THE BROADCAST BRIDGE
Connecting IT to Broadcast

Virtual Production For Broadcast

Part 3 - Creative Image Capture

A Themed Content Collection from The Broadcast Bridge

Themed Content Collection

Contents

Series Overview

By Tony Orme. Editor at The Broadcast Bridge.

Virtual Production For Broadcast is a Themed Content Collection which serves as a reference resource for broadcast technologists. It covers the science and practical applications of all aspects of virtual production for broadcast.

Virtual Production is rapidly becoming the workflow of choice in cinematic and episodic TV production. With large-scale multi-location productions there are potential cost benefits but it is the versatility, creative scope and the improved efficiency it can bring to production spaces, that are the compelling forces driving adoption.

The basic principles of back projection and greenscreen have been with us for decades and are already commonplace in TV production, especially in news and sports, but the creative versatility of virtual production brings fundamental technical and creative differences. The technology and techniques of virtual production are also evolving very quickly and there is not yet a standard approach, with different teams establishing their own approach.

Virtual Production For Broadcast provides a deep exploration of the creative techniques, technology and workflow involved. It discusses what currently can and cannot be achieved, with a specific focus on the unique requirements of broadcast production.

It is essential reading for those evaluating incorporating virtual production technology into new studio design and exploring the creative benefits it can bring.

Virtual Production For Broadcast is a four part series:

Available now:

Part 1. The Foundations Of Virtual Production

Part 2. Planning, Virtual Worlds & Virtual Lighting

Part 3. Creative Image Capture

Future parts due in 2023:

Part 4. Uniting The Physical & The Virtual

- Moving Camera-side And Virtual Objects
- Image Based Lighting
- The Future Is Converging

Shooting Locations For Virtual Production

By Phil Rhodes. The Broadcast Bridge.

Sending out a crew to capture a real-world environment can be a more straightforward option than creating a virtual world, but there are some quite specific considerations affecting how the material is shot and prepared for use.

The best-known applications of virtual production have generally been those using real-time renders of a huge, detailed three-dimensional scene. Allowing the camera to explore that world with the convenience of an in-camera effect is a large part of the attraction of virtual production.

At the same time, some productions don't need a computer-generated world to show what they need to show. A driving montage might need to depict the same street used for exteriors cut into the same sequence, and sometimes virtual production is used to avoid travelling to locations which are entirely real – just far away or otherwise inconvenient.

Fundamentals Of Plate Photography

The idea of using live-action footage as the background for a shot dates back to the earliest days of cinema, with in-car shots of the 1940s notorious for their wobbly backgrounds. Modern stabilisation techniques make it easier to escape the wobble, but some of the same considerations still apply. Lens effects including flare, distortion, softness and geometry errors should ideally be absent from a background plate. No image is ever entirely free of those things, but in a perfect world the plate should be as clean and sharp as possible, with creative lens and filter choices left to the taking camera in the final setup.

That's complicated by the fact that virtual production stages are often configured to cover a very wide field of view. As such, shooting live-action environments for them will often demand more than recording a simple, single-camera background plate. One of the existing 360-degree camera systems could be used, although the high resolution and contrast requirements mean that only the very best are likely to be suitable. More often, plates for virtual production are often shot in much the same way as plates for high end visual effects, using an array of high-end cinema cameras configured so their images can later be combined – stitched – to create a single image.

Cameras And Lenses

The requirements for angular coverage and resolution are a geometry problem, depending on the resolution of the LED wall and the choice of camera position and lens setup. Some of the world's most impressive virtual production stages have very high resolution – often in the tens of thousands of pixels in the horizontal direction. That's only likely to increase as LED video wall panels themselves improve, mainly because tighter LED spacing reduces the likelihood that individual emitters become visible.

So background plate shoots are likely to choose high-resolution cameras, potentially highlighting issues which we might not notice in the course of normal shooting. Among those considerations are diffraction limits, which can limit resolution in counter-intuitive ways. A 12K camera with a Super-35mm sensor is diffraction limited above about f/4, meaning that any narrower aperture may start to create softness. With wider apertures, meanwhile, depth of field limits might require the cinematographer to choose where to set the focus for proper integration into the real-world foreground when the material is later used as a backdrop.

It's also common to choose wide angle lenses to photograph background plates, maximising the coverage per camera at least as much as is required for the proposed setup. Generally, rectilinear lenses will be chosen, avoiding the distortion of a fisheye, although all lenses have at least some distortion which may need to be corrected in software later (see stitching). Wider-angle lenses may also be more likely to suffer corner softness and other geometric distortions. High-performance modern lenses minimise those problems, and are often used for plate photography even when the final shot will use a beloved and characterful lens of history.

For similar reasons, it's normal to record plate photography in a high-quality format. How critical this is depends on the proposed final setup – a shot which will be seen out-of-focus will demand less than one in which we want to be able to recognise at least some detail. Some LED walls are set up to render colour according to the ITU's Recommendation ITU-R BT.709, commonly "Rec. 709." Recordings from high-end cinema cameras invariably contain more colour and brightness information than typical Rec. 709 displays can handle, so conventional camera log formats for background plates will usually capture enough information for good results. Grading, as necessary, is likely to be part of the stitching process which combines multiple recordings into a single image. Background plates will usually be processed for a straightforward, naturalistic look, since the finished scene will be graded again once it's been shot on the virtual production stage.

Rigging Multiple Cameras

Where more than one camera is used to shoot a plate, proper rigging is essential to fix the cameras in their relative positions. Good rigging makes it easier to combine the multiple resulting files into a single image. That will always require a certain amount of post-production adjustment and some degree of overlap between cameras so that the seam can be made invisible. The amount of overlap needs to accommodate changes to the image shape due to adjustments for lens distortion; too small an overlap can create complicated problems.

A number of mechanical configurations are possible. Generally, the aim is to place the cameras as close together as possible and minimise errors due to parallax, which are especially visible when shooting nearby objects. The effects of parallax reduce as the distance between the camera and the subject increases. For applications such as aerial photography, the subject may often be so distant compared to the spacing between cameras that parallax errors are negligible.

Depending on the physical configuration of the cameras and the available space, toe-in or toe-out configurations might work best. A toe-out setup has the cameras pointing outward, whereas toe-in places them aiming inward at a notional point in front of the lens such that their fields of view cross over. Toe-in setups can put the cameras as close together as is allowed by the front diameter of the lens, which is often a comparatively narrow part of the camera.

It may sometimes be easier to mount some cameras upside down with respect to others, and in principle the orientation can be corrected when the images are combined. With rolling shutter cameras (which includes several popular high-end cinema cameras) there may be subtle changes to image geometry when the camera or subject is in motion, and that change depends on the scan direction of the rolling shutter. Similar problems exist when shooting stereoscopic 3D where one camera may be upside down with respect to the other. As such, wherever possible, the safest choice is to keep all the cameras in the same physical orientation.

The key benefit of modern technology is stabilisation, and rigging cameras for plate photography will mean creating an arrangement which can be mounted on an active stabilisation device. That might mean a straightforward handheld gimbal for the simplest jobs, all the way up to a vehicle-mounted stabilised remote head for more advanced work. However the camera or cameras are rigged, the limits of the stabilisation system will control how much weight can be carried, and how large the overall camera setup can be.

Stitching

The mechanical rigging of a multi-camera system will never be pixel-perfect, and some post production effort will be required to assemble the multiple files into a single image. Stitching large plates can be a big job for post-production workstations, often involving long sequences of high resolution images recorded in demanding, high quality formats. The process isn't enormously complicated compared to advanced visual effects, but it is not trivial, and last-minute changes may be difficult. Background plates for visual effects are often only finessed in regions which will be seen in the final shot, whereas virtual production backgrounds must be perfect throughout as it is not known what will eventually be seen. Images are easier and faster to stitch if they're well-shot, and some post production processes may be able to make use of lens grids, which are charts shot to allow analysis (even automatic analysis) and correction of the distortion of a lens.

Playback

Playing back such large and demanding sequences can be a challenge. Many virtual production facilities will have done so before and will be able to advise on how to format and supply the background plate for easy playback. That might involve splitting the image up to be played back by several servers in parallel, or other prerequisites. Because of varying playback and display requirements, even where a production can source its background plates from a library, there might be some work involved in reformatting and preparing that plate for the job in question.

Shooting with a single camera removes the need for stitching, and high resolution cameras with wide angle lenses have been developed with exactly this in mind. Considerations such as distortion, flare and reformatting remain, though, and the single sensor has limited performance in comparison to the much larger total sensor area of a multi-camera array.

Live-action background plates are not a panacea and shooting them well is not a trivial task. Still, they'll often reduce the workload massively compared to creating a fully three-dimensional virtual scene, even to the extent of making virtual production practical where it might otherwise not have been.

Capturing Objects In 3D

By Phil Rhodes. The Broadcast Bridge.

Sometimes, there'll be a need to represent real-world objects in the virtual world. Simple objects could be built like any VFX asset; more complex ones might be better scanned as a 3D object, something some studios have begun to consider as a service to offer.

Populating a virtual world with convincing objects is a lot of work. As we've seen, not every application of virtual production will need to create a full, three-dimensional virtual world, but when we do, the virtual art department might be faced with the need to design and build an intimidating number of assets.

It's not always necessary for everything in a virtual world to have been built from scratch. A lot of library assets are available, albeit mainly intended for game development. While they might look great on a PlayStation, the detail and finish might not be enough to convince a movie or TV audience. Even so, depending on the nature of the object, it might make sense to purchase an existing asset and improve it as opposed to building something from scratch. That might require artists with experience in games and visual effects, although any virtual art department will need those skills anyway.

Inevitably, though, at least some assets must be built from first principles, especially when something which exists in reality must appear in the virtual world. Sometimes, we might need vague shapes to lend a little three-dimensionality to live-action background plates. Where more complex objects are involved, particularly irregular shapes with a lot of complexity, capturing a real world object becomes attractive.

Traditional Modelling

The earliest computer generated imaging programs built their objects from primitives such as spheres, cylinders and cubes. By stretching, moving and rotating those primitives, and by mathematically subtracting one from another, complex modelling is possible. With that in mind, the techniques behind the light cycles of the 1981 Tron become clear. Advantages include the fact that a sphere or cylinder has no resolution limit; it never becomes faceted. It's an inflexible approach, though, making complex objects hard to build, and the idea of modelling surfaces defined by flat polygons quickly became popular.

Polygons are invariably triangular because they are assumed to be flat, the three points can only define a flat shape (camera tripods are stable on uneven surfaces for the same reason, while a four-legged table might rock). It's still possible to create spheres, cubes and cylinders, but what we describe is their outer surface, not their shape and volume. Tools for polygon modelling are a key feature of most 3D software, and techniques for capturing real places and objects are invariably designed to create polygonal models.

Laser Scanning

Virtual production is often about placing the scene in a novel location, and capturing locations has often involved laser scanning. The basics are easy to understand – in a laser rangefinder, distance is measured by timing pulses of light. Repeat that process thousands of times using a rotating mount, and a three-dimensional representation of an area can be created. It can be a reasonably slow process, taking tens of minutes to scan an area with reasonable resolution.

Laser scanners can only scan surfaces which are visible from the position of the scanner. That might mean taking several scans of a single environment to fill the voids, and very complex environments may make it difficult to ensure every last surface is scanned. Also, laser scans (like many 3D scanning techniques) produce a series of distance measurements, sometimes called a point cloud, which must be converted to a series of polygonal surfaces for most uses. Given the cloud may contain millions of points, that's necessarily an automatic process, although it may fail where an area of the model was not scanned clearly.

Laser scanners are commonly used for creating a model of an environment that the scanner is within, at least to within the maximum range of the scanner. Smaller models can be used to create a model of a small object, although that's perhaps not the commonest application.

Photogrammetry

The term photogrammetry covers a wide range of computer vision techniques, but in film and television effects it generally means creating a 3D model of something by taking a series of photographs. It derives from the stereoscopic techniques behind 3D cinema, expanded such that we can use more than two photographs – dozens, even hundreds – to accurately create models of almost anything that can be photographed.

Photogrammetry works on everything from tiny objects to vast landscapes. The reference pictures can be taken fairly quickly using a reasonably low-cost, conventional stills camera, and software to turn those photos into a model, at least in simpler situations, might even be free. Shooting photogrammetry reference images demands some knowledge of how the process works, and ideally enough experience to understand what images are needed. Under-coverage can mean inaccurate models, while overdoing it will slow down make the process of processing the images.

Photogrammetry is often the technique used by cellphones (some have additional laser scanners) and it's one of the most immediately accessible approaches to capturing a 3D model.

Structured Light

One of the most widely-deployed consumer applications of 3D capture involves structured light. Microsoft's Kinect accessory for the Xbox used the technique, projecting a pseudo-random (but known) pattern of invisible infrared light into the room and photographing that pattern using a camera. The degree to which the pattern appears to have moved sideways is proportional to the distance.

Some hand held object scanners use the same approach, often with higher-resolution cameras and higher-resolution structured light patterns than Microsoft's 2010 design. The fundamental limitation is one of range: a structured light scanner can only scan objects which are close enough for the light pattern to be visible. It can be more accurate than photogrammetry, but is most often used for objects small enough to put on a desk.

The Limits

The limits of 3D object scanners generally arise from the fact that most of them – with the exception of laser scanners – are dependent on image recognition. Very dark objects which just don't bounce back much light can create problems, as can complex patterns which may make it difficult for a structured light scanner to recognise the target pattern. Conversely, objects without surface texture can be difficult for photogrammetry software to track. Almost any scanning technique can be confused by ambient lighting conditions.

Transparent and reflective objects often cause problems. Laser scanners may mistake mirrors for being a doorway to a complete, depth-reversed copy of the room being scanned. Photogrammetry may struggle to track transparent or reflective objects, and that can create large errors requiring a manual fix.

Possibly the most limiting caveat is noise. Even the most well-finished wall, which a 3D artist might represent with a simple plane, is likely to be seen by the scanner as a number of triangles due to inevitable random noise in the data. Lots of 3D modelling software has features to automatically optimise the polygon layout of any model, although a scanned model will generally involve more polygons, in a less ideal layout, than an equivalent handmade model. More polygons demand more rendering performance, and rendering performance is at a premium in virtual production, so scanned assets are likely to be handled carefully and only when necessary.

Materials

Experienced visual effects people will happily confirm that the materials assigned to a 3D object are often much more crucial to its believability than the fine detail in its geometry. Most of the scanning techniques we've discussed here are capable of capturing both shape and colour. That's clearly true of photogrammetry, which relies on colour photographs of the subject. Even laser scanners often include a supplementary colour camera to record surface detail. One consideration is that these images will represent the fall of light on the object as it was in reality, which may cause problems later when the 3D object is used in a different lighting scenario (in the parlance of the field, the object has its lighting baked in).

3D scanning is therefore best done under diffuse, even lighting, which may be difficult over a large area. Even that risks misrepresenting reflective or transparent surfaces. Some advanced scanning techniques can estimate surface reflectance very accurately using modulated lighting, and they've often been used to scan faces. Very often, though, really good materials will still mean manual intervention. Advances in real-time 3D rendering technology mean that physically-accurate materials, using real-world values for phenomena such as reflectance and diffusion, are increasingly available outside the world of offline, non-realtime rendering. Physically-based materials tend to require less tweaking and may carry better between software and lighting scenarios.

For Speed And Accuracy

There's no straightforward answer as to whether 3D scanning is the right approach for the creation of any particular asset. Poorly-done 3D scans can create complex lighting problems and impact performance. Well-done scans can also be a fast route to very convincing objects. What's almost certain is that for many applications of 3D graphics, from video games to visual effects to virtual production, are starting to tax the limits of human ingenuity in creating sufficiently detailed environments. Even more realistic environments are likely to become more dependent on object scanning, so there's every reason to hope that the tools will continue to improve as time goes on.

Motion Capture

By Phil Rhodes. The Broadcast Bridge.

One of the creative advantages of virtual production for performers is seeing the virtual environment in which they are performing. Using motion capture techniques extends this into capturing the motion of performers to drive CGI characters. New technologies are rapidly transforming the creative freedom this brings.

As we saw in a previous piece on camera tracking, it's not inevitable that virtual production need involve tracking technology at all – it's quite possible to use the LED video wall in the same way as back projection was used for decades. Motion capture is normally used as a post-production process for visual effects work, although some particularly advanced setups have used it to animate a virtual character in real time, so that an actor – and thereby the character – can react in real time to the live action scene.

Figures Of Merit

Some of the same technology which is used to capture camera position can also be used to track people, although those tasks can have sufficiently different requirements that separate systems are used to track cameras and performers. For a three-dimensional scene to be displayed on an LED wall with proper perspective, the camera position relative to the wall must be known with good accuracy. Tracking a person, meanwhile, can sometimes accept small errors so long as the overall effect is convincing.

Evaluating motion tracking technologies for any particular application requires some knowledge of the underlying principles and the limits of various technologies.

Technologies

The most familiar camera-based optical capture system is an outside-in configuration, with cameras surrounding the action and observing passive, reflective markers on the performer. This configuration can offer a large working volume, with the option to trade off accuracy and volume by altering the location of the witness cameras. Placing cameras to cover a larger space allows more room for the performance, but may reduce accuracy when the performer is far from the cameras.

Inside-out systems place a witness camera on the taking camera which observes markers in the environment. These systems are often recognisable by the scattering of reflective dots or circular barcodes in the ceiling of the studio. This arrangement allows them to cover large areas, but they are usually made to locate one single point per witness camera. Systems of this type are often used to track several cameras in a broadcast studio, but they are generally not capable of tracking the multiple locations on a human figure that would be required to recreate a performance.

Inertial systems measure position by sensing acceleration and deceleration over time. Like the inertial navigation system on an aircraft, they may be subject to some degree of drift over time. Similar inertial reference systems are sometimes built into modern lenses to report approximate camera position for later visual effects work. The compensating advantage is that these systems can work over a large area, often limited only by the range of a radio data link between the performer and a base station. An optical system can only operate in the area covered by a sufficient number of witness cameras.

Similar benefits attend mechanical motion capture devices. Mechanical systems detect the position of the performer's joints using potentiometers or optical encoders. The approach is often combined with other techniques, particularly optical or inertial, which allow the device to establish its overall position in space. Still other technologies, particularly such as those based on magnetic field sensing, may have a capture volume strictly limited by the physical structure of the device. Because magnetic fields pass through many objects, they can locate all of their tracking markers at all times, regardless the position of the performer. Some active-marker systems, which rely on the performer wearing markers which might also rely on a fixed frame to detect the position of those markers limiting space.

Finally, markerless motion capture systems are often based on machine learning (which is not necessarily the same thing as AI). Markerless systems can derive motion capture data from something as simple as a video image of the performance, ideally with reasonable lighting creating a clear view of the performer. At the time of writing (late Spring 2023), the results of these systems were generally not as precise as those using more conventional approaches, although machine learning is a rapidly-developing field and improvements are widely anticipated.

Caveats

Motion capture as a technique for post production visual effects can produce highly realistic results which contribute significantly to the believability of an effect. It can also work quickly, potentially avoiding the hours of exacting work involved in animating something by hand. Actors appreciate the process because the captured motion reflects all the subtlety of a real performance, although sometimes, motion capture may be performed by a stand-in or stunt specialist.

Recording the finest details of motion is also one of the downsides. Where motion capture data must be recorded and potentially modified, it quickly becomes clear that is difficult to edit the unprocessed data. In conventional animation, the motion of an object between two positions is usually described using only those two positions – waypoints – which are separated in time. Changing the speed of the object's motion simply means reducing the time it takes to move between the two points.

Motion capture data records a large number of waypoints representing the exact position of an object at discrete intervals. It's often recommended that motion data should be captured at least twice as frequently as the frame rate of the final project, so that a 24fps cinema project should capture at least 48 times per second. That's well within the capabilities of most systems, but it does complicate the process of editing motion data. It's impractical to manually alter dozens of recorded positions per second and achieve a result that looks realistic.

Tools have been developed to facilitate motion capture data editing. Some of them rely on modifying groups of recorded positions using various proportional editing tools; a sort of warping. Others try to reduce the number of recorded positions, often by finding sequences of them which can be closely approximated with a mathematical curve. This can make motion capture data more editable, but too aggressive a reduction of points can also rob it of the realism of a live performance, risking a more mechanical, artificial look which is exactly what motion capture is intended to avoid.

Often, motion capture used where a performer is working live alongside a virtual production stage won't be recorded, so there won't be any need or opportunity to edit it. Other problems, such as intermittent failures to recognise tracking markers, might cause glitches in positioning that might usually be edited out. Working live, a retake might be necessary, although well-configured systems are surprisingly resistant to – for instance – markers being obscured by parts of the performer's body.

Rigging And Scale

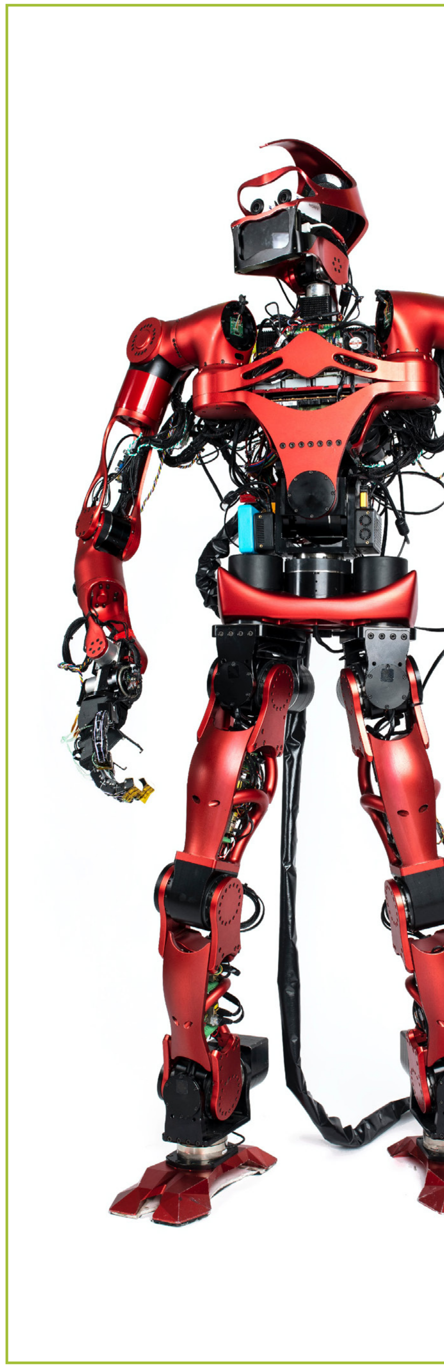
Connecting motion capture data to a virtual character, requires that character model to be designed and rigged for animation. Where the character is substantially humanoid, this may not present too many conceptual problems, although the varying proportions of different people can still sometimes cause awkwardness when there's a mismatch between the physique of the performer and the virtual character concerned.

Very often, the character will be one which looks something other than human. It may be on a substantially different shape, scale or even configuration of limbs to the human performer whose movements will drive the virtual character. Various software offers different solutions to these considerations, allowing the performer's motions to be scaled, remapped and generally altered to suit the animated character, although this has limits. Although motion capture technicians will typically strive to avoid imposing requirements on the performer, the performer might need to spend time working out how to perform in a manner which suits the virtual character. This approach which can make a wide variety of virtual characters possible.

On Set With Motion Capture

Most motion capture systems require at least some calibration, which might be as simple as moving around the capture volume with a specially-designed test target. Some of the most common systems, using spherical reflective markers, may require some calibration for each performer, especially if the performer removes or disturbs the markers. Many virtual production setups rely on motion tracking to locate the camera, even when motion capture is not being used to animate a virtual character. As such, almost any virtual production stage might rely on at least some calibration work, though there is often some variability in how often this is done; performance capture spaces might do so twice daily, requiring a few minutes each time.

As with many of the technologies associated with virtual production, motion capture, where it's used, is likely to be the responsibility of a team provided by the studio itself. Most of the work required of the production will be associated with the design of the virtual character which will be controlled with motion capture. The technical work of connecting that character's motion to the capture system is an item of preparation to be carefully planned and tested before the day. With those requirements fulfilled, using an actor's performance to control a virtual character can provide an unprecedented degree of immediacy. While it certainly adds another layer of technology to the already very technology-dependent environment of virtual production, it creates a level of interactivity which was never possible with post production VFX.




Amazon Studios - Los Angeles.