HARRIS
assuredcommunications

WHITE PAPER

General-Purpose 3D Video Signal Processing

By Paul Briscoe, Principal Engineer
Harris Corporation — Toronto, Ontario

Abstract

3D has been thrust upon the TV industry. This has been enabled, and in fact trivialized, by the fact that handling stereo 3D in side-by side and other "frame-compatible" formats makes it inherently compatible with 2D HD systems. While this has allowed 3D to move forward without the huge capital investments usually associated with a major technology step, many key pieces of the puzzle remain unaddressed. One overlooked class of equipment is general-purpose signal processing, the stuff that "lines up" the signal for use, as well as devices that up/down/cross-convert between formats. These devices, designed for a single, full image in an HD raster, perform full-frame processing. This provides a basic utility for 3D signal correction, but adjustments will impact both pictures, and some adjustments (geometry, etc.) will destroy content. This paper discusses the handling and adjustment of signals in the frame-compatible format, and provides guidance as to the features required for various types of 3D processing in the 2D domain. Also discussed will be the necessary features for handling images in a true 3D manner while in frame-compatible format, as well as the functionality needed to get to/from these formats and dual (L/R) stream signals. Finally, practical 3D troubleshooting techniques are discussed.

Introduction

Overview of 3D Signals

Shooting, producing and broadcasting in 3D bring a new set of things to monitor and adjust, and new ways for things to go wrong. From camera to viewer, maintaining the relationship between the Left and Right Eye images throughout the system is paramount, and because of the sensitivities of the viewer to 3D errors, fixing things when they go awry is more important than ever.

Today's 3D material exists as a pair of images, one from the viewpoint of each eye of a human viewer. Whether captured by a 3D camera rig or synthesized in graphics, the key to 3D perception is the horizontal disparity between the images. When delivered to a viewer's left and right eyes, correctly captured and delivered images will appear in natural 3D with no significant discomfort. The key here is the term "correctly" – this implies that not only is the image captured with disparity appropriate for the interocular separation of the human viewer, but also that it is processed and delivered so that the images, except for their unique POV-induced disparity, are otherwise identical in every way. Since 3D cameras consist of two separate imaging devices, and camera rigs exist in a dizzying variety, this is not a trivial task.

3D signals exist in a number of formats, all of which may require some degree of parametric adjustment. For acquisition and production, discrete left and right eye signals are used. These consist of full-resolution images delivered as either a pair of 1.5 Gb/s HD-SDI signals, or a single 3 Gb/s Level B HD-SDI signal. In both cases, every pixel of the source image is conveyed without any image compression. For storage, distribution and comsumption, it is common to compress the image into a frame-compatible format, where the L and R images are each squeezed 50 percent horizontally or vertically, and placed side-by-side or one-above-the-other, respectively. The resultant full-frame HD image may be conveyed, processed and stored in a regular (2D) 1.5 Gb/s HD-SDI infrastructure. These Frame Compatible (FC) formats are known as Side-by-Side (S-S) and Top- Bottom (T-B). While other methods of coding exist, this discussion will focus on these two primary methods. Because this signal can traverse the entire broadcast chain right to the viewer's set, it offers a simple way to convey 3D to the consumer. The responsibility of reconstructing full-screen images rests with the display device (or set-top box).

Viewer sensitivity to 3D is a double-edged sword – there are areas of extreme sensitivity, and other areas where 3D actually helps conceal errors. Probably the most significant human sensitivity yet known in media comes with 3D, and that's the potential for discomfort due to the Human Visual System (HVS) processing (or mis-processing) of depth and motion. Too much peak-to-peak depth in the image results in an inability to keep it all converged. Objects moving in and out rapidly and/or simultaneous in/out motion can break and prevent reconvergence, and some whole-image motion (combined with the rest) can induce vertigo and motion sickness. The 3D look is in the hands of the director, and whether done well or poorly, technical systems must provide artistic transparency. Although unforgiving when it comes to the 3D-specific aspects of the content, there are also some benefits to having two eyes looking at two slightly different pictures. Consider noise: Gaussian pixel noise may be seen in lower-quality images or low-light environments, and it can be particularly noticeable in dark/flat areas of the image. Fortunately, the noise from the camera is random in nature, and thus the noise from two cameras, while of similar amplitude and spectrum, is uncorrelated. When either image is viewed on its own, the presence and effect of noise may be obvious; however, when each eye views its own image, the HVS tends to average out the two, resulting in a perceived reduction of noise. Various amplitude aspects of the 3D signal exhibit this improved tolerance, as opposed to spatial aspects of the signal, which result in increased sensitivity and reduced tolerance.

Processing Devices

General-purpose processing devices are used throughout the 2D signal infrastructure, providing things such as frame synchronization, signal "proc" functions, gamma and color correction, up/down/cross conversion and image scaling. Although they traditionally have no awareness of 3D, they can provide utility in the 3D workflow. Dual-channel devices can also accommodate discrete 3D signals and are evolving to include 3D toolsets.

Single-Channel Processors

A single-channel processor may be used in one of two ways in a 3D application. In frame-compatible systems, the device can be used to repair overall errors that are common to both eyes, and to make tradeoffs between magnitude/impact of errors on one eye vs. the other. These differential errors will require visual QC to establish the best compromise, while absolute errors can be corrected readily.

The other application for a single-channel processor is in one path (or the other) of a discrete system. In this case, it's used to match the channel to the other. This works well if the "other" channel is correct; however, where for example, limited camera adjustment is available, this may be the only way to match the images reasonably. Should an absolute error remain (the "other" camera was wrong), it can be fixed later when processing for both channels is available, or when in a frame-compatible format. This scenario can also be used where camera rig geometry requires one channel to be flipped in either/both axes.

All of these methods are a compromise, not providing full utility, but offering the ability to make a degree of correction, which may make the difference between usable content and not.

Dual-Channel Processors

A pair of channels in a single device finds their utility in discrete systems, providing full control over each channel. Both independent as well as linked controls may be provided for simple differential and absolute adjustment of the images. Such devices can allow full camera rig utility, enabling correction of any optical inversions, upside-down camera mounting and adjustment (careful!) of interocular distance. These devices may also support 3 Gb/s, and thus be able to handle conversion to/from that format and discrete streams of 1.5 Gb/s, in addition to processing the channels independently. It's important to ensure that in 3 Gb/s mode, the device has the capability of full Level B operation.

3D-Enabled Dual-Channel Processors

These devices provide additional 3D processing capabilities including combining/decombining of frame-compatible formats, allowing conversion between the different FC formats and discrete channel processing. Such devices do the scaling required to shrink/expand the images on the appropriate axis, as well as offsetting the positions such that when combined, they are correctly aligned in the FC format. Conversely, when fed a frame-compatible signal, they will offset back to center and upscale each channel back to a fullscreen discrete HD image for output as two 1.5 Gb/s signals, or 3 Gb/s Level A. Typically full proc and color correction functionality is provided per channel, and frame synchronization may be slaved to ensure that the channels drop/repeat on the same frame.

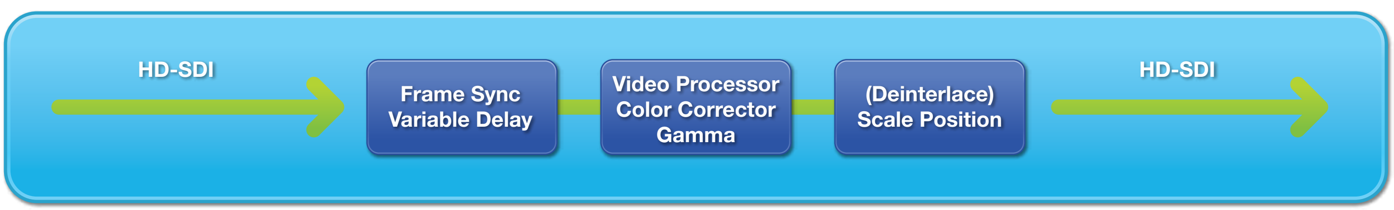

Figure 1 — Single-Channel Processor

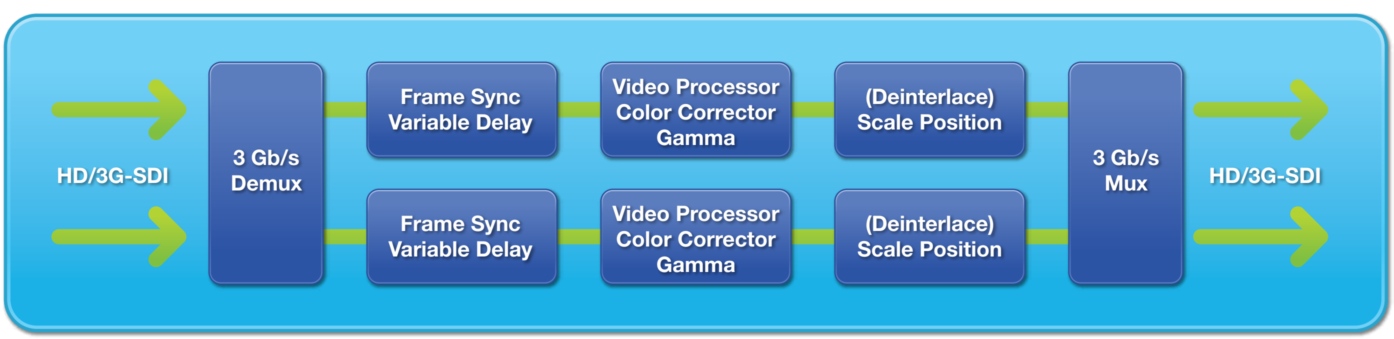

Figure 2 — Dual-Channel Processor

Specific Issues

Picture Alignment and Geometry

Vertical alignment of the L and R images is critical in 3D. There is no depth information carried vertically, and any misalignment of the images in this axis will result in an obvious error for which the viewer can't correct. A small (one or two pixel) error may only produce eye strain and discomfort; however, much more and the difference will become annoying. In the transition region between invisible discomfort and obvious error, the viewer may experience the error snapping in and out, depending on content and action.

Horizontal misalignment, on the other hand, directly impacts the 3D experience. Taking a correctly depthadjusted 3D scene and horizontally displacing one of the eye images serves to move the depth field in and out of the display. The result can be to push the scene back into the display with image separation at the back, or perceptually even worse, bring it out of the display toward the viewer with the obvious problems that causes. Any unintended misalignment of the image in either axis will create 3D fidelity problems that at best will degrade the content, and at worst, destroy the 3D experience. These errors can be spotted with visual QC, but cannot be resolved with 2D processing in the frame-compatible domain. In the discrete 3D domain, picture positioning controls can be used with great care to move one (or both) images with respect to each other to reduce or eliminate the error. While repairing the 3D, moving the images will result in cropping and gaps at the edges. Small (one or two) pixels/lines of offset will probably go unnoticed; however, much more and there will be discernable errors on the gap edge. Additionally, should the gap occur on the image side that becomes adjacent when in a frame-compatible format, the resultant transition will create transients that will stimulate bad behaviour in downstream equipment.

Image geometry problems mostly originate at acquisition due to mechanical/optical alignment issues in the camera rig. These errors can include rotation, amd mismatched zoom and keystoning, as well as vertical and horizontal misalignment. The best way to deal with geometry issues is to fix them correctly at the source, but it's possible to be stuck with a misaligned rig or a recording from same. Other than orthogonal image offsets, downstream repair of geometry errors is mostly beyond the capability of in-stream processors at this time, and may be limited in the future due to the complexity of processing required and resultant loss of quality.

Wrong Convergence/Divergence

At acquisition, one is counting on all things being set up correctly to allow the stereographer to determine and set convergence and depth on an ongoing basis. Misalignments anywhere in the system, from camera to operator display, can result in the wrong decision being made on the fly, the result being the wrong convergence and depth in the recording. Using image offset to push or pull the depth to the correct amount may be possible if the error is small, but larger convergence errors are beyond today's inline processing. As mentioned above, any image offset will result in a crop and gap occurring, with the gap causing the most opportunity for downstream trouble. Here, in the discrete acquisition environment, a 2D processor in one path can accommodate this adjustment. Although this is absolutely the wrong way to deal with convergence errors, if it's already wrong, it may offer some improvement.

Temporal Synchronization

It's vitally important that the L and R eye images remain in temporal sync at all times. As annoying a problem as lipsync errors, L/R desynchronization results in live action taking on a strange "pumping in and out" effect, as the combination of scenic action and camera motion conspire to modulate all of the 3D information to varying degrees. While a visually complex-looking problem, it can be solved by simply synchronizing the channels. If already locked into frame-compatible format the image will have to be decombined, synchronized, and recombined. This processing is best done in the downscaled image format to avoid the loss of quality and introduction of artifacts that additional up and re-downscaling would cause. A strictly 2D processor cannot provide this functionality; however, such processing can be done in a two-channel device.

Eye Swap and Flip

It's entirely possible to switch the eyes in the course of moving signals and files around. Such an error doesn't look obviously wrong until it's viewed in 3D, and may escape casual QA in a 2D domain. Such an error results in an unwatchable picture; despite this, the author has seen many people struggle (and some actually think they see 3D) when trying to watch a swapped picture. A 2D processor cannot swap the eyes in a frame-compatible environment; however, this could be done in a frame-compatible 3D-enabled 2D device, which can reposition the image halves without scaling. Some camera rigs may deliver one or both of the images reversed in either or both of the H and V axes. This is the result of the various layouts of 3D camera rigs. Flipping the images as required by the rig is necessary to determine and adjust the 3D aspects of the image. A 3D processor can, of course, perform independent channel manipulations; however, a 2D device can be used in one channel's path to transform that image to match the other.

Color Correction/Levels

Because 3D is being captured with two different (no matter how identical) cameras, there is the opportunity for colorimetry differences between them, in addition to the usual setup adjustments such as black and white levels, saturation, hue and gamma. Even with all-digital cameras, the analog nature of the optical sensor block is such that matching calibrations need to be done precisely, particularly if establishing and maintaining a specific "look" is desired.

Two types of color correction need to be considered, absolute and L/R differential. Absolute color correction is the normal process used to adjust colors to the correct values on a monitoring device, and compensates for not only the camera but also for the lighting and scenic composition. At acquisition, it is important to capture precisely so that fidelity is preserved and so that later in the workflow, a colorist has the most range of control; after all, their job is less about fixing bad color and more about enhancing it perceptually. Differential matching is all about removing differences between the cameras. Although calibrating each independently against a reference should establish matching, under careful comparative scrutiny, even finer tuning can be accomplished.

Where the images match each other but are incorrect against an absolute reference, it's a matter of applying the same correction equally to both. This is, of course, a natural in the frame-compatible domain, as all amplitude adjustments to the picture are applied to both. In the discrete signal world, two channels of tracking correction are required. In the case where one camera is correct, it's necessary to disassemble the signal into discrete channels to fix the error and use independent channel processing. Once in the FC format, the signal can be processed for conventional analog parametric adjustments just like a 2D signal, since these are common to both images.

Scaling and Deinterlacing
Inherent in the coding from discrete eye signals to frame compatible is the need to scale the picture in one or the other axis. Such scaling is performed with digital filters, which decimate the pixel count by half in the appropriate axis prior to combining the images. Although these filters can deliver very high-quality scaling, it must be remembered that half of the pixels are being removed in the process. When the picture is rescaled back to full size, similar filters reconstruct the missing pixels, but the lost bandwidth isn't recovered. While a single pass of frame-compatible coding doesn't grossly reduce the obvious visual bandwidth of an image if delivered to and re-expanded on a final display device, a second pass of the same will quickly degrade detail in that axis. Concatenating one of each type of system will result in perception of an noticeably softer picture due to the loss of sharpness cues from the unaltered axis.

Here the two-eyes advantage helps as well; 3D tends to impart a perceived sense of added detail to the image (purely by HVS trickery), which appears to offset the loss of resolution to a degree. Going beyond one FC pass, however, definitely will produce obvious loss of detail.

When delivered as a frame-compatible interlaced picture, deinterlacing becomes important, particularly in Top/Bottom formats. The deinterlacer's job is to try to guess the pixel values on the missing lines of each field, based on prior (and sometimes next) field data. Where the horizontal split occurs in the image, the deinterlacer will find unrelated, adjacent field and line data, and the corresponding interpolated pixels will be incorrect. Although this error is only a few vertical lines in magnitude, it can propogate downstream, where it can be further amplified by compression, scaling, etc. This is one of the primary reasons T/B format is used with progressive signals, and L/R is used for interlace. Such errors on the split will, of course, manifest themselves in the viewing environment as artifacts at the top (in one eye) and bottom (in the other eye) of the image.

Cross-conversion of frame-compatible signals is a particularly signal-degrading process. Using the example of 1080i to 720p, the 30i interlaced signal in S-S format is deinterlaced to a full 1080p image, and then the two channels are extracted and horizontally expanded to derive two 1080p images, now reduced in H bandwidth. These are then scaled to 720p, requiring H decimation from 1920 to 1280 pixels, and V decimation not only from 1080 to 720, but also further to 360 pixels (lines). These two images are placed one above the other for T/B format. This signal has had horizontal resolution reduced, and was deinterlaced and downscaled vertically. The resultant image has been equally degraded in both axes, and has had deinterlacing artifacts added in the middle for good measure. It can be seen why it's best to derive frame-compatible signals from discrete channels and, once in compatible format, is best left there.

Other Processing
One particularly dangerous functionality is noise reduction. This is something that is often left on the "low" setting in 2D simply because it doesn't seem to degrade a good picture, and appears to help a noisy one. Noise reduction can indeed improve a 2D picture, but it must be used with great caution with 3D FC signals. NR works by identifying noise pixels separately from the content and processing those pixels in a unique manner, usually interpolating them from adjacent quiet ones. This, of course, causes reduction in resolution, which can't be tolerated in FC formats due to the already-scaled images. Only consider use of 2D noise reduction where the loss of resolution is a better tradeoff than the noise. Also be careful of NR modes; many devices have smart adaptivity, which will misbehave at the image split. Take care also where NR is buried in a device, such as an MPEG encoder. Verify that it is set to an aggression level that does not mishandle the content. Of course, NR in discrete systems doesn't differ from 2D NR, with the caution that it needs to be applied equally to both channels.

MPEG Systems
MPEG transport systems will do an excellent job of transporting frame-compatible signals for distribution without significant issues. One aspect that requires awareness is related to the L/R split in the picture, whether horizontal or vertical. Because MPEG codes the image as though it's all one coherent scene, errors can occur at the split which, upon 3D display, can cause one-eye artifacts on the sides or top and bottom of the picture. These errors will generally be minor; however, concatenating compression systems can result in the magnification of this type of artifact. It can readily be seen in the frame-compatible image by examining the split for obvious compression-type errors.

Monitoring
While 2D T&M gear natively does not provide 3D capabilities, some degree of measurement can be made. Amplitude parameters can be examined despite the scaling of the picture; one simply has to look at the lefthand side of the waveform in S-S mode and the top half of the waveform in T-B to see one eye, and vice versa. Levels and colorimetry can be reasonably measured in this manner. Specific frame-compatible issues, such as edge/split gaps and artifacts, can be examined, although not with total visibility, as even the re-expansion at a display immediately downstream will introduce artifacts as a result of those in the compatible domain. As 3D FC gains traction, many 2D T&M manufacturers are adding special display and measurement modes to enable further probing of the FC signal, and simplified access to the measurements noted above. In this way, many traditional 2D products are emerging with useful 3D FC feature sets.

Conclusion

Conventional 2D processing devices can provide good utility in 3D environments. Frame-compatible signals can be normalized for use in the same way 2D signals can, and some 3D issues can be minimized. A two-channel device for use in a dual-stream environment provides much greater utility for 3D signal correction, and when fitted with 3 Gb/s interfaces, can offer conversion between single- and dual-link discrete formats. When combined with internal scaling and combining capabilities, any-to-any format conversion and processing are available. Trick functionality such as H/V flips and time-aligned frame synch round out the utility of the device. In particular, the ability to perform processing in the frame-compatible domain without the need of scaling up and back down means quality can be retained. Within a 2D plant infrastructure, use of these devices allows total processing control over frame-compatible signals in a manner familiar to operations and conventional workflows, and can provide the control needed to comfortably live with 3D in the 2D environment.

This paper was originally published in the 2011 NAB Broadcast Engineering Conference Proceedings.

For more information, please visit broadcast.harris.com.

HARRIS®
assuredcommunications®

Broadcast Communications Division
9800 South Meridian Boulevard, Suite 300 | Englewood, CO USA 80112 | Tel: +1 303 476 5000
broadcast.harris.com