8/14/01

Empowering Digital Video Over the Web®

Vweb's MPEG Encoder Based Security/Surveillance Camera Subsystem

Amit Gulati and Bill Reckwerdt

Abstract

Vweb's family of MPEG based encoder chips is ideally suited to commercial/industrial security, surveillance and monitoring applications. Vweb's solutions customized for security offer dynamic bit-rate changes, dynamic frame-rate alterations, as well as dynamic resolution changes. In addition, a customized family of security specific low-power encoders in development will allow for a simple and very low-cost camera architecture mating a CMOS or CCD ITU-R BT.656 video interface input with a wired or wireless connection.

Introduction

Digital security monitoring is gaining strong momentum within the professional security industry. The rapid build out of broadband infrastructure, both terrestrial and wireless, as well as the development of standard and non-standard video compression algorithms, spanning a wide range of bitrates, has enabled the cost-effective introduction of digital surveillance.

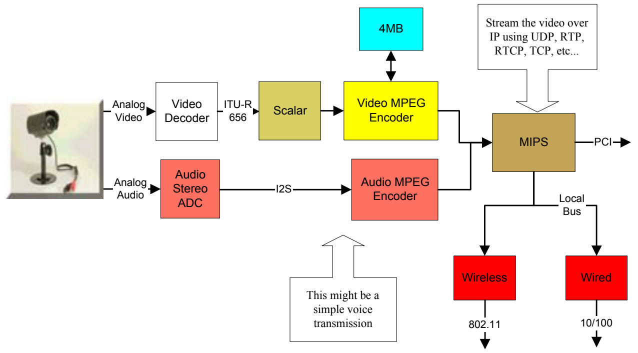
The first products in digital security relied upon analog phone (PSTN) transmission using h.263 to produce a barely acceptable CIF image at low frame-rates. On the other end, higher resolution security cameras relied upon wavelet or M-JPEG based solutions in order to meet the requirements for time-lapse and image fidelity, while compromising bitrate.

In reality, small alterations to real time MPEG based encoders (which offer aggressive bitrates due to the use of spatio-temporal bit-allocation as well as professional quality suitable for consumer DTV) enable all necessary "knobs" required of a security environment; the ability to vary the rate of compressed video data and monitor scene conditions by region. In this paper, we will discuss methods that enable Vweb encoders to deliver all the "knobs" necessary for a cost-effective MPEG2 based security camera application.

Camera Architecture

A network security camera device has two primary interfaces: A camera lens on one end and a network connector (e.g. RJ45) or wireless antenna (e.g. 802.11b). Vweb's family of Encoders support MPEG-1, MPEG-2 and MPEG-4, which allow scalable resolutions, flexible audio encoding, system stream multiplexing, a general purpose MIPS processor for networking and ancillary tasks. For the security application, the MIPS processor can be easily utilized for network layers. A real time OS with native IP stacks is ideal to support a full array of network protocols and QoS methods.

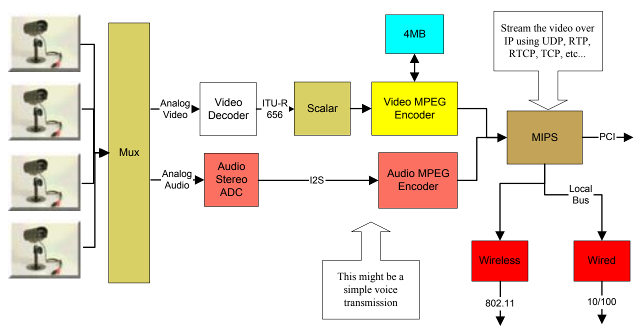
The network security cameras come in two flavors. In the first case, one camera outputs digital video over the network. The camera and the digital compression technology are housed within one unit. The following picture shows this setup.



The one-channel security cameras benefit from several features such as:

- Low Latency (encoding, muxing, transmission)
- Intrusion detection (motion sensing, audio triggering) and alarm verification.
- Variable bitrate with network & operator feedback.

The second case consists of multiple analog cameras multiplexed into one digital box. The cameras share the bandwidth. The following picture shows this setup.



The multi-channel security cameras benefit from several features such as:

- Reduced bandwidth usage.
- Improved compression compared to JPEG or Wavelet due to the inherent spatio-termporal algorithms. I.E. most walls have similar contours within a building.
- Reduced cost

As seen, the digital compression subsystem is common between the two setups. This allows reusability between the two designs.

In addition, security applications require a wide range of audio/speech support. Certain applications require audio return paths for gate entry, etc. Depending on the

5300 Stevens Creek Blvd., Suite 320, San Jose, CA 95129 Tel: (408) 615-1888 Fax: (408) 615-1889

application, Vweb's Encoder Family can support a very low bit-rate narrowband speech encoder such as G.723.1 or G.729a or a multichannel wideband speech coder such as MPEG Layer I/II.

Video "Knobs"

Network conditions, bandwidth limitations, recording resources and operator inquisitions demand varying levels of quality, bitrate and resolution (picture size and frame rate). The Vweb family of encoders can alter their bitrate from 64Kb/s to 15Mb/s while still achieving 60 fields per seconds. Furthermore, the number of field per second can be reduced to 2.

The Vweb family of encoders supports various dynamically controllable knobs that allow a great deal of flexibility desirable for security applications.

Dynamically variable bit-rate

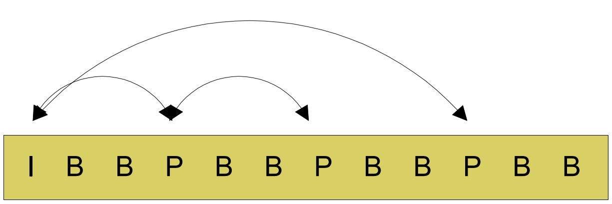
Vweb supports frame or even GOP granularity of bitrate alterations. These changes allow for bit allocation and quantization to be stretched to the rate-distortion limits. Beyond this limit, other means of bandwidth reduction must be employed.

Dynamically variable frame rate (time-lapse)

The MPEG standards were designed for 30fps interlaced or progressive sources. However, decoders are specified to be robust to frame loss. All encoder picture headers contain a presentation time stamp (PTS), which prohibits a decoder from displaying that frame until the decoder clock reaches the PTS value. As a result, sub 30fps bitstream can be created and played back on decoders that obey the PTS by redisplaying the previously decoded image until the desired PTS is reached.

The traditional method of achieving sub 30fps is to drop B frames, which are not relied upon for the construction of any neighboring frames. Dropping B frames in M=3 for example, leads to a 10 fps sequence. Similarly, dropping B frames at M=2 will give a 15fps sequence. At M=1, and N=[30/x], a frame rate of x fps can be achieved by dropping all the P's. This method, however, reduces the compression advantages of MPEG down to Wavelet or JPEG granularity.

M=3



Vweb supports another novel method for achieving any arbitrary frame rate. Vweb's encoders can be set to preserve any captured frame as the anchor frame for any number of succeeding predictive frames (for M=1 and large N). Thus, judiciously dropping P frames results in an IPP sequence at any arbitrary frame rate.

All of the above techniques rely on a MUX that scans and drops the appropriate frames, a task that does not add undue cycles to the MUX solution.

Dynamically variable resolution

At the low-end of bitrates, drastic measures are necessary to reduce the bandwidth utilization. The original 720x480 (NTSC) sources must be resized to 640, 2/3 D1, CIF, QCIF or sub-QCIF when appropriate to achieve the desired bandwidth/quality tradeoff.

Another knob security providers would like to have is latency enhancement (reduction) in an instance of uncertain activity within the scene or more loose latency requirements, when inactivity or known activities are occurring. While encoding latency typically cannot be reduced (it is determined by the number of B frames chosen, or the M value), it can be enhanced within the back-end. The bitstream FIFO can be tuned to respond to more frequent interrupt levels. In addition, extra IP frame headers/extensions can trigger certain networks to treat packets with greater real-time priorities (sometimes at an extra cost that should only be borne out of necessity, such as when the operator/user detects uncertain activity at a camera source).

Intrusion Detection/Triggering

A very desirable feature of security systems is to trigger an alarm when motion is detected, either in a region of the captured frame or in the entire frame, or when a peculiar audio event occurs. Vweb's video capture front-end's support field to field motion detection at 60Hz, regardless of the selected frame rate. However, a specific anchor frame can be used as the frame difference anchor, thus triggering at larger time-lapsed differences. The Vweb solutions can even detect changes in lighting.

Triggering on audio events is rather easily handled by Vweb's audio encoding subsystem. MPEG Audio Layer I/II wideband encoders perform a frequency translation in the front-end, and that information can be shared with a frequency domain notch filter before the masking effects of audio compression hide/remove masked frequencies in each critical band. Narrowband speech encoders also perform a low-latency frequency band translation, which can be utilized by the trigger/threshold mode before the encoder removes artifacts that do not resemble natural speech utterances.

Decoding clients can also be modified to detect motion based on MPEG bitstream parameters embedded into the compressed bitstream. While not as sophisticated as explicit absolute difference based techniques in the front-end, detection at the decoder based on motion vectors, bit-allocation variances, and luminance scaling.

Network Back-end

Vweb's encoders are naturally paired with a general purpose MIPS CPU. This traditional RISC CPU is a load/store architecture with reasonably sized caches and is capable of running a real-time embedded OS with a native IP stack.

Networks can be configured for customer environments, whether they require vanilla TCP/IP, or UDP/IP with RTP/RTCP and multicast. Networks configured to handle differentiated services (diffserv) via MPLS or RSVP can be configured to respond to network congestion/conditions as necessary.

ECN messages have gained popularity as a method of provider feedback to edge devices about adverse network conditions. A flexible encoder such as Vweb's that is closely mated to the network control processor can readily deal with situations by altering its bandwidth requires dynamically, all the while providing the fairest possible video quality achievable under the conditions.

Tips and Tricks

Typically, a feedback channel is necessary to communicate messages back to the camera (such as pan tilt zoom controls, etc), these must be handled by IP receive paths that affect processes on the MIPS control processor. Forward side-information, such as camera condition, pan, tilt and zoom settings and regionally activity measures can be embedded within the user data.

Vweb's family of encoders can take host requests for biasing the rate control algorithm in certain 16x16 regions of the source image. This method is ideal when a corner of the image is the most critical region for monitoring.

Most MPEG encoders, as Vweb's, offer rectangular search areas that provide a wider horizontal search range than vertical. If however, larger vertical motion estimation range is desirable (such as in freeway traffic monitoring), cameras can be oriented vertically and the resulting bitstream can be decoded in, for example, 480x720 windows in a display.

Transition to MPEG-4

MPEG-4 has been touted as an ideal encoding scheme for the security market. While it addresses several markets, only the core profile and above contain the measures attractive for security applications. Initial silicon implementations of MPEG-4 address the lucrative simple profile market (wireless phones, internet streaming) for extremely low bitrate communications. Vweb is addressing the main profile interlaced video definition, but with only 1 VOP (suitable for the video distribution industry).

MPEG-4 provides the facility to encode multiple planes of video, segmented by objects, natural and graphical, for later blending by a decoder that abides by the blending parameters in the bitstream (alpha planes). For security applications, this permits the ability to specify objects of interest and allocate bits judiciously in areas of activity near the object of interest (such as a car, door or gate). However, the art of segmenting a source image (video frame) in real time and classifying it for use by a multi-VOP encoder remains very illusive to both algorithm designers and certainly silicon implementers. It is without a doubt that multi-VOP MPEG-4 main profile level three (MP@L3) encoders will not surface cost-effectively in the near future. DSP based implementations exist, but utilize a crude segmentation scheme and cannot process better than core-profile implementations, unless multiple DSPs are segmented on one board.

By far, clever MPEG-2 based implementations permit rapid cost-effective deployments of robust security network cameras. Differences in theory vs. practical implementations of MPEG-4 today suggest that vendors would be better suited to defer the transition to MPEG-4 until cost and compute effective silicon is realized.

Synopsis

Our patented rate control and motion estimation schemes have provided the ideal tradeoffs between cost effective implementations (low gate counts, low memory utilization, low power) and high quality.

Vweb is committed to being the leading supplier of MPEG based encoding solutions to the security, surveillance, and monitoring applications market. Our solutions are the most comprehensive, leading to a highly integrated overall solution in cooperation with our silicon partners. Reference designs and SDK's are readily available via our sales department (see http://www.vwebcorp.com

VW2000 has been in production since September 2000. It is a MPEG-2 Video Encoder.

VW2010 will be sampling in the 4^{th} Quarter of 2001. It is a MPEG-1, MPEG-2 and MPEG-4 CODEC. It has a built-in MIPS processor.

VW2020 will be sampling in the 2^{nd} Quarter of 2002. It is a MPEG-1, MPEG-2 and MPEG-4 CODEC with Ethernet, Wireless, IDE, USB and 1394.

Contacts

Amit Gulati agulati@vwebcorp.com

Bill Reckwerdt billr@vwebcorp.com