Video Pre-Compression Processing

& Bit Conservation

by

Jed Deame, VP & GM

Teranex Division
Silicon Optix Inc.
http://www.Teranex.com

Abstract

In the current video delivery market content providers are being faced with the ever-increasing demand to provide their content in multiple formats, multiple compressions standards and multiple bit-rates. As the major CODEC providers create ever more powerful and bit efficient compression algorithms, bit rates are being driven down. A fine line is being drawn between acceptable video quality and lower bit-rates to provide more economy in serving the media over the distribution channel.

There is a growing distribution chain over multiple delivery methods - cable modem, DSL, IPTV, digital TV, data-casting and wireless devices. Each will require different bit rates and associated quality levels. Today's bit-rate sweet spot will not be tomorrows.

Also, the source content can take many formats - with video quality ranging from digitally mastered tape to consumer grade analog tape. And the content itself can be film-originated material from a telecine or video originated material with interlace artifacts.

The solution to this problem is a video pre-processor which accommodates the various film or video based material; understands the varying possible output formats and resolutions; and provides powerful noise reduction processing which enables the downstream processing to be minimized and the resultant compressed video quality to be maximized.

The performance and flexibility of the Teranex Video Computer platform enables these complex functions to be combined into a single compact, cost-effective platform. This software-based real-time platform enables subsequent high quality video processing and the flexibility to add capabilities over time easily via software.

Introduction

The basis of pre-compression processing is the minimization of noise present in the source material so that the compressor may better utilizes the small amount of bits that it is allocated to work within. Noise is accumulated onto a video sequence as it progresses through the production process. Analog to digital conversions, generational losses, compression losses and many other production factors can lead to an eventual increase in noise and corresponding loss in picture quality.

Video is becoming available on a myriad of devices - not just the traditional TV served with an over-the-air signal. Digital TV is being implemented to eventually replace analog transmissions and utilize the spectrum more efficiently. Cable television has embraced digital transmission to better utilize the available bandwidth in their network. And of course, the Internet is providing streaming and downloadable digital video content to the consumer and business desktops worldwide. With this wider distribution model, compression has become necessary to best utilize the bandwidth required to provide this video content.

One of the challenges in compression is the huge amount of reduction which is required to be performed by the encoding algorithm. A standard NTSC signal provides 30 frames per second (fps) at 720 horizontal and 486 vertical pixels with a pixel depth of 20 bits per pixel. This yields a data rate of over 200M bits per second (bps). The job of the encoder is to produce the best possible image quality at the following typical rates:

Target Bit Rate	Compression Factor
56 Kbps	5,400 to 1
300 Kbps	720 to 1
1 Mbps	200 to 1

Each of these target rates are typically provided at varying sizes and frame rates which certainly affects the compression ratio's significance. Obviously, the more bits applied - the better the resultant quality *should* be. If bits are squandered by the encoder faithfully replicating noise components of the source material then the viewing quality and/or bandwidth requirements of the resultant stream will suffer.

Noise Analysis

Computers and humans view the digital video content in a very different manner. In Figure 1, the human perception of a roller coaster is shown on the left side. On the right in the same figure, is a representation of the same image to the computer. This computer view of pixels is useful to understand the CODEC's challenge in encoding material. Noise, which would be difficult for a human to see in a dark area of the video, is readily apparent to the computer, which detects the variance in the pixels value. (Note: For this plot only the dark pixels are represented by larger number values which produce a taller 3D value than a brighter pixel which is represented by a smaller value. Latter images will show the inverse of this - the brighter pixels will have larger values and taller 3D values.)

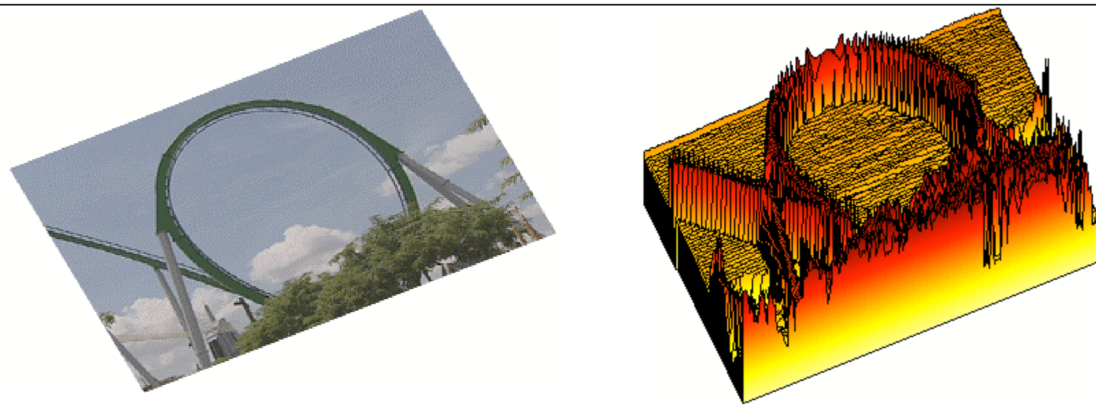


Figure 1: Different Perception of Same Image
(Human perception on the left, Computer perception on the right)

Compression - Frame types

In its simplest form, compressed video can be represented as an encoded frame of video followed by a number of "difference" frames. This first frame of video is called an "I" frame and can be thought of as a stand alone compressed image - much like a JPEG compressed image. The difference frames, or "P" frames, which follow are used by the encoder to attempt to only encode those portions of the "I" frame which have changed as the video plays. After a certain number of frames have been compressed, or a scene change is detected, the compressor will make a new "I" frame and the process will start over again.

Figure 2 shows a stylized version of the compressed stream with a stand alone "I" frame on the left followed by "P" frames as the video plays. The bright object in the "P" frame is the roller coaster moving along the track. The rest of the frame has no movement, and when differenced with the "I" frame, produces a black area - or an area with pixel values close to zero.

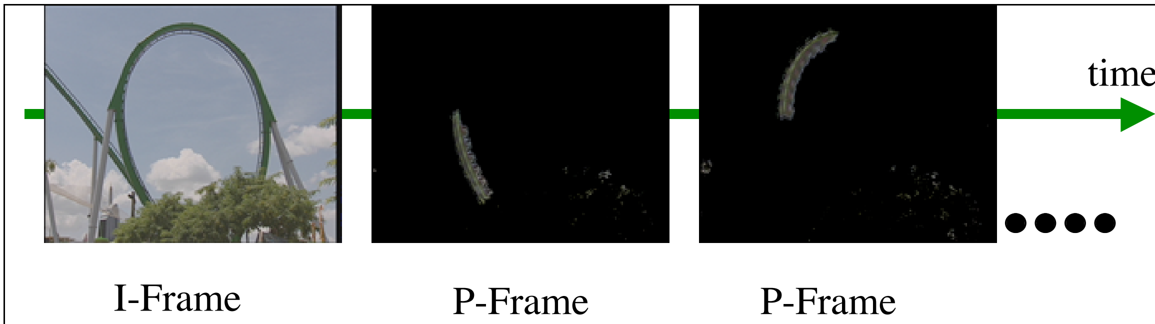
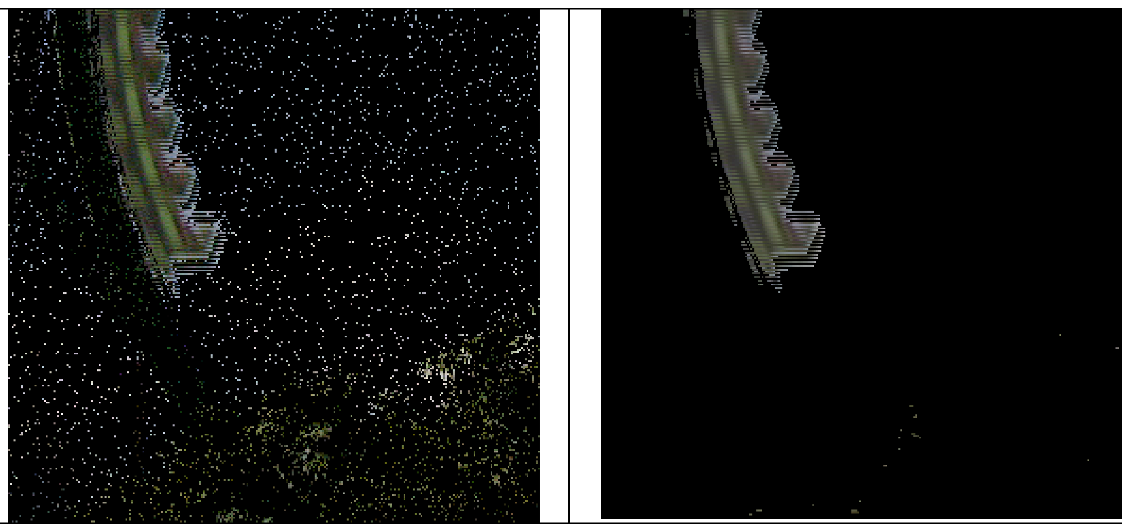


Figure 2: A Stylized Version of the Compressed Video Stream

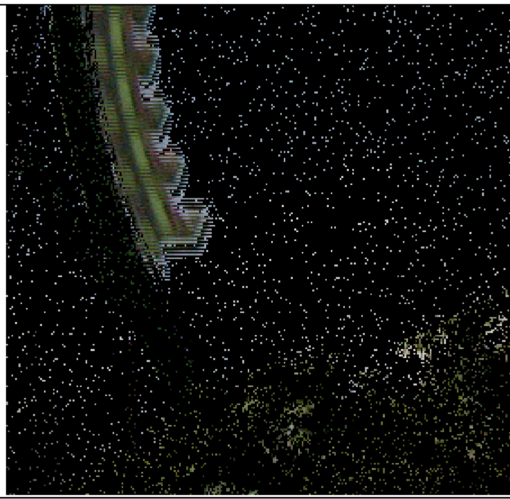
This is how it is supposed to work, assuming the source material is very clean. Unfortunately this is rarely the case. If we zoom in on the object in the "P" frame and the area surrounding it, we see a more typical image on the left side of Figure 3. As can be seen, the image has much random noise. This noise is being coded as motion and is using up valuable transmission bandwidth. If we apply Teranex imageEnhance Pre-Compression Processing, the image on the right in Figure 3 is produced.

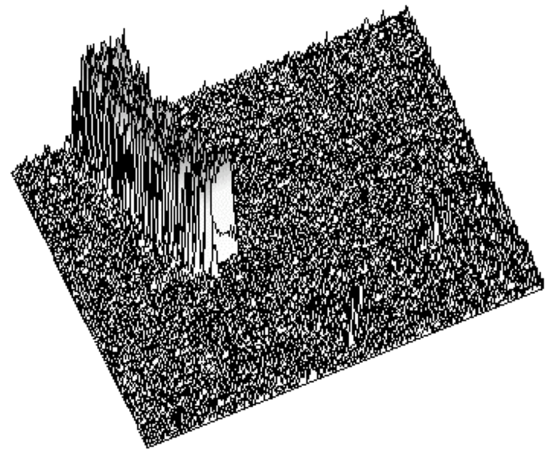


Original DigiBeta Video | After Pre-Compression Processing

Figure 3: Difference Frames Without and With Noise Reduction

To examine the input noise more closely, a 3D plot of the original DigiBeta source image in Figure 3 is shown in Figure 4. From this plot it can be seen that there is significant variation in the floor of the video and this noise floor will drive the CODEC to send much more data in the compressed stream to allow for these variations. This extra data will drive the bandwidth requirements up, or conversely, it may drive the CODEC to use valuable bandwidth for the noise instead of some of the desired motion portions of the video, resulting in image artifacts. Remember that this one frame is replicated 30 times per second while the video is playing and the noise will be randomly distributed throughout the frame.

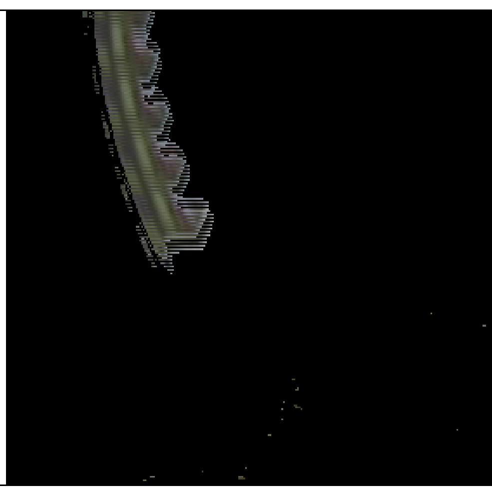


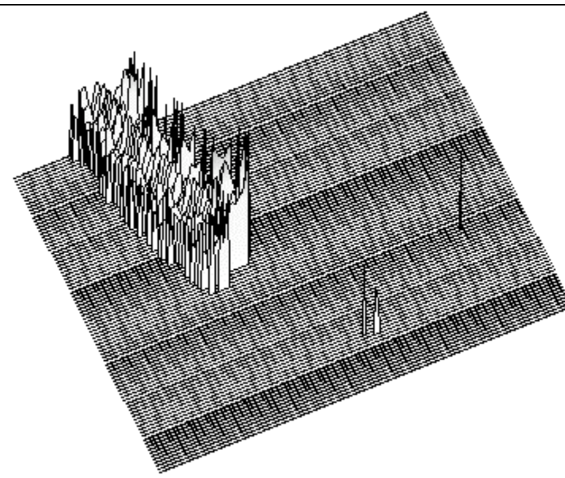


Original DigiBeta Video Material	3D Plot of Pixel Values

Figure 4: Noisy video material

The processed frame from Figure 3 with the resultant noise reduced image is shown in Figure 5 along with its 3D plot. As can be seen, the base level or "noise floor" of the plot is very flat. As this noise reduced video is compressed, the CODEC will not see much variation from frame to frame and hence will have a difference image with mostly black - or small difference pixel values. The CODEC does not use any bandwidth or storage for difference areas that are black.





Teranex Noise Reduced Video Material	3D Plot of Pixel Values

Figure 5: Cleaned video material

The Pre-Compression Noise Reduction Tool Set

The noise examples shown above and many others are addressed by the utilization of pre-compression processing filters. In particular, the following noise filters will be discussed.

- Temporal Recursive Filter
- Adaptive Spatial Median Filter
- Brick Wall Filter

Pre-Compression Noise Reduction

Pre-compression processing is a step that enables the storage of highly compressed data with minimal quality loss. Unlike simple noise reducers employing spatial averaging, pre-compression noise reduction encompasses motion adaptive Temporal Recursive Filtering, adaptive Spatial Median Filtering and Brick-Wall Filtering. The purpose for this type of processing is to assist video encoders by removing or attenuating random and impulse noise, and to reduce the total energy in the output stream via the reduction of high frequency signal content.

First, the source video must be filtered to remove any undesired noise. This stage of pre-compression processing consists of removing impulse noise and attenuating the impact of random noise. Examples of random noise are listed below.

Random Noise Sources:

- Residual noise and grain - archival or generational
- Low contrast film dirt and scratch noise, horizontal or vertical
- Film grain
- FM sparkle
- Compression artifacts
- Cross color from composite decoders (e.g.: VHS/U-Matic Noise)

Random noise is removed and/or attenuated by motion adaptive Temporal Recursive Filtering. Motion adaptive temporal recursive filtering compares adjacent frames, classifying each pixel into one of three categories: noise, motion, or no motion. Pixels whose comparisons are less than a threshold are categorized as having no motion. For those pixels whose comparisons are greater than the threshold, spatial processing is employed to determine if the pixels belong to an object in motion, or if the pixels are to be categorized as random noise.

To minimize artifacts that may be introduced due to improper handling of objects in motion, pixels that are on the boundary of these motion objects are not temporally filtered. Pixels classified as not being in motion receive temporal filtering via an average value, median value, or some other mathematical combination of temporal pixels to smooth out peaks and valleys in the data stream. This attenuates the effects of low level random noise that has been introduced into the image. Pixels that have been categorized as noise are replaced with spatial filtering techniques, again via some mathematical and/or morphological combination of neighboring pixels.

The second kind of noise of interest is impulse noise. These artifacts are not as detrimental to compression as random noise because they are non-periodic and appear typically less often in the video stream, but they need to be removed nonetheless. Examples of impulse noise are listed below.

Impulse Noise Sources:

- Tape drop out
- Satellite glitches
- Analog clamping errors
- Bit errors in digital transmission

Impulse noise is removed by specialized algorithms that locate the desired artifact and replace the affected pixels with information from surrounding pixels. These artifacts are typically much more difficult to locate, and are harder to mask since they may affect numerous pixels in the frame. Adaptive Median filtering solves this problem by comparing adjacent pixels in the video frame.

The Adaptive Median Filter performs spatial processing to determine which pixels in an image have been affected by impulse noise. Strictly applying a median filter may destroy small structure (e.g., whiskers on a cat), as well as eroding away edge boundaries. Replacing strictly a maximum or a minimum valued pixel may remove structurally sound pixels as well, or not remove all of the noise when the impulse noise is spread across more than one pixel.

The Adaptive Median Filter classifies pixels as noise by comparing each pixel in the image to its surrounding neighbor pixels. The size of the neighborhood is adjustable, as well as the threshold for the comparison. A pixel that is different from a majority of its neighbors, as well as being not structurally aligned with those pixels to which it is similar, is labeled as impulse noise. These noise pixels are then replaced by the median pixel value of the pixels in the neighborhood that have passed the noise labeling test. A prime benefit to this adaptive approach to median filtering is that repeated applications of this Adaptive Median Filter does not erode away edges or other small structure in the image.

The ext stage of the process is bandwidth limiting. This is where Brick-Wall Filtering becomes important. It has many times been proven that bandwidth reduction prior to compression has substantial benefits. This is true particularly as the bit-rates are reduced. MPEG encoders begin to create artifacts such as blockiness and "mosquito noise" if they are tasked to encode high detail video at low bit-rates. The compression process indiscriminately "chops off" frequencies, via DCT quantization, generating significant compression artifacts. To prevent this catastrophic breakdown of the imagery, brick-wall filtering may be applied.

A brick-wall filter is typically a very high-order low-pass filter with a sharp cut-off above the designated frequency that removes excessive detail in a controlled fashion - detail that would be otherwise eliminated by the compressor resulting in visible artifacts. By controlling the manner in which the detail is removed, compression artifacts can be minimized. To minimize the blurring that can occur when high frequency content is filtered, a "boost" can be applied after the brickwall filter to accentuate the remaining edges in the filtered image.

Input Source Detection & Handling

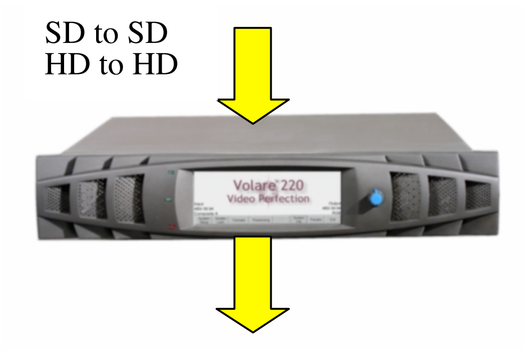
Another important step in compression pre-processing is determining whether the source is video or film originated. This is an important step in the interlace to progressive conversion that takes place at the beginning of any high quality digital process. If interlaced material is detected, a de-interlace algorithm is applied. This is necessary because the odd lines of the image are spatially shifted from the even lines, due to inter-field motion. This misalignment must be compensated for in order to create visually pleasing progressive images. If film originated material is detected, a 3:2 film frame matching algorithm is applied.

There are many different varieties of de-interlacing algorithms in use, ranging from linear interpolation to motion adaptive interpolation to the high-end per-pixel motion adaption. It is important to remember that the quality of the de-interlace algorithm is directly proportional to the amount of processing power available. With ultra-high performance real-time software-based converters, very elaborate and compute intensive algorithms may be implemented.

The Teranex Pre-Compression Solution

Teranex has developed the ImageEnhance pre-compression processor to address the noise reduction, film and video processing required for DTV, broadband and other emerging distribution methods. The Teranex pre-compression processing solution is based upon a software reconfigurable processing engine that enables truly supercomputer performance in a rack mountable form factor. Figure 6 shows the functional operations performed by the ImageEnhance system.

One important feature to look for when selecting a pre-compression processor is flexibility. Since the number of DTV formats in use is increasing rather than decreasing, there is a significant advantage in pre-processors that are updateable to emerging formats, so as not to obsolete the substantial investment into a processor. This "future proofing" concept is just beginning to become popular with the introduction of real-time software-based pre-processors, such as the Teranex Video Computer, which can be updated via CDROM to support new features and capabilities as they emerge.



1. Film/Video Handling
2. Interlace to Progressive
3. Spatial Filtering
4. Temporal Filtering
5. Brick Wall Filtering
6. Aspect Ratio Control
7. Proc Amp

Figure 6: Pre-Compression Video Flow

Teranex TVP Technology

In order to implement the functions described above in a flexible, programmable, interchangeable system, a radically different video processing architecture has been developed. As a result of over one hundred million dollars in R&D investment from the U.S. Government, a parallel processing supercomputer technology has been developed that is capable of extremely high compute power with minimal component real estate. The Teranex Video Processor **(TVP)**, is a fine-grained, massively parallel, two dimensional mesh computer architecture that is uniquely efficient for two dimensional, video processing applications. Technically, TVP is a Single Instruction Multiple Data (SIMD) architecture; and due to advances in semiconductor technology and processor, highly efficient SIMD systems capable of real-time video processing are now practical to implement in compact 19" rack mount units.

Effectively, TVP is a collection of thousands of RISC type computers operating in unison. The computers, called Processing Elements (PE's), are arranged in a two dimensional grid – each PE linked to the four PE's closest to it in the grid. Each PE has its own data memory, and all PE's run the same program simultaneously. In the latest generation of TVP, called P4k, there are 4096 PE's on each chip – arranged in a grid 64 PE's wide by 64 PE's high. Each chip can execute over 400 billion operations per second (BOPS). Furthermore, the architecture is linearly scalable – in the 2RU Video

Computer, the chips are combined in a two-dimensional arrays of 50,000 PE's, yielding computing performance of 5 trillion operations per second (TeraOps).

The Teranex Video Computer Platform

The TVP technology, packaged in a compact 2RU rack mount system, enables Teranex to provide a reconfigurable platform which can support all existing and emerging DTV formats using the highest possible quality algorithms. The end result being a noise reducer that can adapt to new formats and processing techniques as fast as they evolve, and provide a truly future-proof solution. The same hardware platform that performs as a noise reducer, and film/video handler, can be reconfigured to perform as a DTV format converter with a simple software key. The Video Computer platforms are available in 2RU, 3RU, and 6RU sizes with performance up to fifteen trillion operations per second.

The Video Computer platform is the basis for all current and future Teranex products. The programmable nature of the processing engine enables the system to be updated or upgraded via CD-ROM. This reconfigurability enables additional features or enhanced algorithms to be installed in the same basic platform.

Conclusion

The Teranex ImageEnhance Pre-Compression Processor provides a unique solution to video noise reduction. By implementing a range of noise reduction filters, any subsequent compression stages can more efficiently allocate bits to produce the best possible image. Additionally, the intelligent handling of film or video originated material ensures that the filtering is employed correctly based upon the provided video. Lastly, all of the described functions are implemented in a real-time software-based system, enabling continual improvements to be added over time as new requirements emerge.