N92-22453·1.

A VIDEO EVENT TRIGGER FOR
HIGH FRAME RATE, HIGH RESOLUTION
VIDEO TECHNOLOGY

Glenn L. Williams
NASA Lewis Research Center
Cleveland, OH 44135

ABSTRACT

High resolution, high frame rate video cameras using new image sensor technology will soon be able to make significant contributions in scientific and engineering research and long term scene monitoring. Researchers have in the past relied only on high speed photographic motion picture cameras for recording their scenes, with the attendant loss of real-time access to the images, in order to retain the advantage of high resolution. When video replaces film, while the images are available in real time, the digitized video data accumulates very rapidly, leading to a difficult and costly data storage problem. One solution exists for cases when the video images represent continuously repetitive "static scenes" containing negligible activity, occasionally interrupted by short events of interest. Minutes or hours of redundant video frames can be ignored, and not stored, until activity begins. A new, highly parallel digital state machine generates a digital trigger signal at the onset of a video event. High capacity random access memory storage coupled with newly available fuzzy logic devices permits monitoring a video image stream for long term (DC-like) or short term (AC-like) changes caused by spatial translation, dilation, appearance, disappearance, or color change in a video object. Pretrigger and post-trigger storage techniques are then adaptable to archiving the digital stream from only the significant video images.

INTRODUCTION

Ongoing and future microgravity experiments aboard the Space Shuttle or Space Station Freedom require high resolution high frame rate video technology (HHVT) to replace high speed photographic movie film which is heavy, bulky and which cannot be processed in space [1].

In the 1990's, advances in semiconductor CCD and CID array camera sensor technologies will permit fabrication of high resolution video cameras with frame rates much higher than the commercial broadcast television standard of 30 frames per second. The high resolution of these cameras will mean these cameras will compete with or replace photographic film camera technology. Digitized output from such a high rate video stream presents a difficult data storage problem, because data can be produced at rates 300 Mbyte/sec or higher. When this data is sent to onboard storage, total mission video data storage requirements will easily exceed one terabyte. Without careful attention to cost of storage and transmission, such vast volumes of data will become very expensive to support.

As a means of coping with the potentially huge HHVT data storage requirements, an advanced technology development effort at NASA Lewis has created an architecture for a Video Event Trigger (VET) using digital technology packaged on printed circuit hardware capable of being placed inside a personal computer.

Designed to detect onset of motion within less than 5 milliseconds after a new video frame is available from a digitizer, the system will support acquisition of many seconds of video frame storage when coupled with high

density frame store memory capable of continuously recycling uninteresting video frames. With pre-trigger and post-trigger capabilities, such memory could store an entire sequence of images including all the subtle details and changes visible just before the main event.

Built into the VET design is the capability to trigger on sudden image changes while ignoring slow changes, or to trigger on any short term or long term image changes based on differences from stored static reference images.

BACKGROUND

Research and development in the area of detecting and characterizing motion in the video images is not a new idea [2],[3],[4],[5]. Many U.S. and foreign patents cover implementations of scanning and processing video images to filter out noise, locate centroids of motion, and perform data compression of images by coding only areas of images where motion occurs. At the NTSC RS-170 standard video rate of 30 frames per second, hardware exists for quantifying and locating objects in motion within video images. Television scenes from missiles attacking Iraqi targets are recent demonstrations of such technology.

Our need for a Video Event Trigger stems from the HHVT requirements for up to 1000 frames per second of video data, where the ability to trigger on motion within one video frame time would be ideal. In such a case, before the very next video frame is fully digitized, the event would be sensed and special controls would go into effect to begin marking the data for long term storage.

To review, a video camera output drives a coaxial cable with a long sequence of analog signal voltages. Each sequence represents one scan of a raster line across the video scene, and a multiplicity of these sequences, controlled with horizontal and vertical synchronizing pulses, represents one complete video image, or frame. A device called a frame grabber acquires a sequence of raster lines, represents a video frame, and digitizes the signals with an analog-to-digital (A-D) converter, storing the resulting values temporarily in random access memory (RAM). Each 8-bit digital value from the A-D represents a number standing for the brightness and color of a localized region ("dot") of the video image, or "pixel". Screen images are composed of hundreds of thousands of pixels all laid out in columns and rows (such as 512 columns and 480 rows) so that the eye sees a complete picture. At a rate of 30 frames per second, this represents an average data rate of 7.37 million bytes per second. For image sensor devices to be used in the newest cameras, the sensor may actually drive several channels of flash A-D, with the digitized data temporarily going directly to a random access memory.

Of course, the natural consequence of continuously acquiring and digitizing live video becomes a problem of what to do with all the data. At the rate of 30 to 1000 frames per second or more, a real-time frame digitizer would create tens of millions to hundreds of millions of bytes of pixel data per second, to be supplied continuously to high density storage devices, such as magnetic or optical media, or to solid-state RAM. Video RAM or video tape (such as VHS cassettes) or video disk are among the choices of data storage technology in the late 1980's. Note that solid-state RAM storage space represents a high cost limited size memory having nearly zero latency and no moving parts. Whereas, moving magnetic and optical media can store enormous amounts of data permanently, but have quite long latencies. The obvious choice is to take advantage of RAM for temporarily buffering the interesting images during the time necessary to bring the moving media on line.

THE VIDEO EVENT TRIGGER

In typical settings, the human operator in the past has relied on his own observation of the scene or on external devices or externally processed electrical signals, such as from pressure or temperature or acoustic transducers, to create the video event trigger.

A review of processing rates of even the fastest digital processors will reveal that attempts to calculate triggers in software by analyzing frame-to-frame differences will not meet millisecond response requirements. Software algorithms, even on the fastest of computing devices, require many milliseconds up to seconds to analyze the multiplicity of pixels and determine whether the latest video frame has some new "interesting" change happening.

The nature of detecting "interesting" changes is not merely a semantic issue, since changes in color or motion may often be masked by considerable image clutter and may require some algorithmic processing of the image to interpret what is happening. Merely subtracting one video frame from another or looking for motion on the edges of a blob and/or calculating the movement of the centroid of the blob are ways that may fail to generate useful triggers with any one configuration of software rules about what constitutes interesting motion.

By comparison, hardware circuits can be customized to operate faster than software, by embedding comparison algorithms into silicon gate structures special to the application. However, embedded Artificial Intelligence algorithms to process images seeking trigger conditions could yield unsatisfactory results due to an underlying problem in the nature of sensing what is important.

Everyday examples of this problem include:

"Does this frame differ from the last one by more than 5%?"

"The blob didn't move - it changed color!"

"The stripes on that plaid shirt confuse my correlator."

"The cat's iris closed in the bright light, but otherwise stayed in the same place."

The trigger event is some form of change in the picture, determined by any or all of the following changes:

o Motion of the object of interest or in the area of interest (and ignoring motion elsewhere -- for some scenes may have objects moving elsewhere in the video image at all times, e.g. fans or flashing lights).

o Change of color in the area of interest.

o Appearance or disappearance of something in the area of interest.

Some not inconsequential complications to be accommodated are:

o The movement to be detected may only involve a very small object, which could be lost in pixel noise.

o Two full frames must be available for comparison. The comparison may not begin until after the end of a frame. The net result is that the trigger process lags one or more frame times behind the most recent frame being acquired.

o The process of motion interpretation/perception in visual images, which is accomplished in the human brain by millions of neurons making complex rapid-fire decisions, is exceedingly difficult to emulate in circuitry and/or software of reasonable cost and complexity. Software based Artificial Intelligence techniques of the late 1980's are very poorly suited to the task of understanding each and every conceivable video event without major intervention and reprogramming with new computer codes for each task.

One approach to controlling data storage volume and cost is to develop an algorithm which in real-time constrains storage to only include the digitized images of important events. These are defined as images acquired only when there is localized motion around some significant physical event. In NASA's microgravity experiments, video events can often happen after minutes or hours of inactivity prior to the event.

In a subset of cases the video images represent continuously repetitive "static scenes" containing negligible activity, occasionally interrupted by short events of interest. In these cases, minutes to hours of redundant video frames can be ignored until something new happens.

Common examples of such scenes include:

o automobiles crashing into walls in safety tests, where the front bumper comes into view
of the camera only in the last fractions of a second.

o long term automatic monitoring of parking lots.

o long term monitoring of entrance to secure or hazardous areas.

For our purposes, we require video event triggering in milliseconds, and simultaneously a low cost, small volume detector. The best compromise appeared to be a fast parallel processor which is autonomous but not elegant. In concert with hardware speedup, there is the need to customize the detection process by adding an ability to make incremental or gross corrections to the algorithm on a frame-by-frame basis.

Fortunately, recent advances in the industry have yielded new techniques and algorithms for performing rapid evaluation of images. These new circuits fall under the collective name of Fuzzy Logic. There are also analog neural networks, which rely on analog components such as capacitors and resistors and operational amplifiers, and digital neural networks, which rely on digital processing of numbers to make similar decisions. Digital neural networks provide extremely consistent decisions without being affected by analog effects such as temperature, charge leakage, and power supply noise. However, to date all neural networks require extensive "training" using "typical" data. But video events are characterized as one-shot unpredictable changes that often are difficult to assign to any particular appearance (as in the above examples).

In our architecture (Figure 1.), a computer containing a "frame grabber" is attached to a video camera, and video frames are acquired using standard hardware setups. Via software and hardware in the computer, control of the acquisition process results in video frames being presented sequentially to the VET logic control circuitry where they are captured and stored into temporary memory buffers (1, 2, ..., 5). Buffer 5 holds the oldest frame, buffer 4 the next oldest frame, and so on, with buffer 1 holding the newly acquired frame, the one to be compared to all the others.
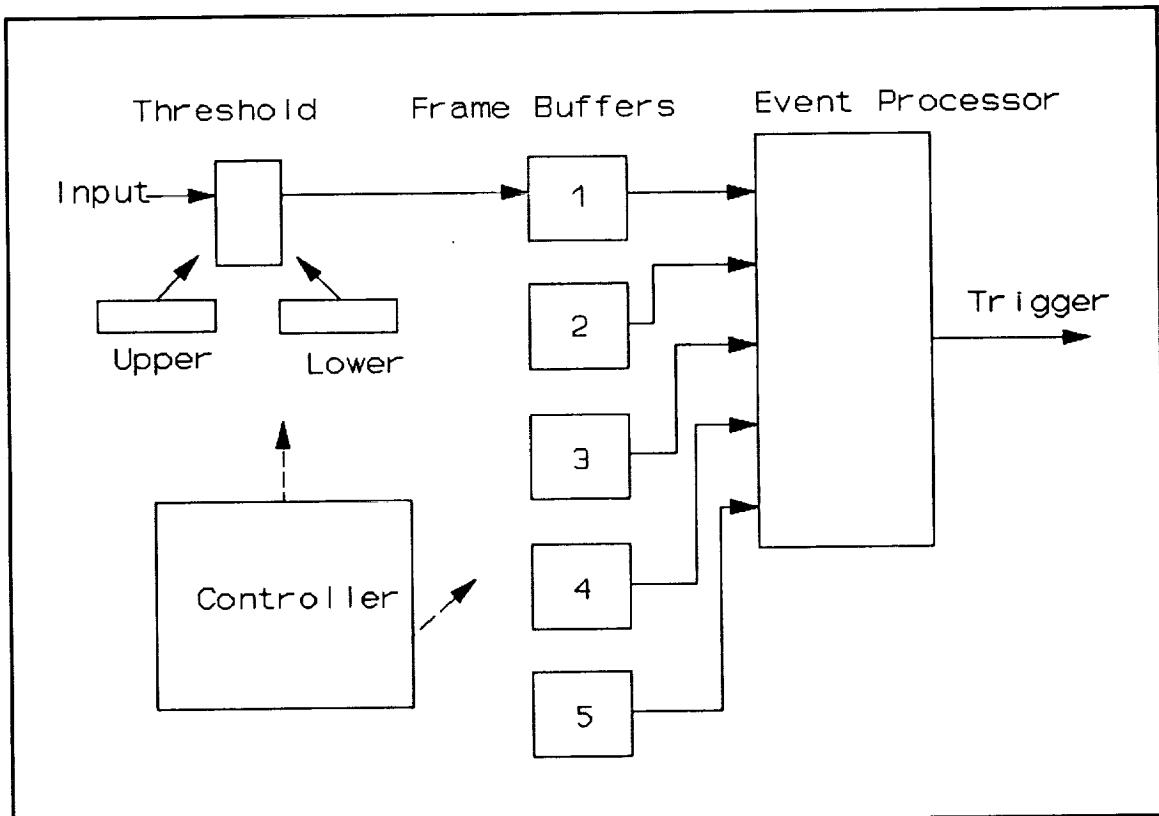
Figure 1 Block diagram of the Video Event Trigger Subsystem.

During system initialization, with the video running continuously, the first few frames are assumed to contain no motion and are stored one-by-one into the k buffers, to be used as the "learned" reference frames. Learning in this case merely means loading up the frame buffer memories, one of which can be loaded every frame time.

There are two methods of storing and comparing old and new video frames. In the first of two methods, each new video frame is compared against a set of k-1 older frames via subtraction and fuzzy logic rules. If no motion is detected, the oldest of the k stored frames is discarded by rearranging pointers to the video frames. Effectively, all the frames are shifted down, and the newest frame is assigned to the first of the reordered k frames. Then the next frame is acquired and the cycle is repeated. This method has the advantage of being able to ignore very slow changes in the video scene, similar to a slow DC drift in a one-dimensional analog signal. Only dramatic changes in the latest video would constitute enough motion to set off a trigger. This is similar in function to "AC Coupling" on an oscilloscope.

In the second method, only the first of the k stored reference frames is ever updated. Each new video frame loaded into buffer 1 is discarded immediately after use. The remaining k-1 stored frames are permanent, non-changing reference frames. The method assumes images can contain either slow changes or rapid changes. ANY changes are important to the motion detection process, and if they occur, they must be reported when a certain threshold is exceeded. This is similar in function to "DC Coupling" on an oscilloscope.

These methods parallel the architectures of one-dimensional FIR and IIR digital waveform filters. However, we avoid the use of recursion, i.e. feeding output images back into the input data stream. We thus avoid problems with instability and limit cycles. But otherwise, our techniques have a close similarity to more commonly digital filter architectures, for which a large pool of documenting literature exists.

We enhance the operation and reduce complexity in the processor by globally thresholding (clipping) the video levels with two digital binary comparators and two "cut" levels. We then have a "window" comparison of the video. "Below Level", "Above Level", "Inside Window", and "Outside Window" are the four choices that result. These levels are merely programmed into storage latches under computer control. These two levels represent the "alpha-cut" (variable sensitivity) levels that determine which levels of gray (or color attributes) will be reduced to a binary ONE by the comparator. All other levels converted to binary ZERO. Then, after the operator's selective adjustment of the alpha-cut levels, the event processor uses only these clipped images. The processing load is thereby greatly simplified, but doing so is not a requirement in the general case, should a design require full gray-level sensing.

ADVANTAGES OF THE VIDEO EVENT TRIGGER

Advantages have already been cited above, in some cases. A summary here would include:

o Able to rapidly sense motion in a video image, in real time operation.

o Relatively inexpensive to construct, and can qualify video images for motion without a large amount of expensive and time consuming hardware or software customizing.

o Ultimately capable of operating stand-alone from any computer and/or microprocessor.

MEMORY REQUIREMENTS

Most applications require minimizing the cost of storage of video images. The quantities of data resulting from high frame rate or image resolution conflict with the need for low cost storage.

In a particular example, assume frames of video data can be stored sequentially and cyclically, a frame at a time, in video RAM storage. High speed, large volume RAM memory boards are commercially available and could in theory be modified for this purpose. Assume the existence of a memory controller, which makes sure that the storage is cyclic, in such a fashion that the very oldest frame is overwritten (lost) each time by a new frame being stored into the memory. We make the "obvious" assumption that the oldest frames carry no data of any value (nothing happened).

Upon an operator "arm" command, the hardware inside the controller starts filling a memory buffer with "pretrigger" data from the digitizer. Once the minimum requirements of the pretrigger buffer have been satisfied, the remaining portion of the buffer is treated as post-trigger data. During the interim, until the trigger signal arrives, the memory is controlled in a fashion similar to a continuous loop magnetic tape recorder. Since the memory has a maximum capacity, the oldest data is continuously replaced with the newest data until the trigger point.

When a video event occurs, the trigger pulse signals the RAM controller to begin a new phase of frame storage algorithms. In this new phase, some of the oldest frames (still redundant) are overwritten by new, interesting frames containing motion. But a selectable number of frames of medium age are retained in memory because they may contain images of precursor activity important to the history leading up to the event. At the trigger time instant, the act of triggering sets a digital logic switch which causes the wrap-around to cease. Thereafter, when the remaining amount of post-trigger memory is filled, the recording stops. The net effect is that the memory holds the entire useful record of the transient, both before and after the trigger point, depending on the size of the "pretrigger memory" setting. All that is required is a little pointer arithmetic to unwrap the data in memory (Figure 2.).
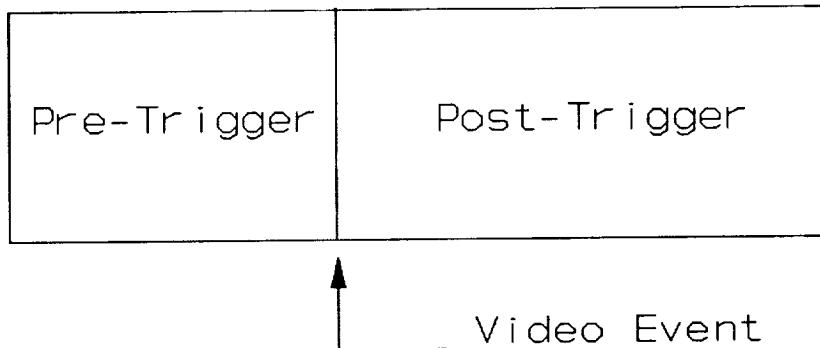
Figure 2 Frame store memory can be divided into pre-trigger
and post-trigger portions.

In practice, what the operator directly or indirectly programs into the controller is "pretrigger count" or "pretrigger percentage", the value of which is used to compute a difference. This difference is the total memory size to be used minus the portion to be allocated to pretrigger data. Upon being armed, the controller first blocks and ignores triggers while filling the pretrigger memory portion. As soon as the pretrigger portion is filled, normal triggering is instantly enabled without skipping any frames. Under normal operation, the memory is filled by the wrap-around method described above, with the difference value used to control the size of the post-trigger buffer.

ACKNOWLEDGMENT

BIBLIOGRAPHY

[1] High Resolution, High Frame Rate Video Technology, *NASA Conference Publication 3080*, May 11-12, 1988.

[2] A. K. Jain, "Image Data Compression: A Review," *Proc. IEEE*, Vol. 69, NO. 3, March 1981, pp. 360-362.

[3] J.R. Jain and A.K. Jain, "Interframe Adaptive Data Compression Techniques For Images," Sig. & Image Proc. Lab., Dep. Elec. and Comput. Eng., Univ. CA, Davis, Aug. 1979.

[4] A. Rosenfeld and A. Kak, *Digital Picture Processing*, Second Edition, Academic Press, 1982.

[5] R. Schalkoff, *Digital Image Processing and Computer Vision*, John Wiley and Sons, Inc., 1989, pp. 190-192.