Camera Tracking

an industry op-ed by Steve Wright

A Technical Wonder

I remember once being given the grand tour at Digital Domain, the powerhouse VFX studio in Los Angeles. As we were walking from building to building and room to room I saw nothing but computer workstations. Finally we wound up in a sound stage where my guide pronounced this to be the last soundstage left at DD, used primarily for large dance numbers. I was thunderstruck. The world of models and miniatures, motion control rigs, cloud tanks, pyro experts, stunt men, matte paintings - the meat and potatoes of visual effects - had all concatenated down to one single technology - computer animation.

Of course, live action plays a prominent role in visual effects since many things are still best shot live. But we live in a visual effects world that routinely puts CGI and live action in the same shot. But there is a colossal barrier to blending the worlds of live action and CGI, and that is the need for identical camera moves. If a T-Rex is to be shown thundering down Main Street then the 3D camera used to render the T-Rex had better match the live action camera move. If not, the T-Rex will not match in perspective or camera angle and will drift and squirm all over Main Street. Not good.

It turns out that “reverse engineering” the camera move of a live action shot is a non-trivial problem. I have, in the early days of my visual effects career, tried to match a 3D camera to a live action plate by hand, simply because camera tracking software had not been invented yet. It is a maddening task because there are literally an infinite number of possible combinations of camera location, orientation and lens sizes that would appear to match the live action.

Utilizing extremely advanced and powerful mathematical algorithms, camera tracking is the major modern miracle that makes it possible to put 3D into live action, live action into 3D, add set extensions, or any of the long list of other visual effects that combine elements of live action and 3D into the same shot. It is the “glue” that binds the 3D world to the live action world. In order for a 3D object to appear convincingly with a live action element, they must both share exactly the same camera moves and lenses.

Figure 1 - a million ways to move the modern camera

This is not an issue for shots with a locked off camera, but no movie director worth his riding crop will want locked off visual effects shots. Cameras are no longer limited to the tracks and dollies of yesteryear. Now they ride on great unmanned booms 50 feet in the air, on the front and back of camera cars, they are hand-held; they are steady-cams, gyro-stabilized helicopter cams, cable cams, even a “birdy cam” that rides on a miniature remote control helicopter (check out Figure 1). There have never been so many ways to move a camera around to get that truly cool shot. There has never been a greater need to reconstruct the live action camera move so that it can be matched in the 3D part of the picture.

Camera tracking used to be the sole province of the “match move” department, a murky collection of computer geeks that scurry around late at night working their incantations on ultra-expensive camera tracking software with exotic names like “Boujou”, “Voodoo”, “SynthEyes”, and others. While the match move department is alive and well, the paradigm shift in compositing is that there are now very capable camera trackers integrated into commercial compositing programs. Not as crafty or powerful as the uber-expensive dedicated programs, but capable of doing a yeoman’s job for many shots. Today’s compositor is now a match move artist too! So let us take a look at how camera tracking software actually works.

Here is the first key point about camera tracking - the camera must be moving in order to do camera tracking. It must be moving because the tracking software uses the parallax shift of multiple tracking targets in the picture in order to triangulate the target’s location in 3D space. No camera movement, no parallax, no 3D information. The second key point is that the tracking software only locks onto static objects - the corners of buildings, the landmarks on a mountain - and ignores moving objects like people and cars.

Camera tracking is actually a two stage process. The first stage is the actual tracking, where the computer rummages through all the frames in the shot searching for features that it can lock onto and follow for a good portion of the shot. The second stage is the “solve”, where the computer grinds on the tracking data it collected to calculate the camera’s position, orientation, and lens for each frame of the shot.

The digital artist assists the computer during both stages. During the tracking stage the artist can delete tracking points that are confusing the computer and plant some of his own to guide it along. During the solve stage the artist can give the computer any known information about the scene (the size of an object, the distance from the camera to an object, the type of lens used, etc.) to help guide the computer towards a successful camera solve. Tragically, in most situations, the digital artist will have no information from the scene and must make his best guesses. In spite of this, modern camera tracking software can solve most shots with great accuracy and even provide the type of lens used including, incredibly, information about the lens distortion.

After the tracking is done we are ready for the solve. The computer runs forwards and backwards through the tracking data multiple times using successive refinement to converge on its best solution for the camera data. In addition to the camera data it also generates what is called the “point cloud”. Each of the trackable features in the scene is triangulated into 3D space so the computer knows their XYZ coordinates. This collection of 3D points is exported along with the camera data and displayed in the 3D portion of the compositing software (or a dedicated 3D program like Maya).

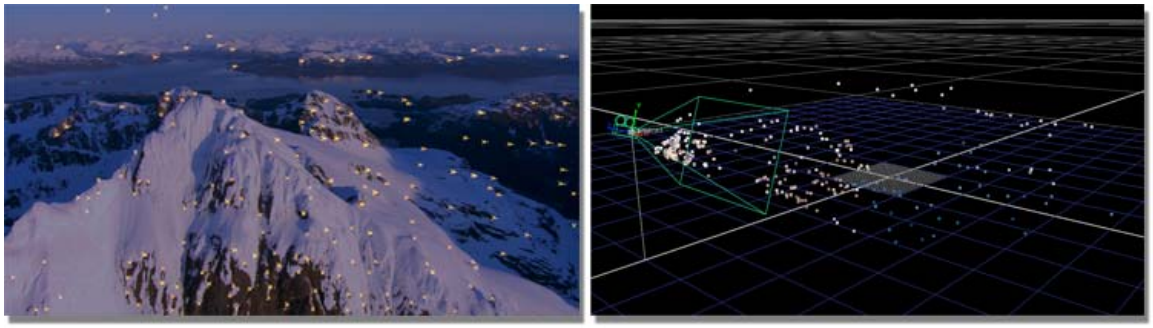

Figure 2 - tracking markers and the resulting point cloud

In Figure 2 above, the picture on the left illustrates a moving helicopter shot over some majestic mountains with yellow tracking markers superimposed. These mark the features that the camera tracking software found attractive and locked onto to collect the tracking data. The picture on the right shows the resulting point cloud showing the 3D location of each of the tracking markers relative to the camera position.

The purpose of the point cloud is to give the digital artist reference points in 3D space to line up the 3D objects. If we wanted to place a castle on the top of the mountain we would use the point cloud as a guide to line up the position of the castle, then render it with the 3D camera from the camera solve. The perspective on the castle and its motion would match perfectly with the live action scene in the final composite.

For me, it was 2D motion tracking that had always been one of the great wonders of computer visual effects as it will magically match the movement of a background plate, something I cannot do by hand. Camera tracking, however, builds on this basic 2D tracking with an entirely new layer of sophistication that compounds the miracle by deriving 3D information from a 2D scene which makes it possible to seamlessly integrate 3D and live action into the same shot. Without it, visual effects would be static, indeed.

Steve Wright