Television System

Klaus Holtz and Eric Holtz
Autosophy
San Francisco, California

Abstract

A first television system based on the Autosophy information theory is now being tested. The new television marks a major theoretical break from conventional television based on the Shannon information theory. In conventional television bit rates are determined entirely by hardware parameters, such as screen size, resolution, and scanning rates. The images shown on the screen are irrelevant, such that random noise video requires the same bit rate as any other video, whether blank or rich in content. In the new Autosophy-based television, in contrast, bit rates are determined entirely by video content, essentially motion and complexity within the images. It is the imaging hardware that becomes virtually irrelevant. A very high degree of visually lossless video compression is possible because only moving parts of the video are transmitted. Transmitted codes represent multi-pixel image clusters in a pre-grown hyperspace library. The system can dynamically and seamlessly reduce resolution of fast-moving objects when necessary to accommodate bandwidth restrictions. Ideally suited to the Internet environment, the new television features high resistance to delayed or dropped packets, a universal hardware-independent communication format, and optional "codebook" encryption for secure communications.

Introduction

Modern digital television is designed according to the Shannon information theory[1] published in 1948, in which the bit rate is determined entirely by the imaging hardware, i.e., screen size, resolution, and scanning rates. The image content is irrelevant. A random noise transmission requires the same bit rate as a blank screen transmission, an action sequence, or a static surveillance scene. Any attempt to reduce the bit rate results in inevitable image distortions or loss of resolution. Most modern video compression methods can do little more than try to hide those distortions from the human observer.

A breakthrough came in 1974 when Klaus Holtz formulated a new information theory[2] in which bit rates are determined by data content. Later, in 1978, Jacob Ziv and Abraham Lempel[3] adopted one of the basic Holtz tree algorithms for use in data compression. In 1984 Terry Welch[4] modified the algorithm using a dynamic library. The resulting V.42bis data compression standard,[5] introduced by British Telecom, is now implemented world-wide in virtually all modems. Lossless still image compression methods using similar methods are the GIF and TIF formats, both also widely used on the Internet. Despite such widespread adoption there is a general lack of understanding of the technology's true basis and potential.

Meanwhile the need to explore that potential has become urgent. The Internet has brought a packet-switching network into most homes and offices, but the range of content it can deliver is severely constrained by bandwidth limitations. The transmission of large screen video at an acceptable image quality requires excessively high bit rates. Applications like Internet teleconferencing and movies-on-demand remain out of reach. In 1991, for example, the FCC tried to introduce a new broadcast High Definition Television standard, but high cost and poor image quality has slowed public adoption.

In 1995 Klaus Holtz proposed a new type of television[6,8,9] based on his Autosophy information theory.[7] Bit rates in the new television are determined only by video content, essentially motion and complexity within the images. The imaging hardware, i.e. image size, resolution and scanning rates, become largely irrelevant. In addition to very high compression ratios, the new television may also solve many of the other problems associated with packet-switching video. The new Autosophy television system is now being built and tested in both software simulation and hardware. Preliminary test results are very promising.

An Autosophy Television System

Basically, Autosophy defines "information" as data which is not already known and which may therefore create new knowledge in the receiver. This is done through "communication" using data packets, called "Tips," each producing an "engram" of knowledge as a minimal extension to what is already known. The result is that Autosophy communicates only motion and complexity within the images, while the imaging hardware parameters become virtually irrelevant. Everything already known by the receiver need not be re-transmitted.

An Autosophy television system contains a matching image buffer in both the transmitter and the receiver, which holds the current image frame. A new input image frame, from the television camera, is compared with the current image frame to detect the pixels whose brightness has changed. The new pixel brightness values of the changing pixels are stored into the image buffer. The screen location addresses of the changing pixels are accumulated in a change buffer. The encoding process combines the changed pixels into spiral clusters using a fixed hyperspace knowledge library. The output is a universal 64 bit code packet, which defines a group of changing pixels in a spiral cluster which can be anywhere within the image frame. The video code packets are randomly mixed with other data packets (representing sound, text, or random bit files) for storage or transmission. The receiver retrieves the spiral image cluster from the code packets using a duplicate fixed hyperspace library. The changing pixel clusters are used to update small areas in the output image frames.

In this way, only the changing portions of the input video are transmitted. The hyperspace library combines the changing screen clusters into spiral patterns for improved compression. Transmission bit rates are determined by the amount of motion or change from one image frame to the next. Static portions of the video are not re-transmitted. A totally static video image requires no transmission at all.

Video Compression Strategies

The new Autosophy television system uses several strategies to remove redundant "information." The compression is essentially lossless until the available channel bandwidth is exceeded. After that, the system switches to visually lossless compression. Unlike other video compression schemes, like JPEG and MPEG-2, that will not lead to visible image distortions. Visually lossless compression strategies include:

--- Removal of noise jitter before encoding. The least significant bits of noise jitter from the camera are not visible to the human observer and can be removed without visible image distortions. A dynamically variable threshold is used when comparing the new input frame with the previous image frame and any change in pixel brightness less than the threshold is ignored.

--- Transmitting only changing portions of the images. Since both the transmitter and the receiver have a matching image buffer, there is no need to re-transmit the static portions of the image frames. Transmission bit rates are directly proportional with the amount of movement within the video. Random noise video would require excessive bit rates, while slow-moving video requires very few transmissions.

--- Using a static hyperspace library of the most often used imaging patterns so as to combine changing pixels into larger clusters. The library is similar to the one used in the V.42bis modem standard, but grown prior to video transmission. Using normal input images as input, a Holtz serial tree network is grown in which the most often encountered image patterns migrate towards the top, while less often used imaging patterns are pushed towards the bottom and eventually dropped entirely from the tree network. The new television contains a matching hyperspace tree library in both the transmitter and receiver. Unlike the V.42bis modem standard, once grown the library will not change during transmission. The hyperspace library contains millions of different imaging patterns in seven levels of resolution. A generic library allows open communications, while specially grown libraries can provide unbreakable "codebook" encryption for secure video communications.

--- Using spiral scanning patterns to encode changing image portions. Instead of scanning the video images from left to right and top to bottom as in conventional television, the changing image portions are scanned in a tight spiral pattern around a pixel that has changed its brightness. Spiral scanning is more efficient because most images are composed of large areas of similar brightness and color.

--- Reducing the resolution of very rapidly moving objects in the video. This backup method is used only as a defense against extremely rapid movement when the transmission bit rate is limited. The human eye and brain can perceive extremely fine color resolution or rapid movement, but not both at the same time. The resolution of very rapidly moving objects in the video may be reduced temporarily without visible artifacts. Once the rapid movement subsides, then normal fine resolution is restored. Such backup strategies are necessary to defend against rapid bursts of packet transmission in cases of severe flashing, scene cuts, rapid panning of the camera, or transmission of random noise.

Features for Internet Video Transmission

In addition to very high video compression ratios, the new Autosophy television has features that make it ideally suited to the packet-switching Internet environment. The features include:

--- A universal hardware-independent communication protocol. In Autosophy communications the video hardware parameters (such as image size, resolution or scanning rates) are virtually irrelevant. A universal 8 byte packet protocol was developed to transmit any type of data, including sound, text, or any other random bit pattern. All these data types can be randomly mixed together into standard Internet TCP/IP packets for transmission. The 8 byte codes may also be stored in a storage medium for play-back at a later time. Simultaneous mixed data transmission is necessary for Internet video because there are times when video is mostly static with continuous narration, while at other times video may change rapidly with less sound. This allows video transmission in a hardware-independent format. The transmitter and receiver may both have entirely different image formats, image sizes, color resolution, or scanning rates and yet remain compatible. This allows television technology to evolve towards larger and larger screens and higher resolution, while always remaining forwards and backwards compatible.

--- High resistance to transmission errors. In Internet communications data packets are often delayed or out of sequence. Data packets may arrive with transmission errors, or packets may be dropped in a congested network.

Re-transmitting defective packets in the TCP/IP protocol is not possible in real-time streaming video or HDTV broadcast. In most streaming video, including the MPEG-2 standard, a single transmission error or dropped packet produces very disturbing visual effects. Autosophy television, in contrast, only transmits motion within the video. Defective or lost data packets may cause tiny spots on the screen to freeze when they were supposed to change. The resulting image distortions will be barely visible even in very noisy transmission channels.

--- Built-in "codebook" encryption for secure video communications. The privacy of communications on the public Internet is becoming a major problem. Privacy and intellectual property rights must be protected. By growing special encryption libraries the new Autosophy television can provide truly unbreakable "codebook" encryption for secure communications. Without a matching hyperspace library video transmissions or video recordings cannot be retrieved.

System Setup

The testing procedure used a set of 8 input images, known as the Waterloo Repertoire ColorSet, downloaded from http://links.uwaterloo.ca/colorset.base.html.

Clegg	Artistic image of yellow flowers	814x880x24
Frymire	Artistic image of a frog	1118x1105x24
Lena	Photograph of a woman	512x512x24
Monarch	Photograph of a butterfly	768x512x24
Peppers	Photograph of bell peppers	512x512x24
Sail	Photograph of windsurfers	768x512x24
Serrano	Artistic image of a woman	629x794x24
Tulips	Photograph of yellow flowers	768x512x24

A first task was to grow a hyperspace image pattern library from the 8 input images. Each image was divided into 4 by 4 pixel clusters, which were then converted into string codes for input to the learning tree library. Each cluster was learned several times at a different resolution, ranging from 8 bits-per-color to only 2 bits-per-color. The algorithm builds a tree network similar to the Holtz serial network[9] implemented in the V.42bis modem standard. The algorithm climbs the tree nodes until no further node extensions can be found. A new node is then appended to the tree network. Whenever a node is found in the tree, the node is swapped with the node stored one location higher in the node list. The result is that the most often used image patterns tend to migrate towards the top of the list, while less common image patterns are pushed down and eventually dropped from the list. The longer the process is allowed to run the more efficient the library will become. After an unattended overnight run in the computer, the network nodes are retrieved to a file and used in the generation of the final output hyperspace library. In software-only applications a Sussenguth tree[10] is added for faster library searching.

The library is grown to be resolution-independent. Very high resolution of 8 bits-per-color is used normally only for transmitting high resolution still images. A resolution of 5 bits-per-color is equivalent in quality to normal NTSC-type television. Lower resolution is used to temporarily reduce the resolution of very fast moving video.

A resolution of 2 bits-per-color is used only to prevent jamming of packet switching networks by totally random noise video.

Since the library does not change during video transmissions, a generic library may be used to transmit any video content in any application. This library would be part of the video encoder-retrieval software or hardware. In software-only applications the Sussenguth tree is used (as in the V.42bis standard) to speed-up video encoding to only a few second per frame depending on the image size. Software-only encoding will not be fast enough for real-time video encoding. However, software-only image retrieval in a fast PC can keep up with real-time speed. In a real-time hardware encoder system the library is stored in a Content Addressable Memory (CAM). A hardware video retrieval system only requires a library stored in an inexpensive Read Only Memory (ROM) chip.

For secure video transmissions over the Internet a private encryption library is grown prior to the video transmissions and distributed through keyword-encrypted Internet uploading. This may provide truly unbreakable "codebook" encryption for secure private networks. Without a correct copy of the transmitter's library the retrieval of intercepted video is virtually impossible.

Each video encoder system requires a "quality profile," which must be set up in advance before any communication. The profile is stored in a file identified by name, such as Internet Video Profile, Teleconferencing Profile, or HDTV Broadcast Profile. The file contains information about the resolution of the camera and the available channel bit rate.

--- The image resolution, in bits per color, must be specified to the encoder to avoid excessive and unnecessary transmissions. This would avoid a situation, for example, in which a normal NTSC-type camera (with 5 bits-per-color resolution) is used with a system set up to encode 8 bits-per-color. A very high bit rate would result because the system would effectively only be transmitting the lower 3 bit random noise from the camera. A threshold value must be therefore specified for each camera or input medium to reduce unnecessary transmissions due to random noise in the video.

--- The quality profile must contain information of how to deal with excessive movement and complexity in the video. While a new image frame is scanned from the camera, it is constantly compared with the previous image in the image buffer to determine the amount of movement in the video. As long as the amount of movement is within the transmission bit rate of the channel, then the compression is lossless and high resolution is maintained. If very fast motion in the video is detected (from scene cuts, flashing, or rapid panning of the camera), then the image resolution must be reduced until the rapid motion subsides. How to deal with such rapid motion depends on the bit rate of the transmission channel or the available storage capacity. Low bit rate channels (such as the Internet) or low capacity storage media require more aggressive strategies than higher bit rate channels. The quality profile, however, must ultimately be determined by the human observer. Very rapidly moving video need not be reproduced at high resolution.

Whatever reductions in resolution are not visible to the human observer can be tolerated.

--- The quality profile may also specify the frame rate to avoid motion artifacts. For very rapidly moving objects in the video the frame rate may be dynamically increased to reduce flicker or perceived motion jumping.

While the encoder must specify its quality profile the receiver does not require such information. That would make future video transmissions or recordings entirely independent of hardware parameters, such as image size, resolution or frame rates. Both the transmitter and the receiver may have entirely different screen sizes, color resolutions, or frame rates and yet communicate in a universal format, which would always remain forwards and backwards compatible. That would allow future television to evolve into larger screens with better resolution without any need to define new standards.

High-resolution still images can be transmitted through the Internet using the same hyperspace library. Special encryption libraries would make such transmissions secure. A more efficient 5 byte-per-cluster code can be transmitted, like a random bit file, to improve the compression ratios of still image transmissions.

Conclusion

While much work remains to be done, preliminary test results are very encouraging. This may become the key technology for enabling transmission of live television in packet-switching networks like ATM, the Internet, or the future Information Superhighway. It may also improve the transmission of a future High Definition Television system.

At this point in time, compression ratios are the most important consideration for the transmission of live video over an Internet DSL connection. In future, though, other considerations like error resistance, universally compatible video formats, and encryption capability may become much more important. There is no need to resort to lossy video compression, like JPEG or MPEG-2, with its unacceptably poor image quality and motion distortions.

The actual video compression ratio calculation depends on the video content (motion and complexity); the size and resolution of the input video (rows x columns x bits/pixel x frame rate); the average number of pixels per cluster; and the overhead control bits in the packet protocol. Basically, the compression ratio is calculated as "hardware" (Shannon bit rate) divided by "content" (Autosophy bit rate). Because there is no fixed relationship between the TV monitor resolution and the images actually shown on the screen, there cannot be any absolute calculation of compression ratios. Only long-term averages for whole movies, for example, are possible.

There is virtually no downside. While a 2.6:1 data expansion ratio is theoretically possible for totally random noise video, any data expansion is extremely unlikely under real operating conditions. For real-time teleconferencing over the Internet or surveillance applications, compression ratios may be many thousands to one.

For real-time television transmission via the Internet, compression ratios depend on the video content. Under extreme conditions (rapidly flashing video, frequent scene cuts, or rapid panning of the camera) compression ratios must be increased by temporarily reducing image resolution until the conditions subside. This will not result in visible video distortions because the human eye cannot perceive fine resolution in very rapidly moving video. The system performance can also be greatly improved by using higher quality cameras with less noise jitter.

The new Autosophy television may have the same impact on video communications as the V.42bis standard did to modem communications. Inexpensive chipsets may, for the first time, allow for high quality video transmission on the Internet. Future television cameras may contain integrated chipsets so as to output their images in a highly compressed and universally compatible standard.

References

1. C. E. Shannon, A mathematical Theory of Communication, Bell Telephone B-1598, Vol. 27, July and Oct. (1948)
2. Klaus Holtz, Here comes the brain-like self-learning no-programming computer of the future, The First West CoastComputer Faire 1977 (1977)
3 J. Ziv, A. Lempel, Compression of Individual Sequences via Variable-Rate Coding, IEEE Inform. Theory, IT-24 (1978)
4 Terry Welch, A Technique for High Performance Data Compression, IEEE Computer, June (1984)
5 ITU V.42bis Data Communication over the Telephone network Data Compression Procedure, Geneva (1990)
6 K. Holtz, Packet Video Transmission on the Information Superhighway using Image Content Dependent Autosophy Video Compression, IS&T 's 48th, Washington DC (1995)
7 Klaus Holtz, Eric Holtz, Diana Holtz, Autosophy Information Theory provides lossless data and video compression based on the data content, EUROPTO'96, SPIE 2952, Berlin (1996)
8 K. Holtz, Advanced Data Compression promises the next big Leap in Network Performance, EUROPTO'98, SPIE Vol 3408, Zurich, Switzerland (1998)
9 Patents: 4,366,551 Associative Search Method; 5,917,948 Image Compression with Serial Tree Networks
10 Edward H. Sussenguth, Use of Tree Structures for Processing Files, ACM, Volume 6, Number 5, May 1963

Biography

Klaus Holtz received an Electrical Engineering degree in Hamburg Germany. From 1966 to 1969 he built a large computer center for medical research at Stanford University. In 1974 he formulated the Autosophy information theory for self-learning databases. In 1975 he patented the data compression algorithm in the V.42bis modem standard. He has published over 25 papers on information theory, data compression, encryption, and self-learning databases. He is the President of Autosophy. in San Francisco. You can contact him via Email holtzk@autosophy.com or on website: www.autosophy.com.