Gary Demos
Chief Technical Officer and Founder, DemoGraFX
4640 Admiralty Way, Suite 500
Marina del Rey, CA 90292
(310) 578-7848
(310) 578-7849 (Fax)
gdemos@demografx.com

Type: Long Paper. ***Potential of Motion Imagery Systems***
Demo: Yes
Demo Title: ***Moving Image Compression Quality***
Demo Description: DemoGraFX will be demonstrating proprietary layered compression technology. Video images captured electronically at 720p/60 of the NASA Shuttle launch (compressed to 10.1 Mbits/second) and Andrews Air Force Base air show (compressed to 9 Mbits/second) will be viewed side by side (butterfly) with the original footage for comparison purposes.

DemoGraFX will bring our own demo equipment (including CRT monitors for viewing).

Potential of Motion Imagery Systems

Abstract

Motion imaging is an important technology in today's world. This technology allows us to collect vital security information thereby providing safety for our citizens and our country. It is also an important communication media to serve information to the public. The success of protecting and disseminating information is only as good as the systems that are used to obtain the information.

With the advent of the digital age, the electronic moving image has made several substantial recent advances in the entire system. The improvements extend from the camera, through recording, editing, archiving, distribution, and presentation. Presentation systems have greatly improved in quality, especially in projection displays, which can now be bright, sharp, and stable.

Various system parameters of the moving image are being explored to extend beyond historical television capabilities. The frontiers include increasing frame rate, resolution, the range of color, the dynamic range, and reducing the noise floor.

When conceiving the moving image system as a chain from camera to eventual display, compression has proven an essential ingredient. Often, the camera, the display system, or the compression system has limited the quality of the image system. However, technology now exists such that no weak system link need limit the system. A consistent level of quality can be achieved from the camera through the system to the display.

In addition to basic quality improvements, other improvements have extended the range of moving image capture. An example is the stabilized long zoom lenses for cameras. In projection, the TI DLP Micromirror projector has demonstrated stable bright images. Alternate technologies are attempting to increase projected resolution.

Overall, the delivered quality of moving images made electronically is now beginning to rival, in some ways exceed the quality of film-based moving image systems.

Introduction

The moving image has been transformed in the past two decades from analog electronic television, and chemical filmstrips, into the world of digital processing. The highest resolution imaging has been film-based, but the typical film frame rate is a low 24 frames per second (fps). Video has been sampled at the higher rates of 50 and 60 interlaced fields per second from inception. Only recently, however, has interlaced been removed in some color video systems. The result is a dramatic improvement in the quality and usefulness of the electronic frames. While interlace is still a widespread problem, the trend toward its elimination is clear.

Until recently, video had either low resolution and low noise, or high resolution and high noise. However, electronic high-resolution cameras are now beginning to rival the best films in noise floor relative to film grain levels. The dynamic and color range of electronic cameras is also beginning to rival that of film negatives (original motion picture film is photographed as an inverted "negative" which s reversed in printing).

Electronic color video cameras with frame rates as high as 72fps, with interlace eliminated, clearly demonstrate the quality which is possible in the moving image. With these major steps forward within the camera, the rest of the imaging system comes to the fore.

On the digital projection end of the system, the digital T.I. micromirror has demonstrated the potential of high-brightness and clarity having a wide dynamic range and color range, as well as being stable in illumination level. Although these projectors are now expensive, their existence proves the value and feasibility of such projection systems. This challenges both T.I. as well as the rest of industry to find ways to reduce the cost of these systems, as well as continuing to improve their resolution, frame rate, and dynamic range.

Video as a Complete System

Video cannot be adequately conceived except as an entire system chain. The delivered image quality will be the result of the concatenation of all of the processing steps applied to the image from original light capture through to the final displayed light. As with all system chains, the quality and behavior of the system will be most substantially affected by the weakest links. In typical digital video system chains, the single weak links far outweigh the affects of the concatenation of stronger links. However, all of the links affect the result, and when the weak links are greatly improved, the other links must be more carefully scrutinized.

At present, the two weakest links are typically compression and interlace. In the presence of interlace, and poor quality compression, nearly all other system aspects are overshadowed.

Compression quality can now be very high using improved coding technology (such as that developed by the author and his colleagues), even at the low bit rates typically used for distribution of digital video. Thus, for high-resolution video absent of interlace, the remaining system ingredients become significant to delivered quality. These include the camera, the processing performed in post-production, and the processing performed subsequent to decoding, but prior to display. Also, any other concatenated production compression will be significant (such as that used by camcorders and tape systems).

Cameras

Cameras are themselves a system. Cameras have an optical and lens system ahead of the sensors, including dichroic and other color filters. The sensors themselves are usually analog photon to electron converters, such as CCD’s. The next step will be low-noise analog amplification and some analog processing. Then analog to digital conversion is performed. Once digital, the camera will perform adjustment, and compensation processing. Also, noise reduction and sharpening are preformed.

In CamCorders (such as DV and DV-HD), the picture is then compressed and stored digitally onto helical tape.

All of these camera elements have a substantial effect on the image quality. The best of today’s cameras can perform all of these steps at uniformly high quality, even at high frame rate.

One area requiring additional attention is the temporal sample. A moving image camera will open a shutter during most of the frame time. The shutter will close as the frame advances, however. This form of temporal sample is less than ideal. In some cameras, the shutter duty-cycle is 50% or less, resulting in gaps in the viewing. Approaching 100% is highly desirable, but is not a complete solution, since the uniform sample is not an ideal theoretical sample, and is prone to some types of temporal aliasing. As the frame rate increases, this aliasing becomes less problematic, but it cannot be eliminated without changing the nature of temporal sampling in cameras.

Computer CRT Displays

Displays and monitors on computers have long exceeded television set displays in quality, but not in size. At present, a typical desktop cathode ray tube (CRT) computer monitor is very inexpensive (typically $300), and is capable of 10-bit dynamic range and one MegaPixel resolution. Such CRT displays are also sufficiently bright that 60 Hz flicker is a significant bother to most viewers. For this reason, computer displays have used a refresh rate exceeding 70 Hz since the mid 1980's.

Computer Display Processing Boards

With CRT desktop computer monitors, the image quality is limited by the display system within the computer, and not by the CRT. Typical computer display boards are presently limited to 8-bits of dynamic range.

Further, when decompressing pictures through such boards, such as when playing a DVD, which requires decoding mpeg-2, the display system performs numerous other processes. Such processes include attempting to de-interlace to find the original frames at 24fps, or use "bob and weave" to choose between presenting a duplicated field or a weave of a pair of fields if the frame reconstruction is not evident to the detection algorithm. This de-interlace and un-3-2-pulldown process becomes problematic when field-rate motion is introduced, such as when pan-and-scan is performed in a video effects switcher that updates the pan position each field.

Another related process is the conversion of color space information stored as luminance and two chroma channels (known as U and V) within mpeg-2, into the RGB color space used by the CRT display. This computation is also usually performed using 8-bit precision, resulting in additional precision and quantization degradation to mpeg-2's 8-bit decoded pixel values.

Another process typically performed by the display board in the computer is up sizing of low-resolution video images (such as the DVD mpeg-2 decoded picture, "streaming" video, or other decoded non-mpeg-2 decompressed video). An up-sized picture on a computer screen will be fuzzy or soft in appearance, since the computer screen has much higher resolution than the video being decoded and enlarged.

Another problematic attribute of decoding pictures for a computer screen is the lack of synchronization between the video 24fps movie rate (when successfully detected), and the screens refresh rate, which might be set at 75Hz or 80 Hz. Frames are repeated in unsynchronized patterns, and in some systems, frames are updated when being displayed, resulting in "image tearing".

The best quality on a display board will be achieved by fully synchronized refresh above 70 Hz (such as 72 or 120 Hz), a 10-bit DAC, and a one-to-one resolution correspondence between the decoded pixels and the displayed pixels.

Flat Panels

Flat panel displays are rapidly improving, but at present they typically only have 8-bits or less of dynamic range. The larger flat panels often have only 6 or 7 bits of dynamic scale. Worse yet, some of these displays allocated the bit codes in a non-uniform way, yielding larger steps for dark and mid tones than for bright pixel values.

Insufficient dynamic range bits will result in visible contour banding when presenting low-noise images. Noise can be added to remove the contour bands, but the result is a noisy presented image. Thus, there is no reasonable substitute for having sufficient bits in high quality display presentation.

Large Screen HDTV Sets

With HDTV large-screen TV's, it is quite difficult to determine the supported scan formats. Some operate at 1080 interlaced lines, some at a range of 480 to 540 progressive lines, and some operate at 720 progressive lines. The information about this critical display system format attribute is effectively being hidden by the TV marketing strategies. Without this information, the usefulness of a given model of HDTV large-screen TV becomes difficult to determine. Further, the demonstration images used are often selected such that interlaced artifacts are not apparent, hiding the interlace artifacts which would occur on a broader range of images. It is also difficult to judge an HDTV set when the source material format is not identified. In the US, both 1080 interlace and 720 progressive formats are supported in the HDTV system selected by the FCC. Without knowing which format is being used, and with the images being selected to remove scenes where interlace is visible, it becomes very difficult to interpret the quality of the image on a given display system.

Projection Systems

The T.I. Digital Cinema Mircromirror Projector is the current front-runner in projected image quality at high brightness. This system supports 10-bit dynamic range, a wide color gamut, and has a resolution of 1280 wide by 1024 tall. Of particular note is that the pixel path is entirely digital from the input of the projector through all the internal processing, all the way to the light, which is digitally modulated by the mirrors. The resolution of 1280 x 1024 is quite useful, but it is anticipated that higher resolution systems will also be developed in the future using this or other technologies.

Other systems, such as the NTT and Kodak experimental systems using the JVC DILA, are exploring increasing resolution, with recent experimental demonstrations showing in excess of 3MPixels.

Of special interest with the T.I. Micromirror projector is the ability to present high frame rate images, well in excess of 24fps. Since this system does not flicker, a 720line non-interlaced image at 60fps is quite pleasing to the eye, while providing for smooth rendition of fast motion.

Compression

The most efficient compression uses motion compensation, where the similarity between adjacent frames is used to help reduce the bit rate required. An example is mpeg-2 which is in wide use for digital television and DVD. Once the motion-compensation is applied by using motion vector references to nearby frames, a correction is coded. In the case of mpeg-2, this correction is a quantized discrete cosine transform (QDCT).

However, in mpeg-2, both the motion compensation and QDCT are performed only to 8-bit precision. Further, the quantization used in the QDCT yields effectively even lower precision. Thus, the effective dynamic range and noise floor of typical mpeg-2 systems is less than 8-bits. The difference between the original image, which may be 8 or 10-bits, and the decompressed mpeg-2, which is less than 8-bits, is due to the errors in motion compensation and the QDCT. The QDCT errors, in particular (known collectively as "quantization errors") create visible patterns and noise on the original image, resulting in substantial degradation to the original 8 or 10-bit picture quality.

Compression can utilize higher precision in both the motion compensation and the QDCT steps. Compression could utilize 10 bits or more for these steps (such as in the compression system developed by the author and his colleagues). This allows a match between the dynamic range of the best high definition cameras, displays and projectors, to the compression system (which are now up to the level of 10-bits). Since compression can support even high bit resolution, up to 12-bits, compression technology is capable of supporting future cameras, displays, and projectors having even wider dynamic range than is presently available.

Layered Compression

Layered compression utilizes spatial resolution layers to provide multiple decoding resolutions from a common compressed signal. In addition, layering the compression has the benefit of reducing the computation necessary to decode the lower resolution base layer, in comparison to decoding the full resolution.

It is common practices in QDCT compression to utilize a quantization "tilt" matrix, which reduces the number of levels used for high frequency details. It is a general property of compression systems such as mpeg-2 that low amplitude details are lost in the compression, due to quantization, and also related to the use of the tilt matrix. However, this is highly undesirable for quality preservation of details. Low amplitude details, such as skin texture, can be crucial to medical applications. Wall texture, carpet detail, and grass detail, are also commonly lost with mpeg-2 compression.

This it is desirable to structure compression so that low amplitude details in an image are preserved. It is necessary to apply these criteria both to single layer coding, as well as layered coding, in order to gain ability to preserve full detail. In higher spatial resolution layers, the low amplitude details can be boosted to offset any loss from compression. In single layer compression, the careful use of spatial detail preservation in processing is required.

In addition to spatial resolution layer, signal quality layering can also be employed (also known as signal-to-noise, or SNR, scalability). With such layering, arbitrary layers of accuracy can be added until a desired level of truth to the original data is achieved. This feature can be highly valuable to both automated analysis and visual scrutiny.

Another dimension of layering is time. It is possible to ignore bi-directionally-predicted (B) frames in mpeg-2-style compression, to support a reduced computational load in decoding and a reduced bit rate in transmission. In this way, B frames can add additional temporal detail to motion by allowing the addition of layers of temporal samples. For example, 18 frames per second can be utilized for predicted (P) frames, with three intervening B frames. Such a structure could be sampled at 18, 36, or 72 frames per second, utilizing no B frames, one third of the B frames, and all B frames, respectively. Another such structure utilizes P frames every 12 fps, such that 5 B frames are employed to yield 72fps. This structure allows 12, 24, and 36fps as well as 72fps decoding of the original 72fps coded pictures.

Resolution Quality Criteria

Visual acuity of a person with normal 20-20 vision is about 100 pixels (50 cycles) per degree of subtended view. This corresponds to 1000 pixels per ten degrees. A picture will appear sharp when there are at least 30 pixels (15 cycles) per degree. Below 20 pixels (10 cycles) per degree, the picture will appear fuzzy and soft, being slightly out of focus to the viewer.

Thus, a typical computer screen now operates at somewhere between 1000 and 1500 pixels horizontally, by about 750 to 1200 resolution vertically, corresponding to a range of horizontal viewing angle between about 35 and 50 degrees.

Viewing Criteria and Bits Per Pixel Quality

Current state of the art supports 10-bits of dynamic range in cameras, and displays, and can be maintained through the entire image chain, including compression. This corresponds to a viewing dynamic range of about 300:1, with a tonal distinction of about 0.6%. However, human color vision can span about 10,000:1 in dynamic range, with 0.25% tonal distinction. This corresponds to about 12 bits. No current electronic moving image cameras or displays yet support this dynamic range. However, improvements to the dynamic range and noise floor can be anticipated in the next decade. If cameras become available with such extended range, the extended image range can be utilized on lower range displays and projectors through the use of brightness adjustment to select between shadow and highlight detail viewing. This is similar to motion picture film negative, which supports several times the range of brightness capture than that utilized by the corresponding print from that negative. This adjustment feature is a key attribute of film negative, allowing brightness and color adjustment after the fact.

On high brightness displays and projections (100 fl and above), tonal distinction becomes greatly expanded, providing substantial additional information and detail within the image. For example, the amount of rain in a scene looking at an object through the rain can be best estimated if the scene is viewed on a high brightness display with high tonal distinction.

Typical 10-bit viewing works best if the display has at least 300:1 dynamic (white to black) range, and is operated at a 20fl white level.

Standardization Opportunities

Most of the issues discussed here represent opportunities for standardization. When digital moving images are viewed from a system-chain perspective, it becomes evident that all links in the chain must have consistent quality in order to provide the desired quality of the final resulting image for viewing or analysis.

Conclusion

The technical ingredients of greatly improved moving image system chains are now within reach. When selected and managed efforts to establish such chains can yield significant near-term benefit to many common needs for the use of the moving image.

<end>