grass valley

WHITEPAPER

3D Television

April 2010

Many existing Grass Valley products are easily adapted to 3D operation with minimal modifications. Thus production companies can, without heavily investing in new 3D-specific technologies, develop practical skills and experiences which will enrich and empower them as the market for 3D evolves in the coming months and years.

Contents

Introduction 1

What is Stereoscopic 3D? 2

Cameras 7

Signal Routing and Production Switchers 9

Servers 9

Contribution and Distribution 10

Grass Valley's 3D Workflow 11

Conclusion 11

Introduction

Digital projection has enabled an old idea—stereoscopic 3D movies—to give the movies a new lease on life. *Hannah Montana and Miley Cyrus: Best of Both Worlds Concert* took the box office record for a Super Bowl weekend opening in 2008, and more recently *Avatar* has become the first movie to exceed $1 billion USD at the box office, both due to being 3D productions.

Seeing the success in theaters, television content producers and distributors have expressed interest in establishing their own stereoscopic 3D services. Channel launches for 2010 have been announced in both North America and Europe.

Consumer electronics manufacturers are excited by the prospect as they see a way of revitalizing television receiver sales, which are currently depressed by the falling prices of flat panel displays and a drop in demand following the initial surge. In the US, consumer electronics manufacturers are sponsoring some of the proposed channel launches.

There are both technical and creative reasons why 3D to the home is very different than 3D in the movie theater. This paper does not attempt to enter this debate (although some of the technical issues will be touched on for context). It looks instead at how today's technologies can be used and adapted for 3D production, and what special considerations need to be made.

Figure 1 – *Cinema and gaming are driving 3D today; will live 3D sports be the next big winner?*

What is Stereoscopic 3D?

When looking at real world objects, our eyes and brains use two sets of visual information to determine how far away an object lies. Convergence is the angle our eyes have to diverge from the parallel to bring the two images in line, and accommodation is the degree to which eye muscles have to change to bring the object into focus.

In other words, for an object that is close to us, we converge our eyes by turning them in, and we accommodate a very close focus. For an object that is far away the difference in accommodation—the way the lens in the eye is modified—changes, and as the distance tends to infinity so the convergence tends to be parallel.

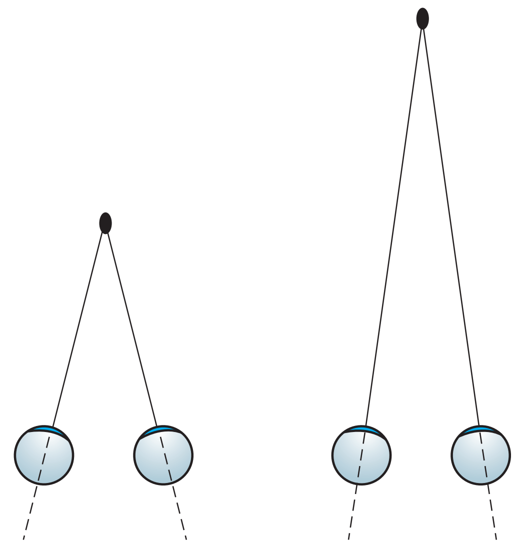


Figure 2 – *Convergence. (Copyright Lenny Lipton)*

Note that in the real world there is never a need for convergence to become divergence: our eyes never point outside of parallel, a phenomenon that is extremely discomforting and one that must be avoided at all costs in the display of 3D images.

To these muscular cues our brain adds other factors, including parallax, the way that nearer objects mask further ones, and experience: we know roughly how big a truck is, so if it appears tiny it is probably far away.

In stereoscopic 3D we attempt to present this depth information on a single screen, by presenting subtly different views of the scene to the left and right eye exclusively: hence the name stereoscopic.

There are a number of ways of ensuring that the left and right images reach the left and right eyes. The traditional colored anaglyph glasses filter the view visible to each eye but this technique is now largely abandoned for high-quality 3D presentations due to better technologies that provides a more consistent and appealing 3D presentation to the viewer. Modern systems use either polarization along with an on-screen polarizing overlay or active shuttering using LCD glasses that run each eye's view on and off in sync with the on screen presentation.

Figure 3 – *Simulation of a polarized 3D image viewed without polarized 3D glasses.*

Figure 4 – *Polarized 3D glasses (photo courtesy of REALD).*

What is Stereoscopic 3D? (Cont.)

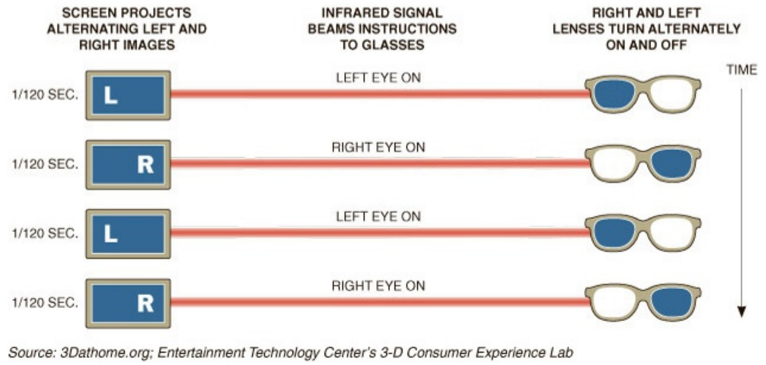


Figure 5 – *Active shutter system showing display and glasses corresponding for proper sequencing of left eye and right eye viewing.*

Figure 6 – *Active Shutter Glasses*

It should be noted that polarized or LCD shutter glasses will reduce the light transmitted to the eyes, and this is an issue that can affect creative and production choices. This will be discussed later.

The key dimension to keep in mind when shooting for 3D is the interocular distance, the measurement of how far apart the lenses of the eyes sit. This is generally taken to be 65 mm, although it does vary from person to person, and will be only around 50 mm for children: there is an argument that the entire geometry of 3D should change for children's television.

The fundamental principle of 3D, then, is that if there is a shift between the left and right images of an object—which will cause our convergence to follow it and therefore place it on the Z axis—we will have perception of depth.

If the left image is to the right of the right image, to converge them the eyes will have to converge to a relatively large degree and the object will appear in front of the screen (negative parallax). If left and right images coincide then the image appears on the plane of the screen. And if the left image is to the left of the right image (positive parallax), then the object appears behind the screen.

If an image is to be portrayed a long way in front of the screen (extreme negative parallax)—the classic 3D effect of an object leaping out at you—then it will cause the eyes to converge a great deal (making the observer appear cross-eyed) which will be extremely uncomfortable for any length of time.

Conversely, if an object behind the plane of the screen requires positive parallax of more than the interocular distance (65 mm, or maybe 50 mm for children) then you will be forcing the eyes to separate which is extremely unnatural and may be one of the principle sources of nausea in 3D presentations.

What is Stereoscopic 3D? (Cont.)

POSITIVE PARALLAX

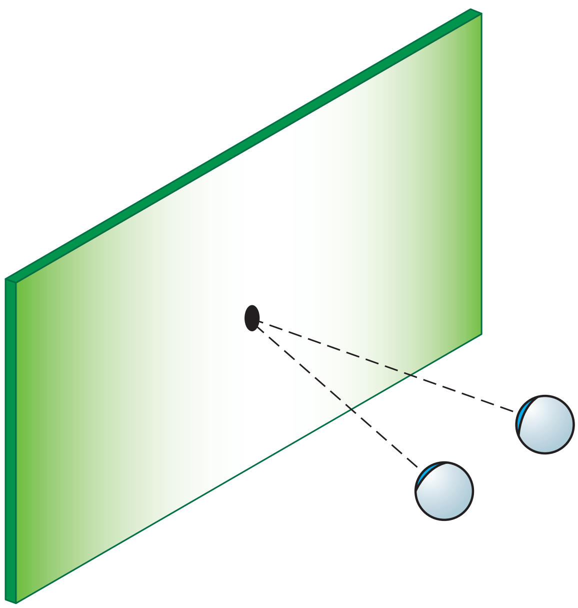


Figure 7 – *Object at plane of screen.*

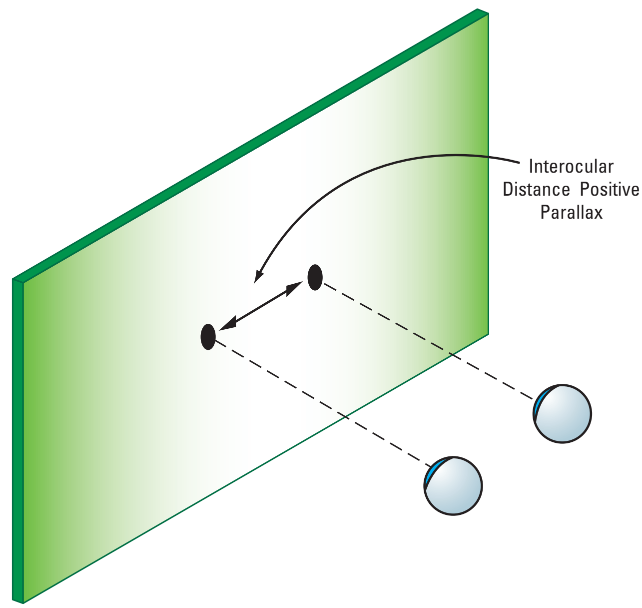


Figure 8 – *Object at infinity*

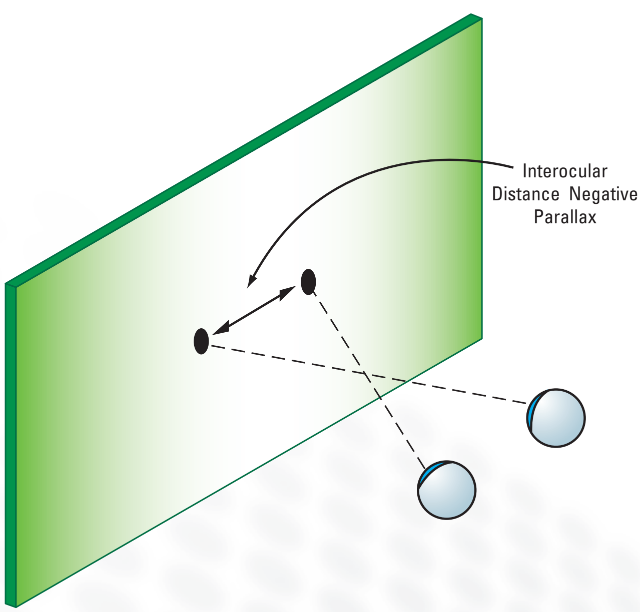


Figure 9 – *Object halfway between screen and spectator.*

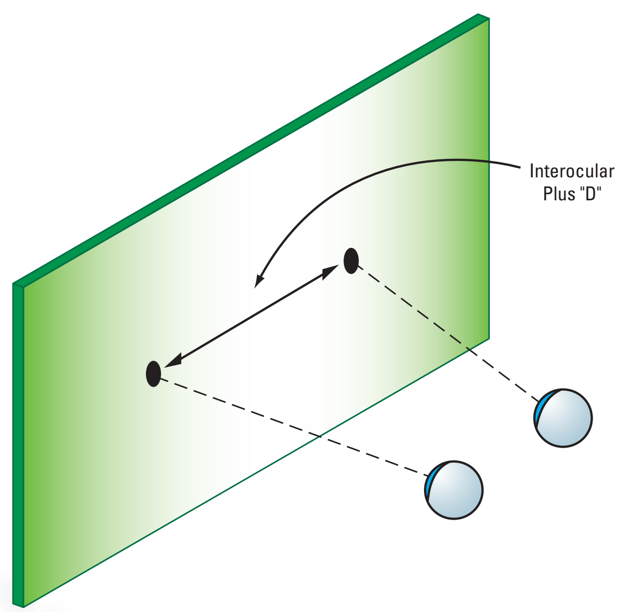


Figure 10 – *Divergence.*

(Drawings Copyright Lenny Lipton)

What is Stereoscopic 3D? (Cont.)

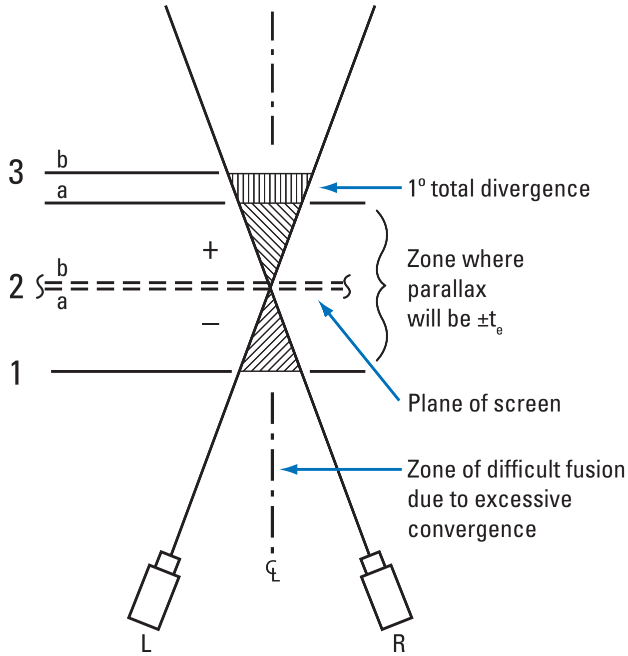


Figure 11 – *Zones of screen parallax. (Copyright Lenny Lipton)*

Note that this limit is an absolute measurement on the screen, whatever the size of the screen and whatever the viewing distance. With the viewing distance in the home typically a fraction of that in the movie theater, there is a greatly increased risk of causing uncomfortable divergence.

The shorter viewing distance in the home also increases the angle of convergence just to focus on the screen: typically it is 0.8˚ at home, 0.2˚ in the movie theater. That, in turn, limits the amount of negative parallax that can be added, so restricts how far in front of the screen objects can appear.

According to the work of David Seigle and John Sanders of In-Three in association with UC Berkeley[1], the director is free to place objects anywhere from infinity behind the screen to 2.4m (8 ft.) in front of it for movie theater exhibition. In the home, this is restricted to 1.2m (4 ft.) behind the screen to 0.5m (1.8 ft.) in front, giving a more compressed view of reality.

[1]www.in-three.com/3DintheHomev2.html

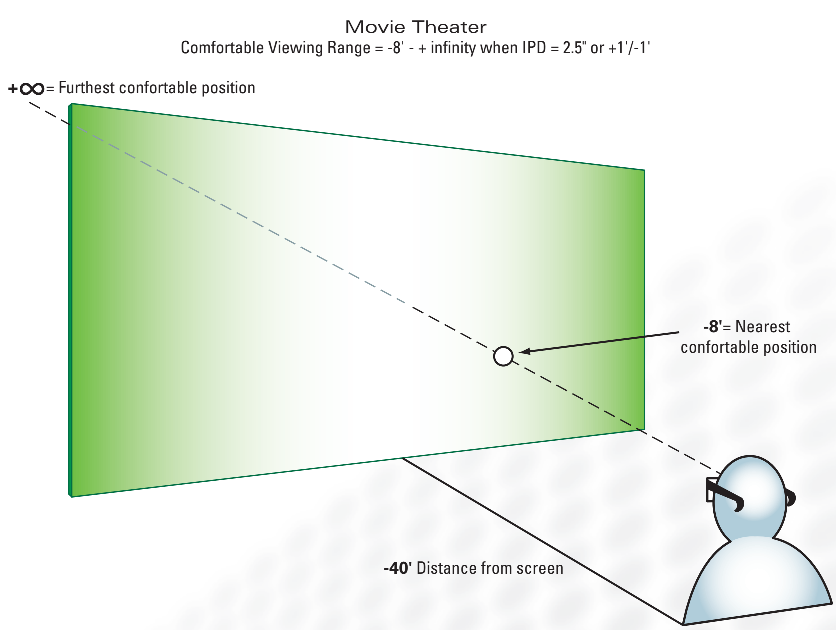


Figure 12 – *With a typical larger screen, controlled environment, and seating placement, it is possible to have a wider latitude with 3D in the theater. (Drawing Courtesy: In-Three)*

What is Stereoscopic 3D? (Cont.)

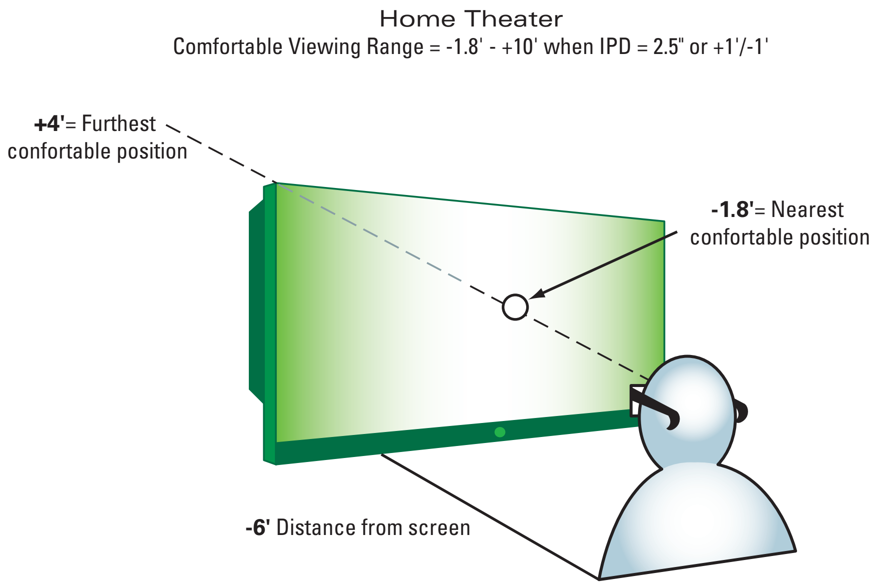


Figure 13 – *Because of a wider range of screen sizes, types, and viewing environment, proper 3D production for the home is more challenging. (Drawing Courtesy: In-Three)*

The final point to note is that stereoscopic 3D depends on convincing our brains to rely entirely on convergence cues and ignore accommodation altogether: although our eyes are moving to bring the two images together, the lenses in the eyes do not need to refocus because everything is actually at the plane of the screen.

In the movie theater this is not a problem because the room is dark and there is nothing else to look at, so the eyes naturally remain on the screen. In the home, however, the light level is higher, meaning that other objects at other distances become visible, so there is a risk of conflict between accommodation and convergence.

The only solution to this, it would appear, is to make the content so compelling that 3D TV viewers have no time to look away. Brian Lenz, of BSkyB in the UK, makes the point that the medium is most successful with event television[2]: big sports games, opera, and ballet—immersive experiences that are enhanced by 3D production.

Having established the fundamentals of 3D TV, the remainder of this paper will look at the technical challenges. By far the biggest can be found in acquisition, so the next section considers cameras for 3D.

[2]www.v-net.tv/video.aspx?id=89

Cameras

As has been established, 3D depends on two separate image chains, representing the eyes of the viewer watching the scene. It is logical, then, that the two cameras have to be separated by the interocular distance of 65 mm.

Before looking at the problems this raises, it is worth spending a moment considering what happens if they are set at a different distance apart.

Research suggests that the result is what you might expect: set the cameras further apart and the result looks like it was shot by a giant head so the images look miniaturized; set the optical centers closer than the interocular distance and objects will appear larger than normal.

This effect seems to be linear. A camera lens center distance of 130 mm—twice the interocular distance—will result in people and objects appearing to be half normal size.

Today's broadcast cameras, and in particular high-quality lenses, cannot be set side by side with their optical centers at a 65 mm separation. The optical mechanics of lens design alone prevent this: lenses need to be large to provide consistent resolution across the full image, and to ensure accurate vertical and horizontal tracking when changing focal length. There are many ways to compromise, but each comes with penalties in image performance or creative control of the image.

Large lenses also capture and transmit more light, which is critical for a number of reasons.

First, remember that the viewer will be watching the final result through some sort of filter system, whether it is LCD shutter glasses (which are never completely transparent) or polarized glasses (which reject all light except that in the defined polarity) so there is a loss of light.

We are all familiar with the fact that, as twilight falls, we lose our color resolution. Physiologically, this is because the eye has two types of light-sensitive elements: rods and cones. Rods are very sensitive to low light levels, enabling us to see in the dark to survive: this is called scotopic vision. Our perception of color comes from the cones, which come in three varieties (very broadly corresponding to red, green, and blue), but these are much less sensitive[3].

The cones give us color, and this is called photopic vision, but that needs more light energy. It may be no coincidence that the highly successful Avatar is, in the main, shot in atmospherically dark settings. Sports fans, used to vibrant colors, tend to notice that 3D is more muted.

To explain the second reason, let us consider the alternatives to mounting cameras at 65 mm centers.

One solution is to mount two high-quality broadcast camera/lens systems side by side, then use electronics to reduce their separation to the required width: in effect, use an image processor to move one view across the frame by the required number of pixels. Side-by-side camera rigs, and this process, work best for objects that are greater than 20m (66 ft.) or more away from the camera, whereas mirror rigs are preferred for scenes where the action is playing much closer to the cameras.

[3]www.colourware.co.uk/cpfaq/q2-3.htm

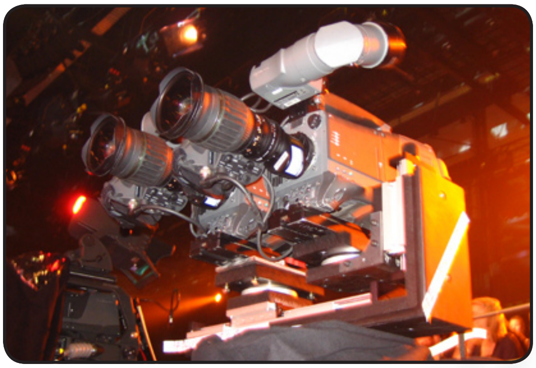


Figure 14 – *Grass Valley LDK 8000 cameras shooting a project in Germany using a side-by-side camera rig.*

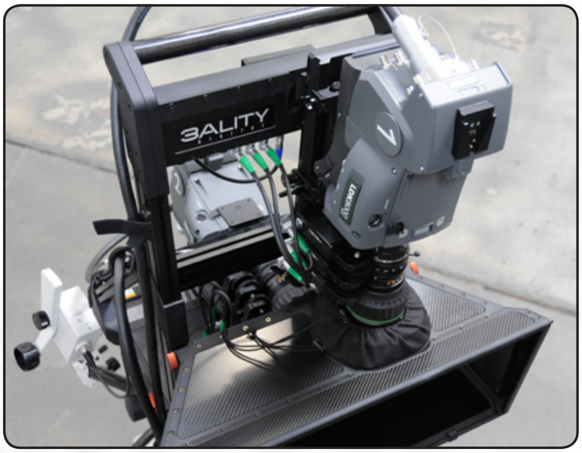


Figure 15 – *LDK 8000 Elite Cameras on a new 3ality mirror system rig (cameras at 90° to each other).*

Cameras (Cont.)

This may well be the long-term solution. However, it does imply another downstream process, which adds to the latency or delay in the system. It also adds to the complexity of production: a recently announced 3D outside broadcast truck for the North American market has six seats for convergence operators—a huge overhead at a time when budgets are ever tighter.

The alternative is to use a mirror rig: an arrangement that uses a half-silvered mirror set at 45˚. One camera points straight ahead through the glass, the other is mounted above/below it a 90° angle pointing downwards/upwards, and sees the mirrored reflection.

This has the advantage that the interocular distance can be set to properly capture the image for each individual shoot, mechanically, then the bolts are tightened with the interocular distance remaining the same throughout the shoot.

The disadvantage is that the half-silvered mirror splits the light in two, so each camera gets half. This is why you need uncompromised lenses: you need to open each by an extra f-stop to compensate for the light lost in the mirror.

That only applies if you have control over the lighting, and can ensure that there is sufficient light available to be able to compensate for an f-stop. In particular, you have to consider depth of field issues. Increasing the aperture in a lens reduces the depth of field proportionally.

This is not an issue in 2D, where the control of depth of field has become part of the creative process. Directors will use depth of field as a tool to focus the audience's attention, whether it is on the critical reactions of the actors in a tense drama or on the skills of a star player in sports. That is part of the visual grammar commonly used in all high-quality cinematography, and that we have all grown to understand and accept.

3D calls for a different visual grammar. If the audience is to perceive and enjoy the 3D stage, it needs to be able to explore it, so much more of the image must be in focus. The risk of the 3D illusion breaking down is even greater if the eye is forced to converge somewhere other than the plane of the screen, only to find that the object at that point cannot be brought into focus. Again, the result for many viewers will be headaches and nausea—certainly not what you want to encourage.

So in general 3D requires smaller apertures and therefore greater depth of field to give more in-focus objects to the viewer. Where the lighting cannot be controlled, or raised to a sufficient level—an outdoor, night-time sports event, for example—then the camera gain itself has to be pushed by +6 dB.

Noise, being random, will be different in each eye, again providing a distraction, so the ideal 3D camera for television will have a very low noise floor, allowing the electronic gain to be pushed in difficult lighting situations. Improvements in the Grass Valley™ LDK 8000 Elite and the LDK 3000 cameras give them a definite advantage of more than 6 dB over previous camera systems.

The other implication of a mirror rig is that, while the signal from the horizontal camera can be used directly, the vertical camera, which sees the scene reflected from the mirror, will need to be flipped in the X, Y, or even both axes depending on the configuration of the rig.

Again, this can of course be achieved as a downstream process, and once set up, need never be adjusted (although accidentally routing around it would be disastrous). But there will be latency introduced if an external image processor is required to be used, which will cause additional synchronization challenges, and keeping the two eyes in sync is enough of a challenge without adding to it.

Grass Valley's LDK 8000 Elite and LDK 3000 cameras include the ability to directly output the image in the usual manner but you can internally select to output an inverted or flipped image as well. Therefore, it makes no difference to the amount of camera latency if the signal is normal, flipped, inverted, or both. This eliminates a source of delay and potential serious errors downstream.

The final issue with a mirror rig is that the two optical paths will inevitably have different characteristics. Even given perfect lenses—which do not exist in the real world—one light path will have been through the half-silvered mirror, the other reflected by it. So there will be chromatic differences between the two which must be corrected.

Once those differences are established, though, the pair of cameras needs to be treated as a single device for subsequent adjustments, for example to track changing light levels in an outdoor environment.

The design of Grass Valley camera control units (CCUs), which communicate via IP over Ethernet, allows a sophisticated master/slave relationship to be established. This ensures that the "differences" between the two camera/lens chains can be matched, then the CCUs locked together under a single operational control panel to allow the pair to be maintained in alignment with the other 3D pairs of cameras on the production.

Signal Routing and Production Switchers

Synchronous switching of routers is a standard part of the Grass Valley Jupiter™ and Encore™ Facility Control Systems product families, so distributing 3D signals as a parallel pair of left eye/right eye digital feeds is not an issue, utilizing Grass Valley Trinix™ NXT or Concerto™ Series systems.

One interchange format that is gathering interest and is a very attractive way of simplifying infrastructures is to multiplex two 720p or 1080i HD signals into one SDI feed. This, of course, doubles the bandwidth, but recent infrastructure products are built with 3 Gb/s bandwidths in mind. In particular, the Trinix NXT router supports long runs of 3 Gb/s circuits (in excess of 100m/328 ft.), which eliminates constraints in building truck layouts and connecting production and server trucks.

At the switcher there will definitely be a change in production style, but in sports there will remain the expectation that multiple levels will be required. While there may be fewer cuts as viewers explore the 3D world themselves, they will still expect to see graphics and statistics to add additional depth to game coverage.

So the first requirement for the production switcher is that it should be able to handle the two left eye/right eye signals as one, under one button, but without reducing its capabilities. Solutions that halve the number of M/E banks available to achieve dual path operations are somewhat limiting.

The Grass Valley Kalypso™ and Kayenne™ switchers were designed for flexible, multi-format production. A function called DoubleTake™ is effectively a split M/E mode, which in this partitioning application uses background and utility busses to carry the two signals in parallel and perform the same effect in synchronization without sacrificing resources. Most other switchers tend to gang M/E banks together for 3D, meaning that a 4 M/E switcher is reduced to 2 M/E when used for 3D productions.

New in the Kayenne switcher is Key Chaining™, a feature that parallels keyer resources and allows setup and control under a single set of buttons to manage keying for 3D graphic elements. Keys may be adjusted separately for accurate convergence. All settings are remembered each time a key is enabled.

The significance of this is that current Grass Valley switchers already installed will perform a great deal of the functionality required for 3D production. In essence, they are 3D-ready today.

Servers

As with the rest of the production chain, the key requirement of a server for 3D is that it should be able to maintain the pair of left eye/right eye files in perfect synchronization, together with the ability, in a live production, to go to a start point and replay the married files instantly.

Grass Valley's advanced server designs already feature the ability to handle video and key channels, but this has been greatly improved with the introduction of the K2 Summit production server and a new feature called ChannelFlex™. This adds the ability to record and play multi-channel feeds and, in the case of 3D, to manage and move the left eye/right eye 3D streams as a single file. ChannelFlex addresses the need for 3D recording and playback without sacrificing channels in the server, so that the K2 Summit server remains capable of playing and recording stereo pairs on each of its four channels.

UK-based OB company Telegenic is very much at the forefront of outside broadcast technology, having recently been commissioned by BSkyB to build its first outside broadcast truck specifically for 3D production.

The truck will be used to produce live and recorded coverage of sports and entertainment events for a dedicated 3D channel to be launched by Sky TV later in 2010.

Telegenic used the K2 Summit to provide 3D replay animations throughout the pilot transmissions and plan to use K2 Summit or K2 Solo servers with the ChannelFlex option as a permanent solution in their 3D trucks going forward.

Contribution and Distribution

All of the above discussion has focused largely on creating the content. We must also address challenges in moving live content over contribution links maintaining perfect synchronization and pristine 10-bit 4:2:2 quality. These needs are met with the introduction of the new VA5004 codec for high-end video contribution over IP networks, which can also be operated in 3D mode. It's a JPEG 2000 encoder using 10-bit 4:2:2 wavelet compression for the highest picture quality. The VA5004 has a 3G interface where the left eye/right eye streams are carried as one signal but compressed as two parallel streams, preserving the full bandwidth of each view. In contribution applications it can be run at rates from 100 to 300 Mb/s and maintain perfect synchronization between the two views.

The final challenge is to take the 3D signal from the venue to the broadcasting center and from the broadcaster to the home. At present, several frame compatible formats that support 3D transport over the existing HD distribution infrastructure have been proposed and are now being examined by standards bodies. The SMPTE task force has made its initial report[4].

The MPEG Industry Forum is actively promoting an extension of the MPEG-4 AVC/H.264 standard, which will essentially transmit one "eye" plus the metadata that will define the differences for the other eye. The estimate is that this will add about 50% to the bandwidth requirement over 2D high definition, or conversely will reduce the bandwidth for a two-channel transmission system by 25%[5].

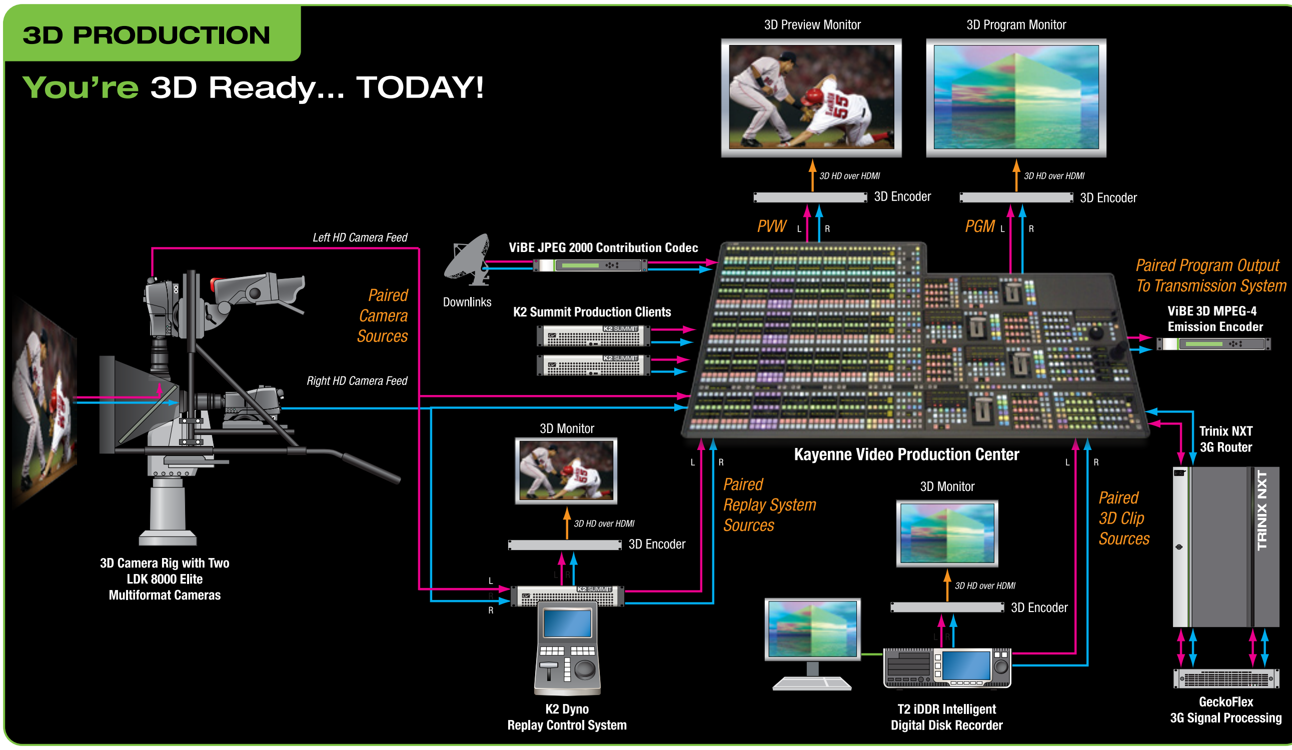
In the meantime, several popular formats have emerged to support 3D's initial market deployment. BSkyB in the UK, for example, is using a system which halves the number of lines in each picture and sends both down a single HD channel with the left eye in the top of the frame and the right eye in the bottom. This has the advantage of being transparent through the entire transmission network, including its installed base of 1.3 million set-top boxes.

Supporting BSkyB's deployment is the ViBE EM3000 encoder which provides compatibility with the Sensio systems for MPEG-4 compressed 3D distribution.

[4] store.smpte.org/ProductDetails.asp?ProductCode=TF3D

[5] www.mpegif.org/m4if/bod/Working%20Groups/3DTV%20Working%20Group%20Index.php

Grass Valley's 3D Workflow



Conclusion

In the last couple of years 3D productions have had an enormous impact on the movie industry, with creative and financial success allied to the technical triumphs. Many of these successful movies have been entirely animated, but increasingly there is a market for live action or mixed environments, with *Avatar* the leading example.

While movies are usually made in very controlled circumstances, where every element can be prepared in advance, there will be some occasions in which television style production techniques will be used. Concert movies like the Hannah Montana special are obvious examples.

In addition, some television content creators and distributors have begun to examine 3D as a new way to entice and retain audiences. While it is too early to say whether consumers will want 3D in the home, with the need to wear glasses and avoid the temptation of real-world 3D objects, it is clear that there will be many opportunities to become involved in trial productions in the near future.

Rather than invest heavily in new, 3D-specific technologies, it is clear that many existing products are easily adapted to 3D operation with minimal modifications. Thus production companies can, without significant capital expenditure, develop practical skills and experience which will enrich and empower them as the market for 3D evolves in the coming months and years.

SALES

Local and regional sales contacts can be found by visiting
www.grassvalley.com/sales

SUPPORT

Local and regional support contacts can be found by visiting
www.grassvalley.com/support