June 20, 2013

Encoded Pan and Tilt Head Considerations for 4K Imaging in an Augmented Reality Environment

A White Paper by Andrew Butler, Senior Product Manager, Vinten.

With all the excitement and momentum toward 4K Ultra-HD production coming out of the 2013 National Association of Broadcasters convention and exhibition, there has been much attention paid to the 4K imagers built into the cameras and the need for high quality lenses capable of delivering enough resolution to project a crisp and error free 4K image onto the camera sensor. There's also the matter of needing adequate bandwidth in cabling and wireless technology to transfer these higher resolution images.

I'd like to discuss an additional need that 4K productions will face in deploying augmented reality: providing enough pan and tilt positional data for the UHD frame at the extreme telephoto end of zoom lenses.

The aim of any augmented reality system, such as those which insert a down-and-distance logo onto the playing field at a sporting event, is to provide a composited image of the real background and virtual graphic that appears to actually exist. To do so requires highly accurate pan and tilt and positional information for the camera. Typically this is provided by an encoded head, such as Vinten's i-Series encoded heads.

How high a resolution does this information need to be? Our goal is to allow a single pixel of the viewing screen to be resolved on the measuring device at the most extreme levels of zoom. Let's do the math on a 1080 high definition production using one of the longer lenses (a Fujinon XA101x8.9 BESM) that a sports production might deploy.

When that lens is zoomed all the way in, its horizontal angle of view might be just 36 minutes, or 1/600th of the 360 degrees the camera head could pan. On a 1080 HD frame, the horizontal pixel count would be 1920. By multiplying the horizontal pixel count of 1920 by 600 we can calculate how many data points would be needed in the head's 360 degree rotation in order to resolve 1920 pixels in the 36 minute of horizontal view, the highest telephoto point of the zoom lens.

The result is that it would take 1.15 million data points through 360 degrees of rotation in order to adequately describe the pan angle of that camera, zoomed all the way in, at the pixel level. Since Vinten's i-Series encoded heads provide a minimum of 1.48 million data points per 360 degrees of panning rotation, we easily deliver the levels of performance necessary to resolve each individual horizontal pixel. (Doing the math to resolve the 1080 vertical pixel count for the 20 minute vertical angle of view provided at the extreme telephoto end of the same zoom lens yields a similar figure of 1.17 million data points through 360 degrees of tilt. Our i-Series heads provide a minimum of 1.48 million data points through 360 degrees of tilt.)

When moving up to provide augmented reality in a 4K, or a 3840 × 2160 pixel frame, double the number of data points is required, to 2.35 million through the 360 degrees of rotation of the pan and tilt axis.

Although this is a higher resolution than even our most accurate heads appear to output at present in fact the encoder technology in the i-series heads is scalable and can deliver up to 12m counts per 360 degrees. Both the i-series heads and VRI interface box architecture were designed to make this upgrade as easy as possible. Contained within the VRI are two interpolation modules (one each for the pan and tilt encoders) which can be exchanged by an engineer in minutes; the head requires no modification whatsoever. We've got 4K handled, with the longest zooms, and with room to spare.

It's worth addressing what happens in an augmented reality system when the positional data resolution is below one data point per pixel. The effect of not resolving it to the pixel level is that when you pan, the graphics that are composited with the video will potentially oscillate between pixel one and pixel two, and the rendering computer is not able to resolve the angle exactly. In a really bad case the rendering computer becomes confused as to where that graphic should be composited and the viewer will see a slightly fuzzy graphic when the camera is paned or tilted. It can even appear as though the graphic is floating slightly over the composited video. Having any of these errors occur destroys the illusion that the graphics are actually part of the real scene that people are watching, and undermines the value of the augmentation.

There are two other considerations to be taken into account of in 4K productions utilizing augmented reality: the zoom and focus data from the lenses and secondly, bandwidth.

If you are shooting in 4K with a 4K lens, will that lens need more encoding points for a given zoom within that range? Probably not, it's unlikely lens manufacturers will have to change the encoding performance in designing new 4K lenses. An HD lens typically has about 55,000 encoder counts between the two stops on the zoom, between full telephoto and full wide angle. Sensor size being equal, you're still working with the same focal length difference between the largest focal length and the shortest focal length of that lens. The 55,000 encoder counts should be adequate for the rendering engine to use a mathematic model of the lens (which may be specific to a single serial number or more generalized to a type of lens) to determine what that lens' field of view is and the magnitude of any distortion in the image to a degree that the compositing looks real.

One final consideration; as 4K video requires higher transport bandwidth, would 4K mean a greater flow of data between the VRI box and the rendering computer, possibly overloading the bandwidth of the serial connection between the two? The answer is no, the way our equipment is architected is, all of the processing and scaling is completed in the VRI box so the data packet transmitted (synchronized to the video frames) has the same format and size regardless of the resolution of the image and therefore encoders. Even in the case where frame rates increase (which may be beneficial to motion quality in 4k) there is sufficient processing power and bandwidth in the VRI system to handle the data required.. Even so, we're looking at Ethernet architecture in communicating this information and the bandwidth capabilities of Ethernet are obviously huge.