The Challenges of Compositing in Stereo
(and how Nuke plus Ocula can help)

Introduction

3D cinema has come a long way since its early days of gimmick-driven plots and headache-inducing visuals, indelibly associated in the public consciousness with the distinctive red-and-cyan anaglyph glasses needed to view the very first attempts. To a large extent, advances in capture and projection technology have combined to eliminate many of the problems associated with early forays in this area. However, unless care is taken throughout the compositing process, it is still possible to produce a stereoscopic 3D (stereo) film that will prove uncomfortable – or even impossible – for audiences to watch and enjoy. In this document we will discuss some of the problems you are likely to encounter when working on a stereo film and explain how our Ocula plug-ins – combined with Nuke's advanced stereo 3D capabilities – have been designed to assist you in overcoming these. We'll start with a brief recap of how the 3D illusion is created in the cinema, then move on to discuss the ways in which it can be broken or become strained – and what Nuke and Ocula can do to help.

Stereo 3D

Outside the cinema, our visual systems infer depth in the environment around us from differences between how our left and right eyes perceive the world. These differences arise from the eyes' separation and are most pronounced on objects that are close to us, and which our eyes consequently view from slightly different angles. When we look at distant objects, however, the distance between our eyes (our "interocular" distance) is small compared to the distance to the object; this means our eyes' views of them are effectively the same. They will also appear at almost the same position in each eye's field of view, unlike nearby objects.

What this means inside the cinema is that we can fool our visual systems into perceiving depth on a flat cinema screen by presenting a slightly different picture to each eye. These days – as in 3D cinema's 50s heyday – this is usually achieved by displaying each eye's image using differently polarised light; the audience must then wear polarising glasses, which deliver the correct image to each eye while filtering out the alternative view. The most comfortable viewing experience – with the least chance of causing eye strain or headaches – would be achieved by ensuring that the differences between views conform as closely as possible to what we would see in real life, so as not to tax our visual systems overmuch. However, in practice this would be impossible, due not only to physical limitations such as the size and position of the cinema screen but also to the creative elements needed to tell a good story, as well as the need to deliver an exciting experience to the audience – which, after all, is what they come to the cinema for. For example, the most realistic representation of a scene could be achieved by shooting with two cameras separated by the same distance as the average person's interocular, then displaying them so as to preserve this similarity. Unfortunately, in most situations this would also make for quite a boring 3D film. Imagine, for instance, a landscape shot whose most interesting feature is the range of distant mountains in the background. Because of the distance, shooting this scene with the camera separation ("interaxial" distance) equal to the average interocular distance would make the two views of the mountains near-identical. As a result, when projected onto the cinema screen they would look completely flat – realistic this may be, but none the less disappointing for anyone who has come to the cinema anticipating a spectacular show.

Instead, in order to maximise the impact of our mountain landscape, the interaxial separation of the cameras used for the shot will often be increased. Instead of the normal human viewpoint, the viewer will then effectively be presented with what a giant – with a proportionately greater interocular distance – would see while gazing at the same scene. Seen with the giant's widely-spaced eyes, even the distant mountains will become properly three-dimensional.

The fixed position of the cinema screen is another reason why emulating reality as closely as possible is not always desirable. In real life, as we look at an object our eyes converge upon it at the same time as their focus adjusts to bring it into sharp relief. When we look at a cinema screen, our eyes are always focused upon the screen, which remains a constant distance away. However, as we follow the action around the scene our eyes are often required to converge on a point somewhere beyond the screen – or even, at times, in front of it. Keeping the point of convergence always on the screen – where we are focusing – would be severely restrictive and would act to prevent the 3D effect from complementing the narrative as it should.

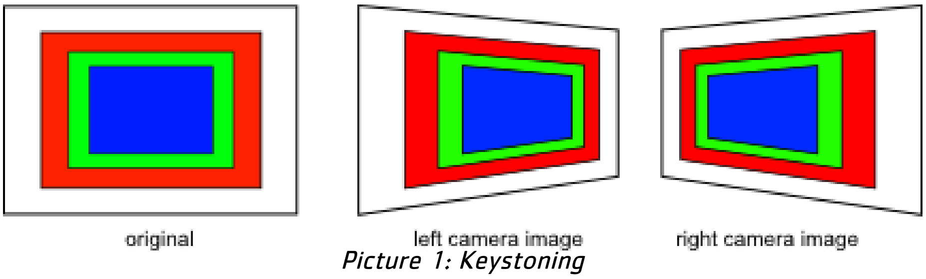
In both cases, the best solution turns out not to be to attempt to duplicate reality, but rather to present the viewer with an enhanced version which makes the optimum use of the 3D effect – a kind of "hyper-reality". Obviously, the more you push the boundaries of what a viewer's eyes expect to see, the harder they will have to work in order to resolve the scene, increasing the chance that eye strain, or worse, will result. Getting the 3D right can be a delicate balancing act.

Let's move on to discuss some ways in which it can be got wrong... and how to correct for them.

Vertical Misalignments

The difference in position of the same point in a scene as viewed by our left and right eyes is usually known as the "stereo disparity" at that point. Due to the horizontal separation of our eyes, most of the time these disparities are almost perfectly horizontal. Vertical disparities, on the other hand, don't generally occur in the real world and, as a result, can be extremely difficult for our eyes to resolve.

Unfortunately, vertical disparities between the left and right images that go to make up a 3D film are, in fact, rather common. Before the advent of digital projection, which allows a single projector to display both images, these were a major source of discomfort for cinema audiences as it was common for the two projectors required to be imperfectly aligned with each other, often causing vertical disparities between the views. However, even now that the projection alignment headache has been removed, vertical disparities can still creep in during filming itself. Most predictably, the two cameras used to shoot the scene might also be imperfectly vertically aligned, or even rotated slightly with respect to one another. Less obviously, there is another way in which the camera configuration can introduce troublesome vertical disparities. 3D films are often shot with the two cameras pointing towards each other, so that their views converge at a point - this mimics the way our eyes would naturally look at an object. When cameras are angled towards each other, though, the physical planes on which the two images are formed – the film-back, say, or CCD in the case of a digital camera – will also be at an angle to one another, as well as to the scene that will eventually be displayed. Viewing the scene at an angle introduces slight perspective distortions, which will not be the same for the two views (see Picture 1). This type of distortion is known as "keystoning" and introduces some vertical disparity between the views – these differences may seem slight, but are more than enough to cause problems when translated to the big screen.



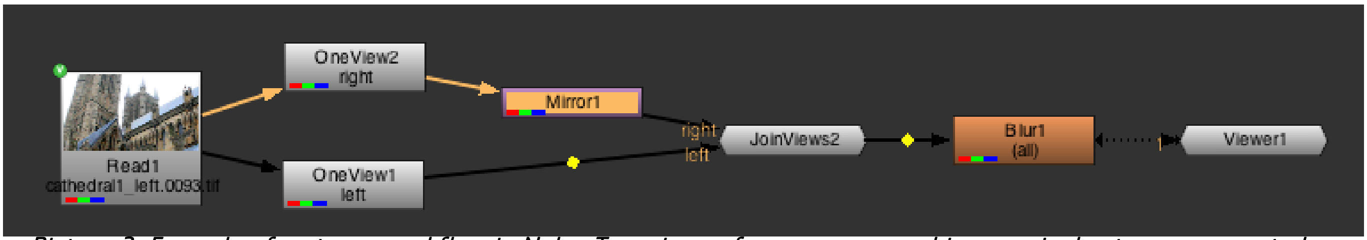
Picture 1: Keystoning

This is where our Ocula plug-ins for Nuke make their first appearance. Ocula is a collection of tools for correcting common problems with stereo footage – the relevant tool here is its VerticalAligner. This first examines the two images to detect matching points in each view, then calculates a transform to apply to one or both views that will make these points line up horizontally as closely as possible. The result is a pair of images in which the vertical disparities have been removed and which is no longer likely to cause headaches in the cinema. Very often, vertical alignment will be the first task in a stereo pipeline; in many cases, Ocula's VerticalAligner can replace a painstaking manual alignment operation with a few simple clicks of the mouse.

Discrepancies between Views

When compositing in stereo, it is essential to maintain the correct correspondence between the two views. Any differences introduced during post production – or which fail to be corrected in the course of it – will make it more difficult for viewers to fuse the two views into the intended 3D scene.

Nuke's stereo workflow has been designed to make preserving the correspondence between views as easy as possible. The left and right eyes' views can be read in as a single image stream and treated as such throughout the compositing process, if desired. This means that a global effect – such as a blur or a colour grade – can be applied to both views in one hit, and will affect both equally. Conversely, in order to execute a position-dependent operation – such as the addition of a CG character to live-action footage – it is important to preserve the distinction between the views in order to get the positioning right. For this reason, Nuke's "Split and Join" nodes allow the views constituting a stereo stream to be separated, treated individually and then rejoined. They can then be processed further as a single stream, or split again – and later rejoined – whenever necessary.



Picture 2: Example of a stereo workflow in Nuke. Two views of a scene are read in as a single stream, separated so that a transform can be applied to the right view only, then recombined. A blur is then applied to both views simultaneously.

The Ocula plug-ins are also very useful for the precise, view-dependent positioning of new elements, needed to ensure they are correctly represented in the 3D scene. A key component of Ocula is its DisparityGenerator, which can produce a map of pixel-to-pixel correspondences for each pair of left and right views. The left-to-right disparity tells you how each pixel in the left view would need to move in order to create the right view, and similarly for the right-to-left disparity. These two maps are not simply the inverse of each other: some parts of the left image might be occluded in the right view and vice versa. Around the edges of the field of view, there will also be areas which are visible in only one of the images.

Nuke can use a disparity map to copy an element drawn in one view – such as a roto or paint stroke - automatically to the appropriate position in the other view, thereby saving you time as well as ensuring that the duplicated element will be correctly positioned. Due to the difference in viewing angle, the element will also frequently need to be deformed in order to have the right shape in the second view. Nuke can transform the element in two ways: firstly, by simply following the disparity vectors from one view to the other. Alternatively – since, in occluded areas for example, the disparity vectors do not always tell the whole story - its "Correlate using Ocula" option will follow the disparity vectors as before in order to find an approximately corresponding position for the element, then refine this initial guess in order to determine the best position for it in the second view.

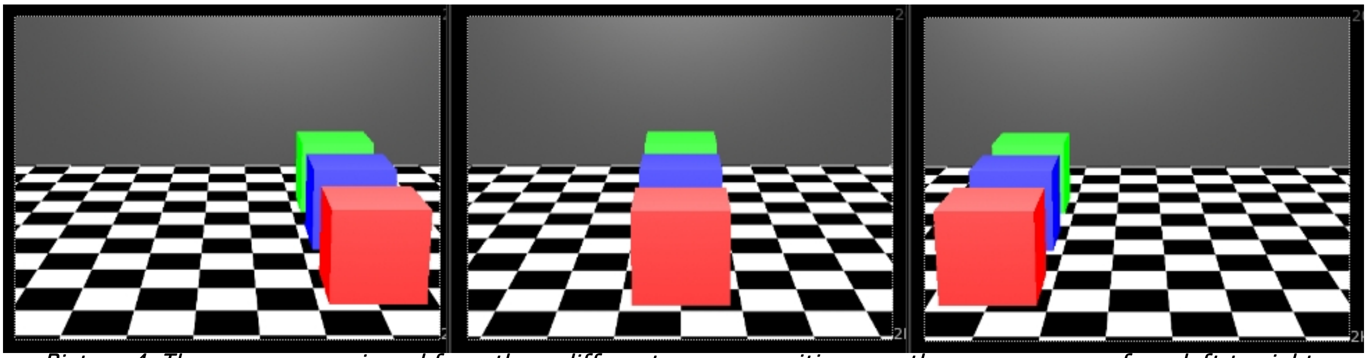
Picture 3: Example of a roto shape in Nuke (drawn around the sign towards the bottom left of the images) which has been copied automatically from the left view of a scene to the correct position in the right view.

Ocula also includes a NewView generator, which can create novel intermediate views at any point between the original left and right. As with the correlation described above, this process begins with the disparity map, but it also deals intelligently with any occluded areas in order to generate the most likely in-between view of the scene. In addition, given a disparity map, the NewView generator can be used to create the left view from the right or vice versa. This can be useful for error correction: getting rid of those discrepancies between views which could affect the ability of the viewer to fuse the two views together and experience them in 3D, rather than as separate images. Suppose, for example, the left camera saw a bright specular reflection at a particular spot which, through the right camera, was barely visible at all. Although physically correct, if left untreated such a large difference in illumination could be distracting or troublesome for a cinema audience. In this situation the NewView generator could be used to transform the right camera's view in order to generate what the left camera would have seen in that region, had the bright reflection not occurred. The corrected region can then be comped back over the original left view to remove the bright reflection.

Depth Grading and Convergence Adjustment

As we touched on earlier, it is the interaxial separation of the cameras which gives a stereoscopic 3D film its apparent depth. Changing the interaxial separation can therefore change the depth of a shot, making the scene appear shallower or deeper depending on whether you bring the cameras closer together or move them further apart. Of course, this is easy to do when the "cameras" are virtual ones being used to render a CG film, trickier to achieve with a real-life camera rig and impossible to change after the event, in post production. (Actually, with Ocula you can get close... but we'll come back to that later.) The main problem is those occluded areas we mentioned earlier.

When a camera is moved in relation to the three-dimensional scene being filmed, different areas of the scene can become occluded or be revealed, as shown in Picture 4. With a stereo pair of images, the majority of the scene will be visible in both views, while some areas around the edges of things will be occluded in one view but visible in the other. If you then wish to shift the left view towards the right one, for example, not only will the angle at which the left camera views the scene change, but the areas that are occluded will also change: some parts of the scene which were not previously visible in the left image will now be seen. These effects will be most noticeable close to the camera, and will decrease as the objects imaged get further away.



Picture 4: The same scene viewed from three different camera positions, as the camera moves from left to right.

So, if the director does decide the depth of a scene needs to be changed in post, the problem is not so much that all the visible parts of the scene need to shift by different amounts: by calculating the disparity between the left and right images, we can actually make a reasonable guess at where everything needs to go. The real problem is the areas of missing information where a part of the scene that was previously hidden becomes visible, and we don't know what the camera would have seen there from its new position. However, we do know what these areas look like in the other camera's view; we also know how the image of the scene surrounding them changes from one view to the other. Again, we can in fact make a reasonable guess at what the image should look like, and this is what Ocula's InteraxialShifter does. The InteraxialShifter can be used to reduce the apparent depth of the scene by bringing the two views closer together, as if the cameras used to shoot the scene had been closer together than they actually were. Because of the occlusion problem – as well as the fact that the estimated disparity can never be a hundred percent accurate – the two new views will not be perfect, but they can be produced quickly and easily and will usually require only a small amount of manual correction.

Unfortunately, even with the InteraxialShifter it is not possible to go the other way and actually increase the interaxial separation of the cameras in post. For one thing, we have no disparity information outside the range of the original image pair – although, by extrapolating that disparity backwards, we could probably come up with a reasonable guess, at least over short distances. Again, though, it's the occluded areas that are the real problem: in moving the cameras further apart, you start to reveal areas that are not visible in *either* of the original images. In this case we really do have no information about how to fill in these missing regions; making an intelligent guess at what should be there becomes much more difficult, if not impossible.

Nevertheless, the ability to reduce the interaxial separation in post is not without value when it comes to producing a watchable 3D film. When the depth of a scene changes, people's eyes take a while to adjust to the new scope of the virtual environment. For this reason, fast cuts between scenes of different depths are undesirable, and should be avoided. Although it is usual to plan out a depth timeline for the whole film before shooting – which should include matching the depths of adjacent scenes as far as possible – the order of shots can change during the edit. In this case, it might be necessary to animate the depth so as to bring the depth ranges of adjacent scenes gradually into line. If this is done slowly enough, the audience should not notice it, but will be spared a sudden, discomforting jump in scene scope.

You might also want to reduce depth in scenes with lots of fast motion, where too great a stereo effect can be confusing and make the scene hard to resolve.

The Foundry Visionmongers Ltd, 1 Wardour Street, London, W1D 6PA, UK • Registered in England and Wales No: 4642027 • VAT Reg. No: 945701320
Tel: +44 (0)20 7434 0449 • **Fax:** +44 (0)20 7434 1550 • **Web:** www.thefoundry.co.uk • **Email:** info@thefoundry.co.uk

Picture 5: Example of an interaxial shift using Ocula. Top row: Original left and right views. Middle row: Left and right views after shifting each towards the other by 40% of the stereo disparity. Bottom row: Anaglyph view of the image pair before and after the interaxial shift.

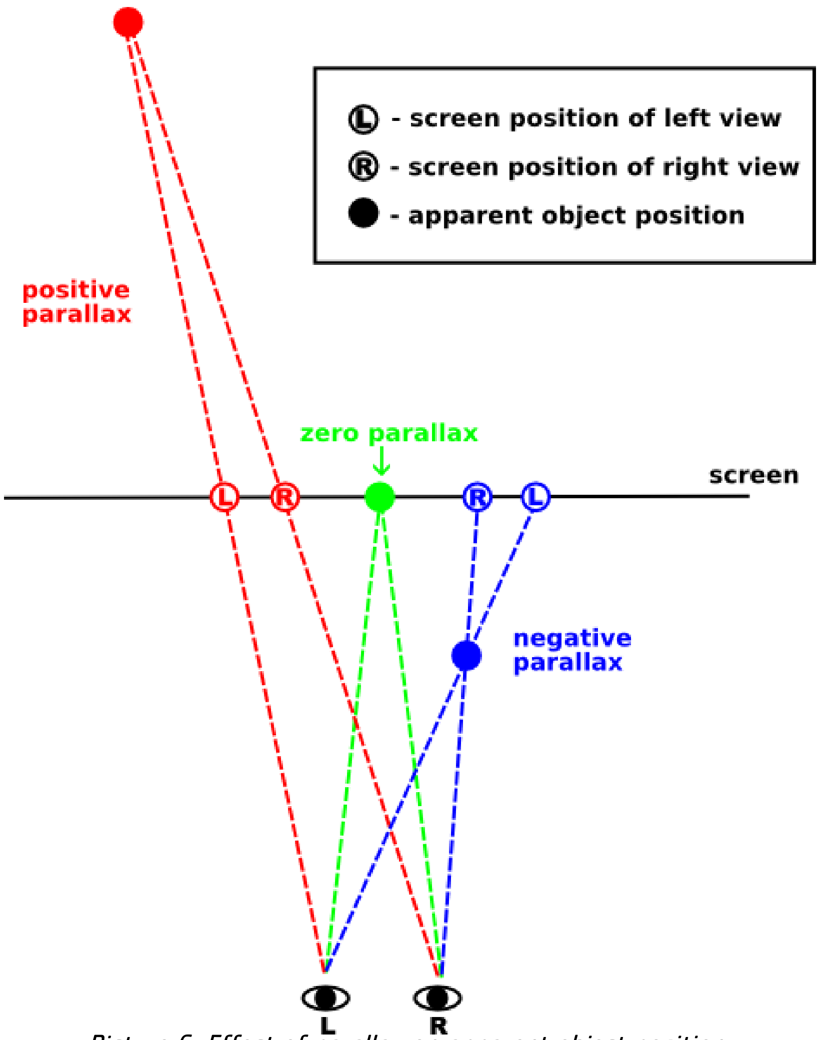
Original images courtesy of Tim Baier, Stereo Consultant (timbaier@vfxgang.com), copyright 2007.

Convergence

As with depth, so with convergence: a sudden shift between scenes can be hard for people's eyes to adjust to. A significant number of such jumps over the course of a couple of hours will make for a 3D film which is very uncomfortable to watch.

When we talk about convergence in this context, what we really mean is zero stereo disparity, or zero parallax. Many 3D films are shot with parallel, rather than converging cameras, which offer side-by-side views of the scene. In this situation, the "convergence", or point of zero parallax, is set in post by sliding the images over one another until the desired focal point of the scene is in the same position in both views, so that the stereo disparity there is zero. When viewed in the cinema, this part of the scene will appear to be at screen depth. Elsewhere, any part of the scene whose position in the left view is to the *right* of its position in the right view (negative parallax) will appear in front of the screen. Any point whose position in the left view is to the *left* of its position in the right view (positive parallax) will appear to be behind the screen, as shown in Picture 5. This means that changing the point of convergence from the front to the back of the scene, say, will suddenly shift the whole 3D scene from appearing behind the screen to appearing in front of it. Such a shift is unlike anything we experience in normal vision and can therefore be difficult for our eyes (and brains) to cope with. As with changes in depth, however, such shifts can be made more palatable by animating the convergence so that the change happens gradually.

In practice, the interaxial separation and point of convergence for a scene are closely linked and it is often desirable to change both together. When changing either or both, however care must be taken to avoid introducing divergence – a sure-fire way to make your audience feel ill!



Picture 6: Effect of parallax on apparent object position.

Divergence

Divergence occurs where the amount of positive parallax between the two views is so great that the audience's eyes would need to rotate away from each other in order to resolve the scene. Obviously, this is not something that occurs in normal vision, and not something that people's eyes are designed to do. Although preferable to avoid it completely, it is generally agreed that a very small amount of divergence – no more than about one degree – can be accommodated by most people and is therefore acceptable for 3D cinema.

This introduces another situation in which it might be necessary to reduce the interaxial separation for a shot. When moving the point of convergence forward, the positive parallax towards the rear of the scene will increase, possibly to the extent of introducing divergence on the furthest objects. If this occurs, the InteraxialShifter could be used to reduce the total depth of the scene and bring the disparity in these areas back within resolvable limits.

Screen Size

Something else that affects the apparent depth of a 3D scene in the size of the screen on which it is displayed. More precisely, the depth the viewer experiences will be determined by the ratio of the screen size to his or her distance from the screen. This ratio tends not to be the same for an artist in front of a computer screen as it would be for a person in the cinema; consequently, these two will not experience the same range of depth if the stereo scene on the computer is simply scaled up to fit the larger screen. In fact, a scene that looks fairly flat and uninteresting on a desktop screen can be much more impressive when transferred to the cinema. However, this also means that divergence can occur in the cinema in scenes where it would not appear be a problem on the smaller screens used for compositing work, so the effect of the screen size must always be borne in mind.

Hyper- and Hypo-stereoscopy

Another reason for caution when varying the interaxial separation is that too great a deviation from the average interocular distance can produce undesirable effects. A very large interaxial separation - "hyper-steroscopy" - instead of merely helping to add depth to distant objects as we discussed earlier, can cause "miniaturisation": it can make a viewer feel as though they are looking not at a real scene, but a scale model of one. This is a result of how the brain copes with the parallax on the furthest objects being much larger than it would normally expect: it reasons that since our interocular distance has not suddenly increased, these objects must therefore be smaller (and closer) than would be the case in real life. Similarly, a very small interaxial distance - "hypo-stereoscopy" - can make the viewer feel as though he or she has shrunk in relation to the scene, so that something which would usually appear small – an insect, for example – will seem gigantic. It's easy to see that these effects could sometimes be advantageous from a creative point of view, and could even be used to tell part of the story in children's classics such as James and the Giant Peach or Alice in Wonderland. However, outside the children's market they will often be distracting rather than helpful to the narrative, so care should be taken to avoid introducing them inadvertently.

"Breaking the Frame"

When something – an actor, say – moves out of shot on one side of the cinema screen, they will disappear from one view before they disappear from the other. This is not always a problem – if the audience's attention is focused somewhere in the centre of the screen, for example, they might not notice what is happening around the edges. If the audience is looking at the actor as they exit, though, it can be confusing when they suddenly drop out of one view and can potentially even break the 3D illusion. The usual solution to this to add black along the edge of one view in order to add a virtual "frame" to the scene and cut off the part of the scene where that eye sees more than the other. An alternative solution from the CG world is to use different interaxial separations for different parts of the scene. In Monsters vs Aliens, for example, Dreamworks dealt with this problem effectively be reducing the stereo effect on characters as they moved out of frame, thus ensuring they remained visible to both eyes for as long as possible. Of course, this is not feasible when shooting live action footage, but under some circumstances it might be achievable in post with the help of Ocula's InteraxialShifter and some clever compositing.

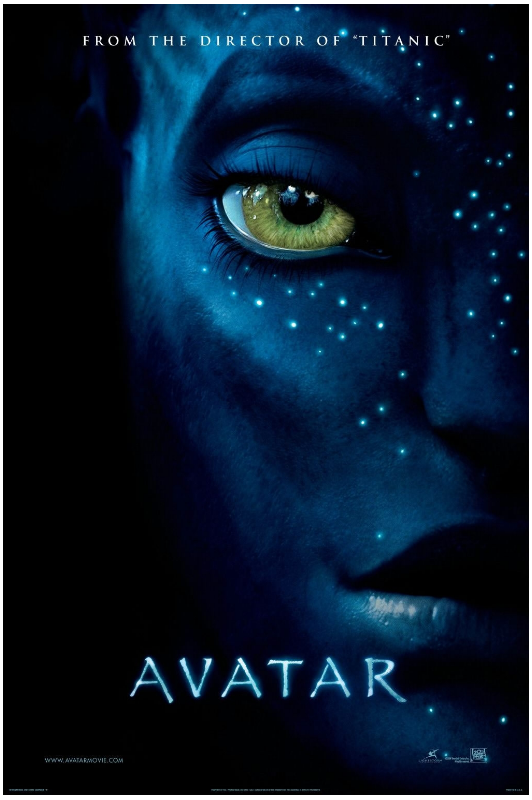
Colour Differences

Colour differences between the left and right views can also make it harder for the viewer to fuse them together into a single 3D scene. The different cameras and lenses used to obtain them – even if, in theory, they should be identical – can produce a slightly different colour cast between the two views that make up a stereo pair. Colour differences are even more of a problem with the beam splitter rigs used to obtain smaller interaxial distances than would be possible with two separate lenses, where the separation is limited by the physical size of each lens. With such a rig, the light used to form the two images is differently polarised, introducing local as well as global colour differences- specular highlights, for example, may be entirely absent from one of the views.

Ocula's ColourMatcher offers two solutions for these two different situations. Firstly, its global colour match will attempt to match the colour distributions of the two views, and can be effective for correcting the slightly different global colour spectra arising from the first scenario. Secondly, ColourMatcher also has a block-based method which is more effective at treating the local – and often more significant – colour differences that result from different polarisations. In this mode, ColourMatcher uses a disparity map to align corresponding pixels between the two images. It then divides the scene into rectangular patches and attempts to match the colour distribution in each patch of one image to the corresponding patch in the other. The patch size can be adjusted as necessary in order to produce the best results. This local method can work very well, even to the extent of reproducing details such as highlights from one view that were not initially present in the other.

Nuke and Ocula in Production

So far, we've covered the theory of what Nuke and Ocula might be able to do for you in the process of finishing a 3D film. At this point, you could be forgiven for asking whether these techniques really work in practice. Well, the answer is yes they do, as clients such as Weta Digital, Framestore and CafeFX will gladly attest. Nuke and Ocula have already been used on several major stereo productions, including The Hole – winner of the inaugural Persol 3D award for the best stereoscopic 3D film of the year at the Venice film festival in 2009 – and James Cameron's eagerly awaited new venture, Avatar.



Both Framestore and Weta Digital composited their Avatar shots in Nuke. In Framestore's case, the sheer complexity of the Avatar work was a major influence in their decision to use Nuke, according to their Head of Nuke Compositing, Christian Kaestner. "Avatar for sure was pushing technology to the limits - not just on the 3D or TD level of the project but also at compositing level," he says. "The number of layers of CG an artist had to deal with per shot - just pure CG layers - was something we had never done before and seeing Nuke deal with it made us very confident that we are looking into something that wouldn't fall apart any time soon. Everything we had thought Nuke was capable of was put into a larger than life scenario with Avatar, and it showed just what Nuke was able to handle."

Weta in particular also made extensive use of Ocula. "The Ocula plug-ins for Nuke have become an impressive and invaluable toolset for us on Avatar," said Compositing Supervisor Erik Winquist. "I wouldn't want to dive into another stereo show without them in the arsenal."

US-based post house CafeFX also chose Nuke and Ocula for their work on The Final Destination. As far as Tom Williamson, VFX supervisor, was concerned, there was really no contest: "Compared to other systems, that are more about brute force and not elegant at all in terms of the workflow, the Nuke workflow is well thought out," he says. "Nuke's View paradigm, and the whole set-up for working on left and right eyes with a single set of nodes that can control each channel independently, yet with shared capabilities, was the killer. We were so impressed that we adopted Nuke and got on the Ocula beta programme."

Another advantage of Nuke is the ease with which its basic functionality can be extended, whether by means of software plug-ins, python scripting or the ability to build up powerful "Gizmos" from existing tools. CafeFX's technical team used the latter capability to develop a Nuke Gizmo that provided compatibility with Zalman's Trimon 3D monitors, which allowed them to interact with the 3D scene in real-time. "We could make adjustments in the composite in Nuke and see the results immediately," Williamson enthuses. "We could also switch between left and right eyes, split off nodes to make individual adjustments, and then playback and tweak the results all in real-time, on-the-fly, in Nuke.

Williamson was equally impressed by his experience with Ocula, particularly its VerticalAligner and the Correlate facility it provides within Nuke. "Vertical Alignment is a fantastic facility of Ocula," he says. "To cure shots where the left and right eyes were misaligned, or gave keystoning effects, we ran the footage through Ocula... Also tremendous is the Correlate plug-in, which substantially reduces the manual labour required by artists doing paint or rotoscoping work. Image manipulations, such as paint, roto and polygon shapes, that you make on one eye are automatically replicated to the other." Not only does this dramatically improve productivity, in his opinion, but it also improves the final result: "Before Ocula, having to do this work separately to both channels was a disaster zone," says Williamson. "The human eye is great at spotting disparities, but with Ocula you can get the disparity down to super-human, sub-pixel levels."

Still from the SoBe LIFEWATER commercial. Image courtesy of Digital Domain.

This view is also shared by Digital Domain, who used Nuke and Ocula to create the 3D SoBe LIFEWATER commercial, shown during the Superbowl in 2009. Jay Barton, the Visual Effects Supervisor for the ad, said that, "The Ocula feature set for Nuke has streamlined our processes for working with stereo imagery. Working in stereo viewing mode with instant access to complex tools for compositing offset image streams and rendered elements has allowed us to provide a much better final product in much less time than previously possible."

Summary

Compositing in stereoscopic 3D presents many new challenges for even the most experienced compositors. Having the right tools for the job – in the shape of Nuke and Ocula – can help you to overcome these problems and achieve the best possible results. Nuke's stereo workflow can remove much of the extra overhead associated with working on two image streams instead of one, while Ocula tools such as VerticalAligner and ColourMatcher can automate previously time- and labour-intensive manual correction processes. In addition, Ocula's stereo disparity generation technology helps to preserve the precise correlation between the left and right views that will ensure audiences will be wowed by your 3D efforts... not pained by them.

The Foundry Visionmongers Ltd, 1 Wardour Street, London, W1D 6PA, UK • Registered in England and Wales No: 4642027 • VAT Reg. No: 945701320
Tel: +44 (0)20 7434 0449 • **Fax:** +44 (0)20 7434 1550 • **Web:** www.thefoundry.co.uk • **Email:** info@thefoundry.co.uk