29 August 2008

The Role of Ocula in Stereo Post Production

Lucy Hallpike

0. Introduction

The Foundry's Ocula tool set is a new collection of plug-ins designed to alleviate common problems in stereo post production. These range from inherent problems with different filming techniques – such as the perspective distortion that occurs with "toe-in" camera set-ups - to the sheer time and effort required to apply all VFX work to two views instead of one.

A key component of Ocula is the Foundry's new stereo disparity estimation technology, which is encapsulated in the plug-in O_DisparityGenerator. Much of Ocula is underpinned by disparity estimation in the same way that motion estimation drives many of the Foundry's Furnace plug-ins, which are already well-established tools in post production.

1. Stereo Disparity Estimation

An image – such as a single frame of a motion picture – is a two-dimensional (2D) representation of a three-dimensional (3D) scene, which by definition will contain structures at different depths. In an image taken from a slightly different viewpoint - for example, that of the second camera used to produce a stereo pair - the same structures will appear slightly shifted. Stereo disparity is the term used to describe the shift that occurs for a particular point in 3D space between the left and right images. In a stereo pair the cameras are offset horizontally, so this usually corresponds to a purely horizontal shift. However, the amount of the shift is not constant but can vary from pixel to pixel; it will depend on the depth of the corresponding object within the scene. If parallel cameras are used, objects in the foreground – close to the viewer – will shift significantly while those in the distant background might not move at all. In addition, as the viewpoint shifts from one eye to the other, areas that were visible before might become hidden, such as background areas that become occluded when a foreground object shifts across in front of them, or a surface of an object that becomes obscured when the object is viewed from a different angle. Similarly, previously hidden areas can be revealed. Figure 1 shows an example of a stereo pair of views of a scene in which different areas are occluded and revealed in each camera's view.

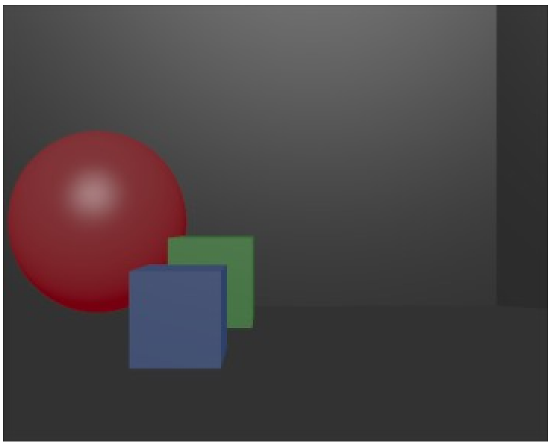


Figure 1a: Left view of a scene

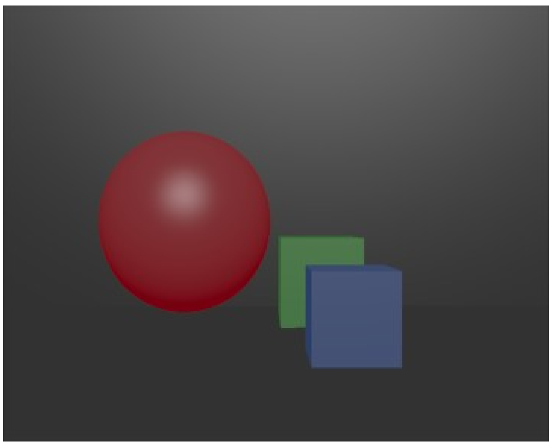


Figure 1b: Right view of the same scene

O_DisparityGenerator builds up a complete picture of the stereo disparity by estimating the change in position of every point in the scene between one view and the other. Disparity estimation is similar to local motion estimation, in that the goal is to estimate how each pixel moves from one image to another. However, local motion estimation is an example of unconstrained optical flow: we make no assumptions about the nature of the scene (other than local smoothness) and pixels are allowed to move in any direction. This is because local motion estimation is usually performed on two frames of a sequence, separated in time. The time separation between the frames means objects within them might have have moved in the interim, and we have no prior knowledge of the motion of these objects. With a stereo pair of images there is no time separation, so the images should be unaffected by motion within the scene. The only "motion" between one view and another will be that resulting from the physical separation of the cameras. This has the advantage of being rigid body motion – scaling, rotation and translation only – and applies to the whole of the image, unlike the motion in a typical image sequence which could include local deformations (imagine a sequence showing a man walking down a street, for example). We therefore know that objects within the scene will be transformed only in this rigid way between one view and the other, and can use this additional knowledge to constrain the motion (i.e. disparity) estimation. In practice, we do this by detecting features in both views of the scene first of all, then calculating the stereo disparity of those features[1]. Given the stereo disparity of these pairs of features, we can then find a transform which maps points in one view to lines in the other view (so-called epipolar lines)[2]. The disparity estimation for the rest of the points in the image is then carried out under the condition that the corresponding pixels in the other image must lie along these lines. Because of this additional constraint, motion estimation done in this way tends to give more accurate results for the disparity than would local motion estimation performed on the same pair of images.

The output from O_DisparityGenerator contains two channels: the stereo disparity calculated from the left view to the right and the equivalent disparity calculated in the opposite direction, from the right view to the left[3]. These two disparity maps provide the foundation for most of the other Ocula tools, which are described in the remainder of this document.

2. Convergence and Depth Adjustment

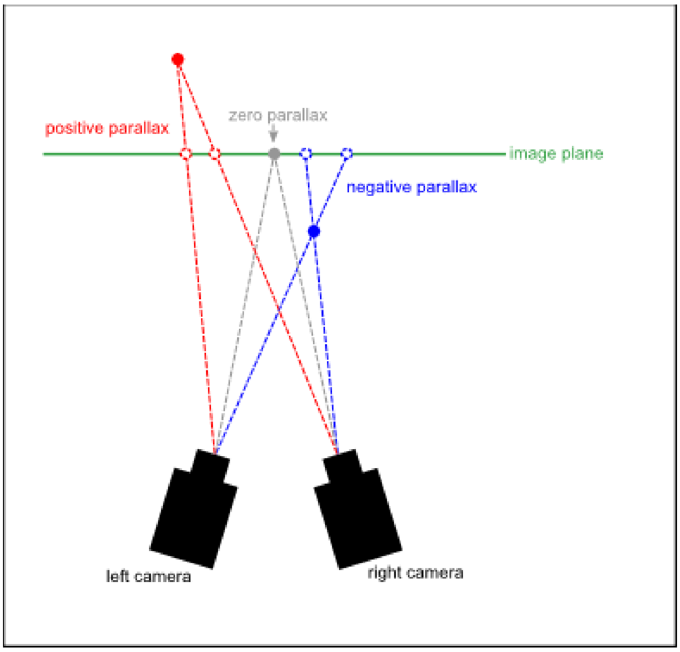
Historically, one of the main problems with shooting in stereo has been the need to make certain decisions about the desired 3D effect before the shoot takes place. To some extent, this problem has been reduced by the development of more advanced camera set-ups which allow for more flexibility during a shoot. However, it is still the case that some 3D properties are fixed after filming and have previously been difficult or impossible to alter as the finished film takes shape. A fundamental property of a 3D shot is the interocular distance; this is the distance between the left and right cameras[4], and determines the perceived depth of the scene. Usually, this distance should be about the same as the distance between the viewer's left and right eyes[5] but it can be adjusted in order to achieve specific effects. For instance, it might be necessary to increase it significantly in order to get a better sense of depth over a distant landscape [7], where filming with the standard interocular separation would result in an essentially flat-looking scene (this is sometimes known as "infinity flatness" [5]).

1. The features in question are points of interest corresponding to recognisable structures within the scene, for example in a scene containing a wooden table they would include the corners of the table top and perhaps some conspicuous marks on the surface of it, such as a dark knot in the grain.
2. In order to be able to map points to points, we would need to know how far away the corresponding 3D points are from the camera. Without this, there is an additional, unsolved degree of freedom which means the points in the first view can lie anywhere along a line in the second.
3. Since occluded areas in one view can become revealed in the other (and vice versa), these two channels are not simply the inverse of one another.
4. The distance between cameras is sometimes, perhaps more correctly, known as the interaxial distance in order to distinguish it from the distance between the viewer's eyes. However, the term interocular is often used more broadly to describe both the inter-camera and inter-eye separations; we have chosen to follow this convention here.
5. About 65mm, for an average adult.

Increasingly the interocular distance substantially can also result in "hyperstereo" or "miniaturisation", where the exaggerated stereo effect means that the viewer feels herself to be massive in comparison to the scene being viewed. This is because the human brain is used to experiencing the amount of stereo separation that results from the separation of one's eyes. Anything greater fools the brain into thinking that the eyes must be further apart in relation to the scene in front of them than normal. Since the distance between the eyes is fixed, the brain can only make sense of this by reasoning that the scene must be smaller than it really is. This can have the effect of making the viewer feel like a giant, or alternatively like a normal human looking down at a scale model [1]. Similarly, a reduced interocular separation can have the opposite effect, making the viewer feel she has shrunk in relation to the scene. Although it is usually desirable to avoid such effects, sometimes they can be invaluable tools in helping to tell a story or fulfil a director's artistic vision (it is easy to imagine how they could be used to good effect in a 3D version of the children's classic Alice in Wonderland, for example, as the heroine shrinks and grows in response to eating certain things). In addition, maintaining the standard interocular separation throughout – while being both realistic and comfortable for the viewer – can make for an unexciting stereo experience. After a while, the viewer's brain will adjust to the stereo effect and they might cease to appreciate the extra dimension once the initial novelty wears off [4]. Varying this separation - and thus the perceived depth - between scenes from time to time can help to provide the necessary stimulation to keep the viewer's brain alive to the differences between the 3D scene in front of them and their customary 2D viewing experience. However, any variation in depth must be used with caution, as people's eyes take a while to adjust to significant changes in the stereo separation, which means that fast cutting between scenes with significantly different interocular separations will be uncomfortable for the viewer and should be avoided [3].

Ocula contains the plug-in O_InterocularShifter which allows the effective interocular distance in a scene to be reduced. It uses the stereo disparity maps calculated by O_DisparityGenerator to create new views which have been shifted towards each other, thereby decreasing the interocular distance and the apparent depth of the scene. The disparity maps tell us, for instance, how much each pixel in the left image must move in order to create the right image. Once we have this information we can instead shift each pixel from the left image only a fraction of the way towards the right one, thus creating a new view of the scene as it would have appeared to a camera slightly to the right of the original left camera. We can do the same to the right image in order to simulate the view that would have been seen from a camera slightly to the left of the right camera. The result is a new stereo pair of images with a reduced interocular separation. A 3D view of these will now show the same scene as before, but with a reduced range of depth. This could be useful during the editing process, for example when trying to match the depths between scenes in order to make transitions more comfortable for the viewer, or simply because the desired depth of a shot has been reconsidered as the final film evolves. It might also help to maintain the apparent depth of a scene during a zoom or camera move.

Figure 2: Parallax resulting from converging cameras



Another fundamental property of a 3D scene is the point of convergence. At this point, objects in the two views will have zero horizontal disparity, also known as zero parallax, and will appear to be the same distance away as the screen on which the scene is being viewed. For a scene shot with cameras that are converging – i.e. pointing or "toeing-in" towards each other - this will occur where a ray emerging from the front of one camera, perpendicular to its image plane, would meet a similar ray from the other camera (see Figure 2[6]). Anything in front of this point will have negative parallax – the object in the left image will be to the right of the same object in the right image – and appear to be in front to the screen. Similarly, objects behind the point of convergence are said to have positive parallax and appear behind the screen.

Convergence can be changed in post with a simple horizontal shift of one image relative to the other (which in Nuke has been implemented for convenience as the ReConverge node). As with the interocular separation, the eyes can take a while to adjust to sudden changes in convergence so it is desirable to try to minimise these when cutting between scenes, for example. Moving the point of convergence nearer to the cameras will have the effect of shifting everything else further away, while converging on a more distant object will bring everything closer. In this second instance, care must be taken to ensure that the scene stays inside the area which can comfortably be seen by the audience. During normal vision, our eyes converge on an object as we look at it, and focus on the object at the same time. However, when viewing a stereo presentation such as a 3D film, our eyes will always be focused on the screen, yet will be required to change convergence as the scene changes or as we look at different parts of it. To some extent, they are able to do this, but when the distance between the focal point and the convergence point becomes too great, viewers will experience discomfort. If an object appears too far in front of the screen, they will feel as if their eyes are crossing and may even be unable to fuse the left and right images into one. Similarly, if

6. This diagram has been simplified in order to illustrate the point. In reality the image planes of converging cameras are not coincident (see section 5).

objects are pushed too far away, the positive parallax between the views could increase to the extent that the eyes would be required to diverge in order to fuse the two images. Divergence never occurs in normal vision; it is generally accepted that a small amount is acceptable in 3D cinema [6], though this can make for an uncomfortable viewing experience, causing eye strain, headaches or nausea for the audience. Beyond this, the 3D effect will be lost and the audience will see two separate images of the object. In order to ensure that a change in convergence does not require the eyes to either converge or diverge excessively, O_InterocularShifter can be used to bring the two images closer together and reduce the overall depth of the scene at the same time.

O_InterocularShifter and Nuke's ReConverge node are also valuable tools in the process known as depth grading, where the convergence and depth of field are adjusted in order to make sure the stereo effect can be comfortably viewed on the screen size for which the finished film is intended. The apparent depth of the scene will depend upon a combination of the screen size and the distance from the screen to the viewer.

3. Automatic Stereo Correlation

An obvious complication of stereo post production is that all work must be applied to two views of a scene instead of one. This includes already painstaking tasks such as paint and roto work, and is further complicated by the need for absolute precision when working with stereo views, as even slight inaccuracies in relative positioning can cause disruption to the eventual 3D effect. In extreme cases, errors could even cause headaches for the eventual audience as their brains try to make sense of what their eyes are seeing. Another Ocula plug-in, O_Correlate, is designed to simplify the process of compositing on stereo views by automatically reproducing work done on one view in the other. For example, with O_Correlate a paint stroke drawn on the left view can be automatically copied to the same location in the right one. Similarly, a roto drawn around an object in one view can be correlated to create a corresponding roto around the same object in the other view. As discussed above, transforming from one view to the other is unlikely to involve a simple horizontal shift – both paint stroke and roto are likely to need to be deformed in order to conform to the geometry of the scene as it would appear from a different location in space. O_Correlate uses the disparity field for the two views (as produced by O_Disparity Generator) to apply the transformation required to translate a paint stroke, roto or other positional effect from one view to the other as accurately as possible, thereby helping to minimise the amount of extra effort required to composite in stereo.

4. Novel View Generation

More complex effects can also be translated automatically from one view to the other, using the view simulation capability provided by O_NewView. The disparity information from the original, unedited views can be passed to O_NewView and used to copy all or part of a transformed image to the other view in a manner consistent with the geometry of the scene. This ensures that the correct stereo effect is accurately preserved in the enhanced version of the scene.

As the name suggests, O_NewView can also be used to create an entire new view from one or both of the originals. For example, it could be used to create the "cyclopean" view of a stereo scene – the view that would have been seen by a single camera positioned midway between the left and right cameras. To create this view – in the normal case where both of the original views are being used – the left and right views would be warped towards each other by half the amount of the stereo disparity between them. The resulting new view would then be a blend between these two warped images. O_NewView provides a choice of blending modes, including "Occlusions" and "Sharp Occlusions" which detect areas that are not visible in both images and deal with them intelligently. "Occlusions" mode will bias the blend towards the image that is most likely to be correct in such regions, while "Sharp Occlusions" will take data only from that image. In a similar way, the disparity information can be used to generate an intermediate image at any point between the two original views. It is also possibly to simulate one view from the other, for example in order to copy an edited part of the scene to the correct position in the other view (as described above).

5. Error Correction

Stereoscopic footage is generally shot with one of two main camera configurations: parallel or converging. Each have their advantages and disadvantages [1] and some stereographers have strong views on the subject of which is the "right" configuration to use [2]. The converging method is more akin to the operation of the human visual system, where our eyes converge to focus on an object of interest, and might therefore seem the more natural choice. The views from parallel cameras do not converge, so the desired convergence distance must be set in post by applying a horizontal shift to one or both images[7]. When converging cameras are used, the convergence can still be adjusted in a similar manner, but is likely to have already been set to the desired value when the footage was acquired. However, although footage shot with converging cameras is less likely to need the convergence adjusted, this method of image acquisition does introduce an additional problem: keystoning. The term keystoning refers to the perspective distortions introduced by the fact that two converging cameras view the scene from different directions, so that their image planes are not coplanar but are angled slightly with respect to each other (see Figures 3 and 4). This introduces some vertical disparity into the pair of stereo views, which must be corrected for in post. While horizontal disparity is what gives us the sensation of depth, vertical disparity is not something our visual systems are used to and should be avoided as far as possible.

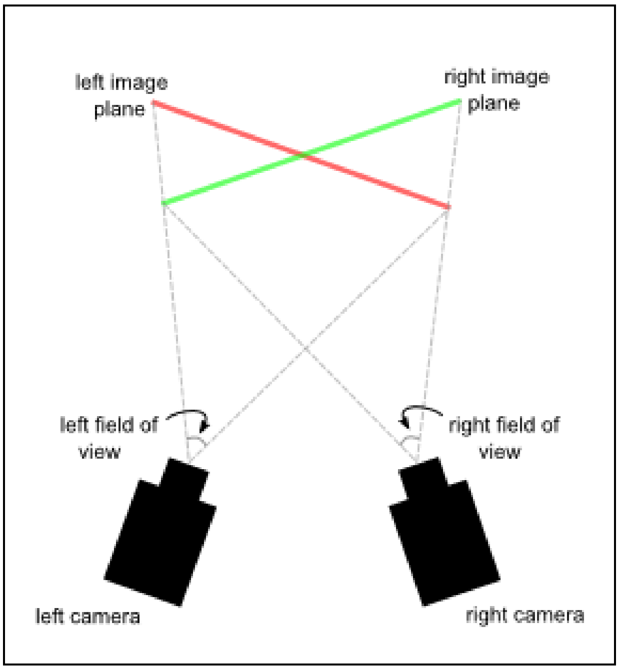


Figure 3: Image Planes of Converging Cameras

7. An alternative to this is to use perspective correction (or shift) lenses, which essentially allow the same correction to be applied during a shoot (see [2]). Such a lens allows the centre of projection to be shifted away from the centre of the image sensor (film or CCD) while keeping the image plane parallel to the sensor in order to avoid perspective distortion [8].

With this in mind, the plug-in O_VerticalAligner was developed in order to allow keystone distortion to be corrected automatically. Unlike the other Ocula plug-ins, O_VerticalAligner does not use the disparity maps generated by O_DisparityGenerator. Instead, it does its own feature detection in each view, calculates the transform that takes one set of features to the other and uses that to work out the amount of perspective distortion present. It then applies a transform to the image which compensates for this distortion and ensures that pairs of features in the two images are aligned horizontally.

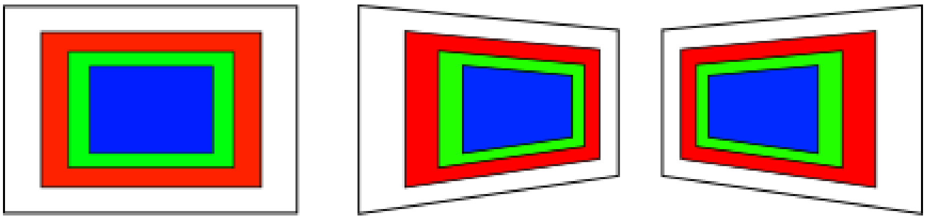


Figure 4: Keystone distortion

6. Colour Matching

Colour discrepancies between the two views of a scene can also make it more difficult for the viewer to fuse objects in the scene or to view it comfortably. Stereo footage is sometimes shot with cameras which are differently polarised; this can introduce an overall colour shift to the resulting pair of images. Slight differences between the physical characteristics of the two camera lenses or image sensors can also be responsible for discrepancies in colour. Correction of these differences in post usually requires skill and can be a painstaking process. However, the automatic colour grading technology behind the Foundry's F_MatchGrade plug-in (included in the Furnace tool set) can help. The algorithm behind F_MatchGrade uses histogram matching techniques in order to transfer the colour distribution of one image to another [9]. This algorithm was designed to be able to cope with extreme differences in illumination – in order to match images taken at different times of day, for example – but is also successful with the more subtle changes that can be apparent between stereo views.

7. Summary

The Ocula plug-ins are designed to make compositing in stereo easier, while ensuring that the final results will provide a comfortable viewing experience for the audience. O_InterocularShifter, combined with Nuke's ReConverge node, allows you to adjust the depth and convergence of a scene in order to adjust the 3D effect and ensure smooth transitions between one scene and the next. O_Correlate and O_NewView help to automate the process of copying work done on one view to the other. They also help to ensure that this is done accurately, minimising alignment errors and thereby minimising discomfort for the eventual audience. O_VerticalAligner corrects for errors in vertical alignment caused by keystoning, which do not occur in nature and, if left in, could cause problems for the viewers such as eye strain, headaches and nausea. Finally, O_DisparityGenerator provides the fundamental disparity estimation technology on which most of the other tools are built.

8. References

(1) Digital Praxis - Stereoscopic 3D: http://www.digitalpraxis.net/stereoscopic3d.htm
(2) Stereo VFX – Convergence: http://www.stereovfx.com/convergence2.html
(3) "James Cameron Supercharges 3D", Variety, 10 Apr 2008 (http://www.variety.com/article/VR1117983864.html?categoryid=2868&cs=1)
(4) Beowulf Review, 3-D Revolution Productions: http://www.the3drevolution.com/3dreview2.html
(5) Journey to the Center of the Earth Review, 3-D Revolution Productions: http://www.the3drevolution.com/3dreview.html
(6) "How a 3D Stereoscopic Movie is Made", 3-D Revolution Productions: http://www.the3drevolution.com/3dscreen.html
(7) "Stereoscopic 3D Film and Animation – Getting it Right", Computer Graphics Quarterly, Volume 41 (3), July 30 2008 (http://www.siggraph.org/publications/newsletter/volume/stereoscopic-3d-film-and-animationgetting-it-right)
(8) "Perspective Correction Lens", Wikipedia article: http://en.wikipedia.org/wiki/Perspective_correction_lens
(9) "Automated colour grading using colour distribution transfer", F. Pitie , A. Kokaram and R. Dahyot, Computer Vision and Image Understanding, February 2007.