Movidius

Warning: Objects On Screen Are Closer Than They Appear!

An Insiders' Guide to 3D Imaging for Mobile Systems

November 2010

Myriad 3D

1 Overview

This white paper provides a background on 3D imaging for mobile devices, including some surprisingly non-intuitive problems associated with capturing, storing, and displaying 3D photos and video. It begins with a primer on 3D image capture and manipulation and concludes with a brief description of how the Myriad technology created by Movidius allows handset designers to incorporate high-quality 3D video capture and 3D playback into their products.

1.1 Citius, Altius, Fortius

Just as computer "green screens" gave way to color displays, and glass CRTs were overtaken by flat-screen LCD monitors, so has 2D imaging begun to give way to three dimensions. Display technology moves ever onward, pushing the boundaries of what's possible. Like the Olympic motto, "faster, higher, stronger," new 3D technologies are pushing new products—and the engineers who create them—to new heights of achievement.

This paper looks at 3D technology, especially as it relates to mobile applications, and at how Hollywood content has raised peoples' expectations. At the same time, mobile devices like smartphones and tablet computers have lowered peoples' expectations of price, weight, and power consumption. While 3D movies pack audiences into the theaters, those same moviegoers are packing smartphones that need to display the same sorts of 3D content in a smaller, more portable, and more affordable form. How can a developer reconcile these conflicting demands?

Warning: Objects on Screen are Closer Than They Appear!

1.1.1 The Challenges of Mobile 3D

3D television is here. The first live 3D sports broadcast was from Britain's Sky TV on January 31, 2010. Ever since the Consumer Electronics Show (CES) that same month, there's been an explosion of interest in 3D TVs.

In movie theaters, *Avatar* blew away audiences around the world with its 3D effects. It heralded a quantum leap in 3D movies, from a headache-inducing collection of not-quite-right 3D images to a truly enjoyable and absorbing experience.

In game consoles, Nintendo's handheld 3DS (coming in 2011) displays palmtop 3D images that do not require cumbersome 3D glasses.

Webcams, smartphones, and ubiquitous wireless Internet connections have led to exponential growth in photography and photo sharing. Video recording and sharing followed by only a few months. Can we assume it's only a matter of time before 3D photos and 3D video hit the mainstream?

Probably, but there are a few technical hurdles to overcome first. First, we need to ditch the goofy 3D glasses, but this has already started to happen. Second, we need more compelling 3D content. This has also begun, with networks like ESPN and Sky already producing and broadcasting live sports in 3D. Third, we need 3D viewing to shift from special-event live venues to casual home operations. Again, this transition is already underway, with most major TV manufacturers gearing up 3D TV production ahead of the 2010 holiday season.

Finally, we need a mobile version of the 3D viewing experience, and the massive wave of user-generated content we see with 2D photos and videos. Once mobile 3D recording and playback are a reality, the boom cycle will repeat itself. Allowing the average user to capture, edit, share, and view 3D content using just her mobile device will spur an explosive growth in 3D content, just as digital cameras and smartphones did for photos and traditional video.

But success in mobile 3D *playback* does not have to wait for mobile 3D *recording*. Even standard 2D video (in other words, the vast majority of the video in the world today) can look more interesting and exciting when played on a mobile 3D screen. The secret is converting the 2D video to 3D, preferably in real-time. This, in turn, depends on the quality of the conversion algorithms, which are non-trivial. Do this wrong, and the results look awkward and cheesy, like those early 3D monster movies from the 1960s.

Early attempts at 3D playback were, predictably, a bit raw and uneven. For a time, “3D video” was synonymous with “headache.” Today, however, the technology exists to fix these early teething problems and deliver a first-rate 3D experience.

1.2 Mobile 3D User Scenarios

What does “3D video” mean? It depends; there are many different scenarios and usage cases, just as there are for 2D video or still photos. For example, is the user recording 3D video or just snapping 3D photos? Is he sharing that video (or photos) with a friend? How will that friend display the 3D images on her smartphone? What if her smartphone (or other playback device) is only capable of showing 2D images?

Here’s a table that outlines just some of the usage cases we might encounter, and the technology behind each required to make it happen.

3D Usage Scenario	Underlying Technology
3D still-image capture	Dual camera capture / Camera rectification / 3D Zoom & Stabilization / 3D convergence processing /3D image encoding
3D video capture	Dual camcorder capture / Camera rectification / 3D Zoom & Stabilization / 3D convergence processing / 3D video encoding
3D live preview	Real-time live 3D preview
3D image playback	3D image decode and playback / 3D zoom and pan / 3D convergence adjustment
3D movie playback	3D video decode and playback
2D movie playback with real-time conversion	Real-time automatic 2D-3D conversion
3D gaming	Conversion of 2D games to 3D
3D sharing	Upload and 3D sharing / 3D-2D conversion for 2D compatibility

2 3D 101: How to Create a Three-Dimensional Image

Seeing in three dimensions is the simplest thing in the world for us as humans, but among the toughest things to achieve as an engineer. It's even tougher when cost, power consumption, and silicon area are folded into the equation—and when are they not?

Creating, storing, transferring, and displaying 3D images and videos are all well-understood from a scientific perspective. The trick comes in applying that research to a cost-effective and manageable device for the mass market. The solution begins with an understanding of the fundamentals behind the technology, and of what makes for a good (or bad) 3D user experience.

2.1 How Do We See in 3D?

In real life, we see 3D images (whether moving or still) because our two eyes are both looking forward, and therefore receiving two slightly different images from the same scene. Not all animals have this sort of depth perception, by the way. Many fish and birds, for example, have eyes on opposite sides of their heads and never see the same thing from both eyes. That's one reason birds constantly tilt their heads from side to side.

Our brains are very good at interpreting these small differences in our two-eyed view of the world. We can tell whether an object is near or far, or whether one object is in front of another. Conventional 2D cameras (and pirates wearing an eye patch) have no such mechanism for depth perception.

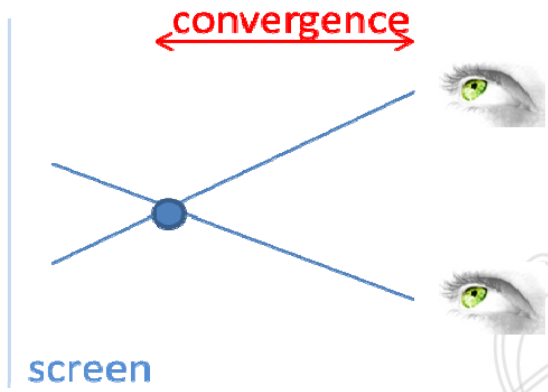
Any scene recorded with two cameras placed side-by-side can take advantage of our natural 3D perception—but only if it's done right. With two pictures and two eyes, the viewer gets the impression of real depth.

2.1.1 Parallax and Convergence

We're all a little bit cross-eyed. Our eyes point slightly toward each other because the two lines of sight cross, or *converge*, somewhere in the distance, called the *convergence point*. (Artists call this the vanishing point.) Our brains infer the distance to an object based on how much our eyes must cross to see it clearly.

Mathematically speaking, the difference between the two images is called *parallax*. Parallax is simply the angle between the two eyes (or two cameras, or two telescopes, etc.) pointed at the same object. Increasing the parallax angle—such as by moving the cameras farther apart—can increase the perception of depth.

In technology terms, an object appearing behind the display screen is said to have *positive convergence*, while one in front of the screen has *negative convergence*. A good 3D scene will include elements of both.



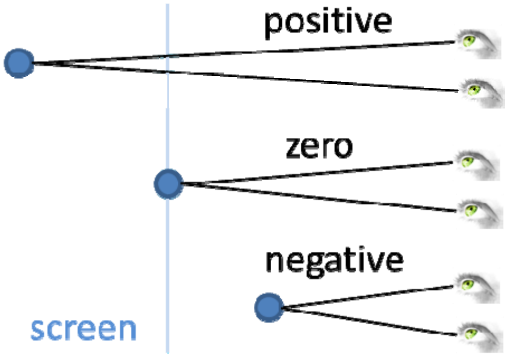
2.1.2 Focus and Accommodation

In addition to parallax and convergence, our eyes also use something called *accommodation* to judge distance and depth perception. Our eyes change focus, or accommodate, to adjust to near objects or distant objects. (As we age, this accommodation may take a bit longer, or might require special eyeglasses.) Just as with parallax, our brains automatically interpret this focusing as depth perception. Focusing on a nearby object cues the brain that the object is close by; focusing farther away sends the opposite message.

2.1.3 Visual Cues and Inference

Between these two natural and automatic reflexes, we perceive depth. But how do we see depth in old-fashioned 2D photographs or videos? How do we tell when something is near or far in a photograph?

The answer lies in the third part of our natural depth perception: inferring subtle visual cues. Whenever the first two methods don't work—such as when we're watching a 2D movie—we rely on our brain's fall-back mechanism. We assume that small objects are farther away than large objects; that distant objects move more slowly than nearby ones; and that close-in objects will block our view of things far away. We also mentally organize objects based on light and shadows. In short, or brains are adept at filling in the missing depth information from a variety of sources, even when only one "eye" is working, such as in a 2D photo or movie film.



2.2 How To Make a Bad 3D Image

That's both good news and bad news. The good news is we can enjoy 2D movie and photos, even though they're missing important depth information present in the real world. The bad news is that it's easy to do it wrong and confuse our brains by delivering mismatched visual cues.

For instance, we'll get a headache if there's disagreement between accommodation and convergence on a 3D display. Our eyes may be looking at a nearby object (the screen), but focusing on objects that appear to be behind the screen. To first-time 3D engineers, this is a neat visual trick that makes objects appear to move off the plane of the display screen. But after just a few seconds it becomes disorienting and ultimately uncomfortable for the viewer. Poor-quality 3D can even be nauseating. Understanding the factors that lead to bad 3D images will help us avoid them.

2.2.1 Convergence Algorithms

One common source of bad 3D images is inferior-quality convergence algorithms. These will produce 3D scenes with too much or too little depth, making the scene unrealistic and causing unnecessary eyestrain as the user tries (and fails) to focus and make sense of the imagery.

2.2.2 Retinal Rivalry

Retinal rivalry occurs when there's an unnatural disparity between the images observed by each eye. These can arise from many different factors, including the display or encoding format (see Section 3.6). The process of resolving differences is known as *rectification,* and is covered in Section 3.3. Retinal rivalry can cause viewer discomfort over time.

2.2.3 Excessive Parallax

As we saw earlier, each eye must accommodate and converge (using focus and parallax, respectively) for an image to be viewed correctly. Achieving the correct parallax angle depends on content, screen size, viewing angle, and distance. If any of these are incorrect, the viewer will experience double vision or blurred images, both of which are uncomfortable and unpleasant.

2.2.4 Safety Limits

Some comfort and safety limits have been proposed, such as maintaining a parallax angle of less than 1.6 degrees [5]. A good convergence algorithm can determine the depth range of the scene, and maintain parallax within comfortable limits.

2.2.5 Stereo Blindness

Sadly, between 4% and 10% of the population is "stereo blind," and will not see any 3D effect. Although for some people this is a treatable condition, for many others it is not, and must be considered an unavoidable accessibility requirement for designers and engineers. [7]

3 Mobile 3D Hardware Technology

In this section we'll examine display technology, camera technology, post-processing (rectification) algorithms, and some extra "gee-whiz" features that can make 3D photo and video viewing more interesting.

3.1 Displays: Lean Forward versus Lean Back

Surprisingly, it's easier to implement good 3D playback on a small mobile device than on a large home TV or movie screen. That's partly because mobile devices generally have only one viewer, and he's very close to the screen. That limits both the distance and the angle from which he'll likely be watching. It also means he's free to move the device around in his hands if it doesn't look right.

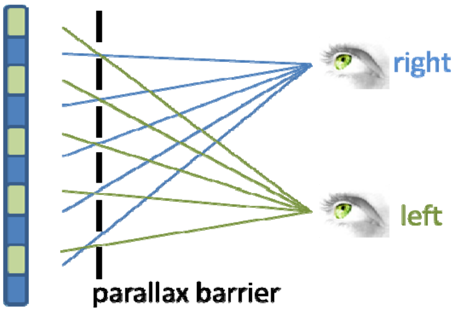
Among other advantages, this means mobile devices don't need 3D glasses. That alone is a powerful benefit to the market adoption of 3D viewing.

How, then, do we display 3D images on a mobile display screen? There are various methods, each with their own benefits.

3.1.1 Handheld Parallax Barrier Displays

A parallax barrier is a physical sheet with regularly spaced holes in front of an LCD display. Its effect is to mask off every other pixel to each eye. The viewer's left eye will see one set of pixels while the right eye sees the other set. If the left and right LCD images are properly managed, the user will perceive parallax, and thus, depth.

The downside of a parallax barrier is that it cuts the effective horizontal resolution of the display in half. (Vertical resolution is not affected.)

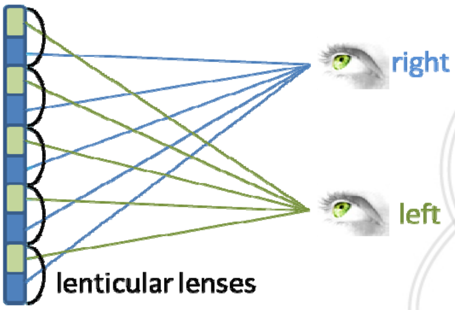


The parallax barrier can itself be an LCD, which allows for an electronically switchable barrier. This solves the "half resolution" problem and is ideal for mobile devices. Switching off the parallax barrier provides full resolution for 2D images, combined with excellent 3D rendering at half resolution. This popular technology is used in the Nintendo 3DS.

3.1.2 Lenticular Lens Arrays

A lenticular array places a grid of curved lenses atop an LCD display. Each lens covers two adjacent pixels. Like the parallax barrier, every even-numbered column of pixels will be visible to the viewer's left eye, while odd-numbered columns are visible only to the right eye.

Unfortunately, lenticular arrays aren't electronically switchable. They must either be permanently fixed to the LCD or manually added and removed to get full-resolution 2D viewing.



3.1.3 Backlight Switching

Instead of a barrier or lens, this method leverages the backlight *behind* the LCD display. Left-eye and right-eye images are displayed in turn, alternating the backlight direction between frames. High frame rates are needed for this to work well, and fast loading of frames to the LCD display buffer is also required. The technology offers good 3D quality, and unlike barriers or lenses, it works at full resolution. This technology is likely to be more popular in the coming years.

3.1.4 Anaglyph 3D Display

Anaglyph displays are probably the most familiar type, and use two different-colored versions of the scene, viewed through 3D glasses with red/green, red/blue or red/cyan colored lenses. This format has the advantage of working on conventional 2D televisions or movie screens, so it's backward compatible with existing playback devices. The disadvantage, of course, is the 3D glasses. Recent advances in this technology have enabled different color pairs with less retinal rivalry and color de-saturation, but these problems still persist.

3.2 Cameras: Eye See, You See, We All See With ICs

How difficult is it to build a 3D camera? Not very, as it turns out—but your viewers might not be happy with the results. As with most things 3D, there are subtle but important physiological considerations to creating an effective and enjoyable 3D experience.

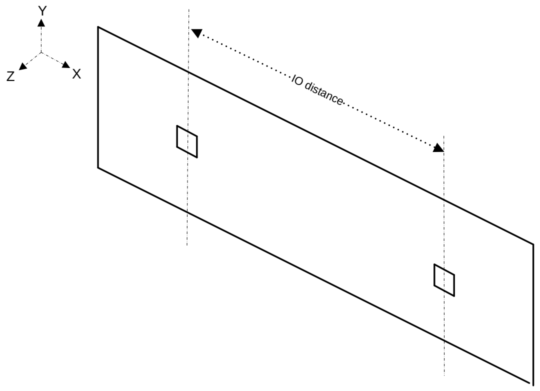


Figure 1: Stereo cameras must be placed coaxially, equidistant from the focus subject, and on the same horizontal plane. In reality, this is difficult to achieve, so the capture device must accommodate and correct for the inevitable three-axis misalignment.

3.2.1 Camera Technology

Cameras themselves, whether still or video, are straightforward. The trick is placing them effectively. As a rule, the two cameras should be coaxial, meaning they should be placed side-by-side and not one above the other. They should also be equidistant from the subject being photographed, so one camera isn't ahead of the other. This is harder to achieve than it sounds, because users can unwittingly tilt or tip a camera without knowing it. For best results, Figure 1 shows how the two cameras should be separated along the X axis, coplanar along the Z axis, and with no significant Y offset between the cameras.

3.2.2 Intra-Ocular Distance (I0)

A primary consideration is the distance between the two cameras. This is called the "interaxial distance," the "stereo baseline," or the "intra-occular" distance (often abbreviated simply "IO"). There's no single best recommendation for this X-axis offset, which affects the capture-depth range and user experience.

The average adult human intra-ocular separation (space between the eyeballs) is about 65mm, so this is a good place to start. It gives the viewer a natural feel when the two camera images correspond to what her eyes would see in the same situation. Using a greater IO value (wider spacing) is called *hyperstereo*, while using a smaller value (closer spacing) is termed *hypostereo*. Excessive hyperstereo or hypostereo tends to give images an unnatural look, known as miniaturization or gigantism, respectively, because the captured depth is disproportionate to the distance between objects.

3.2.3 Camera IO and Depth Range

Looking now along the Z axis, the chart in Figure 2 shows how parallax varies with distance between the subject and the cameras. With a nominal IO distance of 65mm (red line), a subject about 3 meters away from the camera would have a parallax angle of about 2% of the width of the overall image. In other words, the distance from the cameras to the subject is about 50 times greater than the distance between the cameras themselves. As the subject moves away from the camera (or vice versa), the parallax angle decreases. As the subject moves closer, the parallax angle increases.

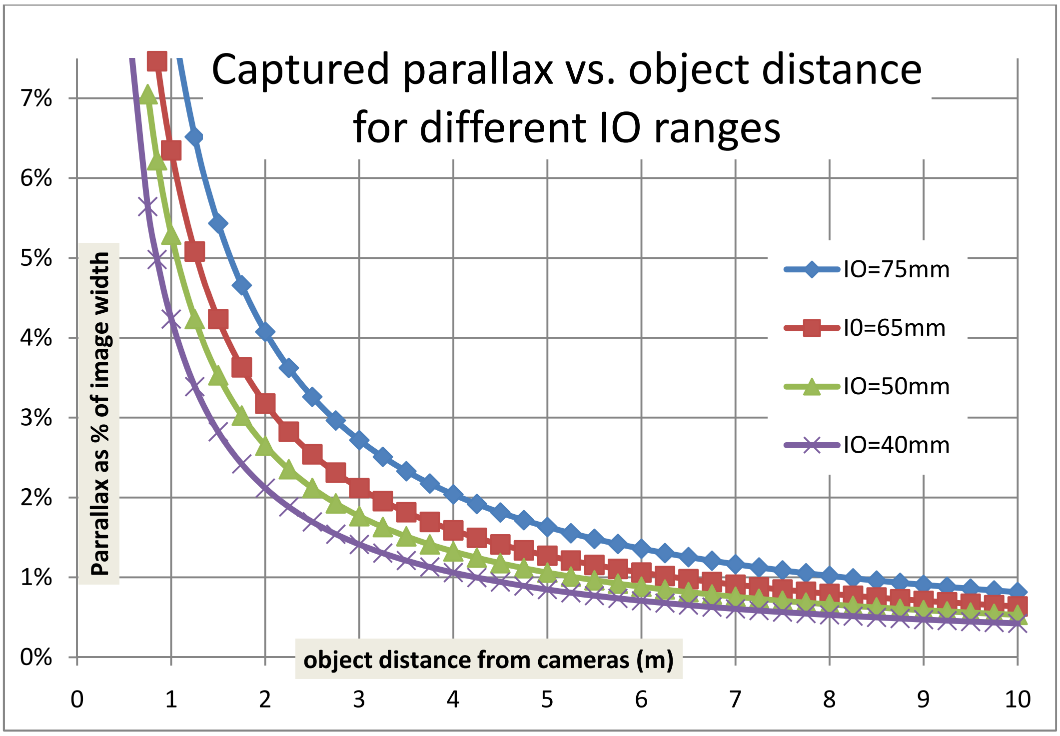


Figure 2: Optimal viewing is achieved with the right balance of IO distance and parallax angle. Too much/too little parallax results in objects at indistinct distances or that appear uncomfortably close.

At the upper limits of parallax, the subject becomes uncomfortably close (the “cross-eyed” effect). That occurs at about 7% of the overall image width, as the chart shows. At the opposite extreme, with minimal parallax, distant objects become difficult to visually sort out; they all appear “far away” but at an indistinct distance.

Decreasing the IO distance (green and purple lines) accelerates this effect; increasing the IO (blue line) delays it. Overall, a short IO distance makes close-in subjects easier to photograph, but at the expense of distant subjects, while a greater IO has the opposite effect.

This useful data allows us to pick the sweet spot for the “near-depth range” somewhere between 1% and 5% of parallax, defined as a percentage of the width of the display. If we stay within this range:

- Objects will be easily discernable at different depths, even on small screens with modest parallax
- Objects at a greater depth may all appear “far away,” but still distinguishable from objects at infinity
- Objects closer in may be difficult to view, though automatic convergence can yield excellent images

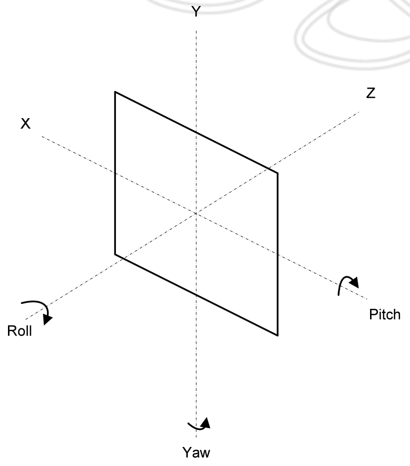
So far we’ve talked about parallax *percentages* not parallax *angles*. The proper percentage depends on the size of the display and its distance to the viewer. Fortunately, these are easily defined for handheld devices. For example, an iPod with a 75 mm screen might typically be held 350 mm from the user’s face.

3.3 Rectification Processing

Capturing 3D images or videos is simple in concept but complex in practice. A number of things can go wrong, some more subtle and fiendish than the others. Correcting for these inevitable problems is what separates good, engaging 3D content from poor, low-quality content. As with most things, image rectification is more easily said than done.

Ideally, all 3D photos or videos will consist of a matched pair of images taken from two well-placed cameras simultaneously. There will always be some differences between the two images due to parallax from the cameras’ two different viewpoints. In fact, there should be differences. That parallax, however, should be the *only difference* between the two images. Any others cause viewer confusion and/or discomfort.

3.3.1 Camera Differences



Mechanical or electrical differences in camera sensors, lenses, or subsystems all introduce sources of disparity between the stereo image pair. Even two "identical" cameras can be different. Bad pixels, different sensitivities, varying gain levels, or other normal manufacturing variations make it effectively impossible to take two identical pictures.

On top of these physical variations, each camera's color balance and/or white balance might be set differently. One image will be more color-saturated than the other, even though both images are taken simultaneously with identical cameras. Something as simple as a shadow falling over one camera can upset its white balance, iris setting, or color saturation compared to the other camera.

Even under ideal lighting conditions, differences in focal length or field of view (FOV) can introduce subtle differences in the relative zoom of an image pair. And finally, both images must be capture *simultaneously*, which isn't always the case when one processor is managing both imager sensors. Software-controlled image capture introduces a delay that upsets the perceptual quality of a 3D pair.

3.3.2 Mounting Differences

As we saw earlier, the two cameras are separated by the IO distance. They must also be level and pointed in the same direction. This may seem obvious but it's harder to achieve than it sounds. Seemingly minor variations in one camera's tilt, pitch, or roll will introduce unwanted (and difficult-to-diagnose) errors in the 3D image. The viewer won't consciously know what's wrong, only that something doesn't look right.

The table below summarizes the six degrees of freedom for mounting a pair of image sensors. Briefly, these are the X, Y, and Z placement of the two cameras relative to each other and the Rx, Ry, and Rz rotation of those cameras relative to each other. As an example, errors in Rx indicate that one camera is tilted slightly downward relative to its mate. This may not be noticeable in manufacturing until the 3D images are assembled.

Error	Meaning	Effect on image
Err_X	X translational error	Very small difference—effectively tolerance on the chosen IO distance
Err_Y	Y translational error	Very small depth-dependant difference
Err_Z	Z translational error	Very small difference in image area
Err_Rx	Pitch error	Can cause significant vertical mismatch between images for small differences
Err_Ry	Yaw error	Can cause extra horizontal convergence/divergence beyond what would normally occur due to parallax
Err_Rz	Roll error	Above 0.5–0.7 degrees of mismatch this becomes quite noticeable

3.3.3 Image Rectification

Fortunately, all of these mechanical and electronic errors and inconsistencies can be cured through the clever application of post-processing software. Some flaws can even be remedied in real-time, either as the photo/video is being recorded or as it's being played back.

This is a huge boon for device manufacturers, who can use less-expensive image sensors and specify less-stringent tolerances for mounting those sensors. The advantages in lower bill-of-materials cost and reduced product scrap are obvious. Moreover, the consumer/viewer gets a better experience, no matter how expensive or carefully engineered the product may be.

3.3.4 Image Area Loss… and Recovery

All rectification loses part of the image. It is an unavoidable side effect of the processing and correction: some portion of the captured image is lost, but usually less than 5%. There are a few approaches to minimize, or even recover, the lost image area:

- Run the sensors at a higher resolution than the capture resolution
- Place a fixed black border around all images
- Use a high-quality zoom algorithm to recover the lost area

The size of the lost area depends on system configuration and tolerances. Rotation errors are the worst; they can significantly reduce the useful picture area. Typical total rectification loss, including rotation, can add up to about 5% of the image area. Maximum convergence area loss will be in the range of 1% to 5% of image width. Maintaining aspect ratio gives a convergence area loss of 2% to 10%.

Some amount of image-area loss is an unavoidable side effect of rectification and convergence. However, the effects can be minimized with a fully featured rectification algorithm.

3.4 Value-Added Features

Image rectification isn't just for fixing errors or unwanted artifacts. It can also be profitably employed as a way to add value to a product or to add flexibility. Some examples are listed here.

3.4.1 Tolerating Mismatched Image Sensors

Both cameras in a 3D image-capture system are usually identical—but they don't have to be. A device maker might want to create a product with two very different cameras with, for example, one 2-megapixel camera and one 5-megapixel camera. The motivation for this is twofold:

- Mismatched cameras can save BOM cost on the second, lower-resolution, image sensor
- The higher-resolution camera can be used for conventional 2D video capture, yielding better 2D video quality, especially in low light

As charming as this design might be, it does introduce some extra engineering challenges:

- The sensors will have inherent differences in field-of-view angle, resolution, pixel geometry, and other characteristics
- The image sensors will have far greater variation in manufacturing tolerances than if they were matched sensors

These extra requirements, on top of the fundamental error corrections mentioned earlier, mean that high-quality rectification is mandatory to deliver acceptable 3D image quality. In particular:

- Color imbalances will be exaggerated, thus requiring extensive post-processing;
- Resolution and field-of-view angles will be mismatched and must be corrected;
- Zoom errors need to be corrected, preferably automatically to allow for wider tolerances

3.4.2 Calibration: Manual or Automatic?

Almost all cameras and image sensors need to be calibrated, whether the user is aware of it or not. Oftentimes this takes the form of a “white balance” button that the user depresses. Other times, the camera may perform its own white- or color-balance procedure without user intervention.

Alternatively, the white balance and/or color calibration can be set once during manufacturing. While this alleviates the need to perform white- or color-balancing in the field, it is detrimental to image quality. An assembly-time procedure also adds time (and therefore, cost) to the assembly process.

Ideally, the image sensors would white- and color-balance themselves as needed. This allows for fine-tuning over the lifetime of the product and enables the product to compensate for environmental factors (heat, cold, impact, etc.). Finally, automatic calibration removes this burden from the user (who is often uninterested or unqualified to perform this function).

3.5 Convergence Processing

“Convergence” is the process of merging two images (or video streams) to produce a 3D image or video. It’s the single most important step in 3D image processing—and the trickiest. Done right, the result is an impressive, high-impact 3D effect that doesn’t go overboard or look artificial.

There are two basic types of convergence processing: fixed and automatic. Both entail balancing positive and negative parallax within a scene for optimal viewing.

3.5.1 Fixed Convergence

The simpler approach to convergence is best suited for small IO distances and a fixed convergence setting. This is straightforward to do, and provides reasonable 3D effect for objects that are close to the camera. The downside is that every object beyond about 2 meters will be captured with positive parallax and appear behind the screen. For a scene with mostly mid- to long-range content, everything will appear deep inside the screen, defeating the purpose of 3D imaging.

On the plus side, cameras and camcorders with small IO distances don’t need complex automatic-convergence algorithms, making them simple, but at the cost of unrealistic and unsatisfying 3D effects.

3.5.2 Automatic Convergence

Automatic convergence produces a more satisfying 3D experience because it accommodates all types of captured images. Close-in subjects are rendered differently than far-away subjects, yet everything will appear to be located near to the screen.

Automatic convergence relies on complex feature-based analysis of the scene to determine the appropriate “convergence point” (the point that will be rendered at the front of the screen). This allows for larger IO distances, giving a more natural feel to a scene and providing a greater range of depth.

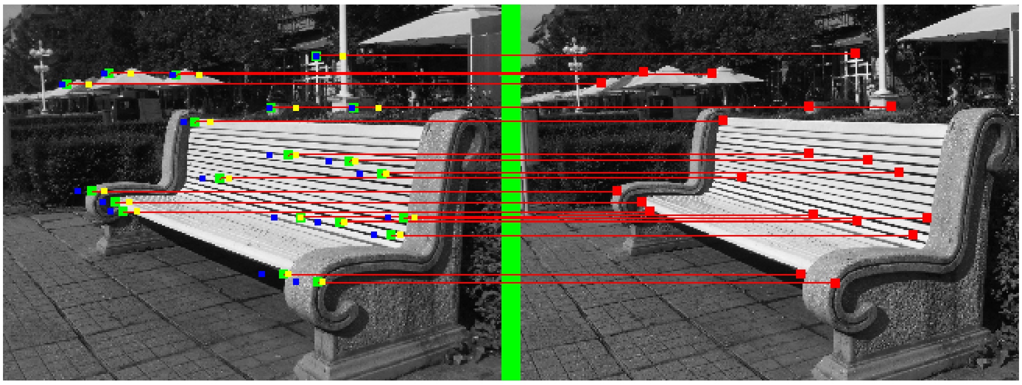


Figure 3: Automatic image-convergence algorithms analyze the scene content and normalize distances to set a “convergence point” in the middle distance for best viewing.

For example, in a scene with a bench set at an oblique angle (such as in Figure 3), the automatic-convergence algorithm will place the convergence point at roughly the center of the bench. The near end of the bench will appear slightly in front of the screen and the distant end behind the screen. Background objects, such as the lampposts and café umbrellas, will appear further behind the screen.

In contrast, a fixed-convergence system would render most of this scene behind the screen—possibly very far behind—resulting in an unsatisfying 3D effect.

Automatic-convergence algorithms can accommodate scenes with any number of near and far objects, and will modify the convergence point accordingly. A scene where the closest object is 5 meters away will still be rendered with some elements slightly in front of the screen and some behind. In effect, automatic convergence “normalizes” distances for optimal viewing.

Convergence processing is just as important for video as for still images. For video capture, the convergence algorithm should work in real-time. It should also adjust convergence *gradually* to avoid any viewer disorientation or discomfort.

A good auto-convergence algorithm should include some preset levels, such as:

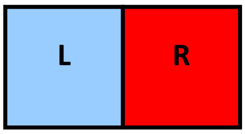
Positive	Keep as much content as possible behind the screen (positive convergence)
Balanced	Capture 70% of content with positive convergence, 30% with negative convergence
Max3D	Capture as much content as possible in front of screen (within safety limits)

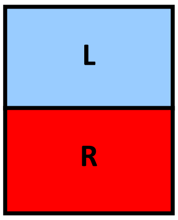
3.6 3D Storage and Transmission Formats

There are almost as many ways to store digital 3D images as there are ways to capture and process them. Here we take a brief look at some of the more popular and practical storage formats.

3.6.1 H.264 AVC with Frame-Packing SEI

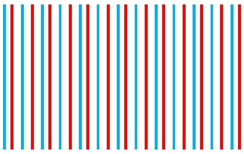
Various schemes allow encoding stereo video within a 2D video frame. The side-by-side frame format (shown at right) is effective, but cuts the effective horizontal resolution in half. A key advantage, however, is its compatibility with existing infrastructure, including YouTube’s preferred 3D video support.





As an alternative, the frame-packed image format (left) allows both images to be stored at full resolution.

The vertical scan-line interleave (right) is commonly used when driving parallax displays. One drawback of this format is that it’s incompatible with chroma sub-sampling as employed in YUV420 or YUV422 video formats. This means that, while it’s commonly used in parallax displays, it’s rarely used in transmission or encoding.



In March 2010 a frame-packing SEI (Supplemental Enhancement Information) extension was introduced to the H.264 standard that allows the particular stereo-packing scheme to be made available at playback time. This helps engineers create systems that support both 2D and 3D video content. [1]

3.6.2 H.264 MVC

This multi-view coding (MVC) extension to H.264 allows for multiple views to be encoded within a single compressed bit stream. Key advantages of this approach are that it’s compatible with 2D players and delivers very efficient encoding (typically only 30% larger than 2D content). [2]

3.6.3 HDMI 1.4

In early 2010 the HDMI consortium publicly released a draft version of its 3D standard formats. This standard is independent of encoding formats, specifying transmission formats over an HDMI physical interface. Full-resolution frame packing and side-by-side formats (both described above) are supported, along with a number of others. [3]

3.6.4 Still Image Formats

For still images, the Multi-Picture Object (.mpo) format has gained significant traction in recent years. It has largely replaced the stereo-JPEG format (.jps), which was popular in the past.

3.6.5 Summary & Key Recommendations

For the time being, H.264 AVC side-by-side is the best 3D format for user-generated video content, for a number of reasons. Specifically, it:

- Reuses existing codec infrastructure for both capture and playback
- Has the same transmission bandwidth as 2D video
- Is the supported format for 3D video upload on YouTube
- Is supported by HDMI specification 1.4a
- Has no resolution loss when used with handheld parallax-barrier displays
- Is technically as good as, or superior to, competing half-resolution 3D formats [4]
- Is popular (used by Fuji Fine Pix, Cable-TV)

In time, H.264 MVC will likely gain traction, aided by its full-resolution display support and bandwidth efficiency. In a market with changing standards, there's an advantage to using programmable devices that can be updated to accommodate the newest standards without changing silicon.

3.7 2D to 3D Conversion

While 3D technology is clearly gaining momentum, the world has a 100-year head start in generating 2D content. It will be many years before the volume of 3D content rivals that of 2D, which means 2D-to-3D conversion is a valuable feature.

Such conversion allows any conventional 2D video to be converted to 3D video viewable on a stereoscopic display. For handheld devices, performing this conversion in real-time is important. Users will want their content converted to 3D on the fly, as it is being viewed.

- 2D content will be more prevalent than 3D for the next several years
- Real-time 2D-to-3D conversion will be key in accelerating that transition

4 Hardware Solutions

So far, we've discussed 3D technology in abstract, hardware-neutral terms. Now it's time to look at concrete solutions to these perplexing problems. One company, Movidius, focuses its efforts on producing hardware and software specifically for improving 3D on mobile devices. The company combines expertise in video processing with proficiency designing mobile silicon, with the aim of improving the end-user experience.

Movidius currently has four different chip-level products based on the Myriad platform for 2D applications, and have established Movidius's reputation for cost-effective silicon for mobile applications. The single-chip MA1133, based on the Myriad 3D platform extends that functionality into three dimensions.

As the block diagram in Figure 4 shows, the Myriad 3D chip handles both the input and the output, for both still and video, in both capture and playback modes. It interfaces with the mobile device's host processor, effectively acting as an accelerator or coprocessor, relieving the host processor of these duties. Myriad 3D is programmable and has its own internal processing engines, making it upgradable in the field. Movidius supplies all the software the Myriad 3D chip needs, however, so developers aren't required to create any 3D code on their own. It's truly a single-chip solution, and it's already sampling to developers working on advanced video for forthcoming mobile phones.

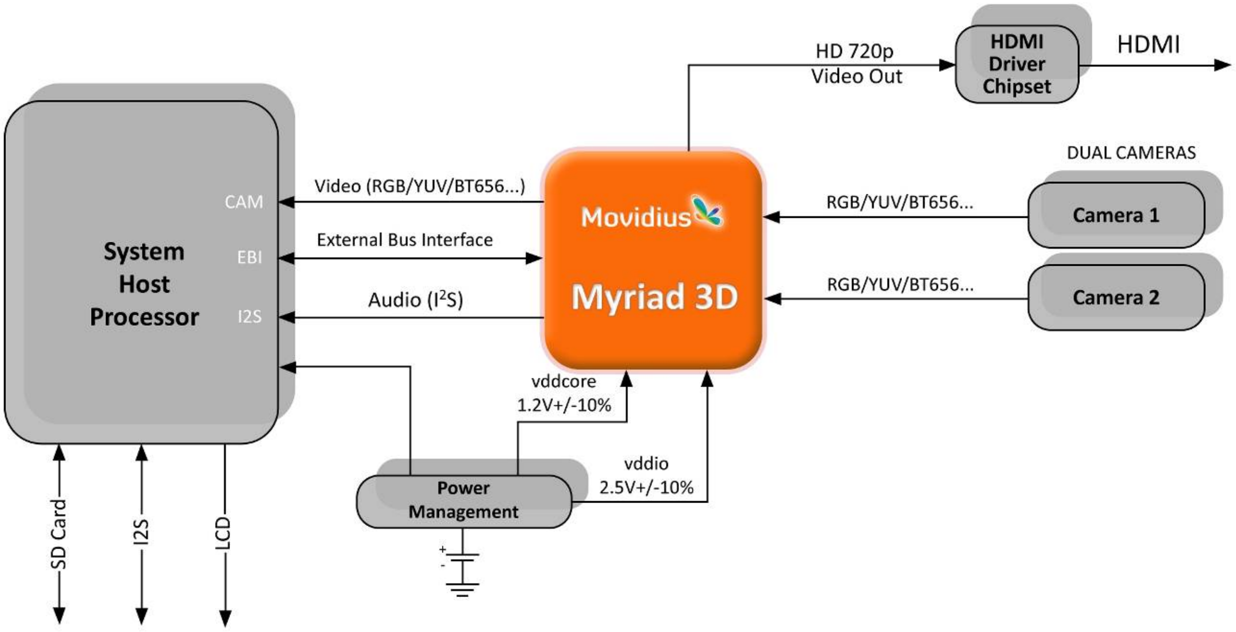


Figure 4: The Myriad 3D chip from Movidius interfaces to a pair of still and/or video cameras, an HDMI display, and the host processor, effectively offloading all 3D processing from the host.

4.1 Movidius Myriad 3D

As simple as the Myriad 3D chip appears on the outside, on the inside it's anything but. It's designed to automatically handle the thorniest 3D-capture, -conversion, and -playback issues, all without host-CPU intervention. For starters, the chip manages two (possibly dissimilar) cameras or camcorders, at 5 megapixel or 720p resolutions, as shown in Figure 5. It automatically stabilizes shaky video, offers live preview, and encodes in a variety of formats. Post-capture, the images or video are intelligently rectified, including compensation for sub-optimal camera alignment, poor lighting, rotation errors, and color balance.

Using Movidius's advanced automatic-convergence algorithms, Myriad 3D can vary the perceived depth of field, providing the best 3D-viewing experience. Headache-inducing artifacts are eliminated and low-budget 3D effects are avoided. Existing 2D video can be converted to 3D on the fly, and enjoy all the same convergence and rectification advantages of native 3D content.

For viewing, Myriad 3D drives HD 720p video out to a variety of 3D display types, including barrier, lenticular, backlight, and anaglyph displays. All this in a small, 8×8 mm package.

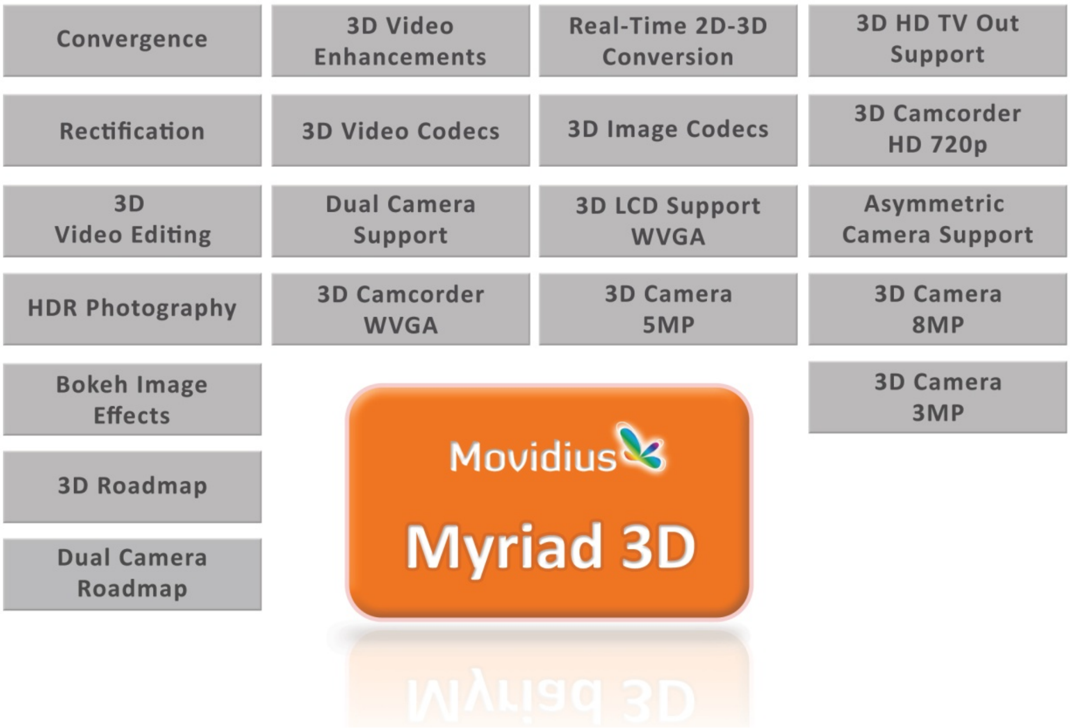


Figure 5: The Myriad 3D Platform enables a wealth of 3D capture, playback, enhancement, and conversion features.

4.2 Total 3D Solution

Like a sprinter bolting out of the starting blocks, a powerful start to the 3D race means a comfortable lead and a victorious finish at the end. A hesitant start means running harder just to catch up. In world-class competition, a slow start spells certain defeat.

Failure in the 3D race comes in many forms: Displaying 3D content with poorly constructed mechanics; using inferior 3D algorithms; battling and removing gross design inaccuracies; improperly handling convergence; failing to account for differences in image sensors; storing and transmitting 3D video in poorly supported formats; and worst of all, failing to live up to users' expectations of the 3D experience. All of these are routes to failure.

The keys to success are a quick start and a competent partner. A tight connection between concept, design, and manufacturing, along with the right components and technology, provides the best starting point for an impressive 3D product. Factor in Movidius' rectification and convergence enhancements and the system BOM is minimized, mechanical tolerances are relaxed, and the user experience is improved.

Easy 3D capture combined with easy 3D playback—even of existing 2D content—and you've got the ingredients for an explosive new product category.

Warning: market opportunities are closer than they appear!

5 References

[1] ITU-T-H.264_201003

[2] H.264 MVC Specification

[3] 'High-Definition Multimedia Interface Specification Version 1.4a Extraction of 3D Signaling Portion', March 2010

[4] 'Dolby Open Specification for Frame-Compatible 3D Systems', Dolby Laboratories, June 2010

[5] 'Foundations of the stereoscopic Cinema', Lenny Lipton, 1982

[6] ' Multipicture Format - CIPA DC-007-2009', Camera and Imaging Products Association (CIPA), 2009

[7] CNET: http://news.cnet.com/8301-19882_3-10435478-250.html

About Movidius

Movidius is a fabless semiconductor company delivering unique multimedia capabilities, including high-definition 3D video. Typical user applications include creating, editing, sharing, viewing, and real-time improvement of multimedia content. Movidius Myriad and Myriad 3D technologies are specially designed for low-power, mobile phone, and consumer products, enabling manufacturers to create highly differentiated products and enhanced user experiences.

Movidius has offices in Dublin, Belfast, Hong Kong and Romania. The company is venture backed, with investors including Celtic House Venture Partners, Capital-E, Emertec Gestion, AIB Seed Capital Fund, and Enterprise Ireland.

movidius.com youtube.com/Movidius Twitter: @Movidius

Copyright © 2010. Movidius Ltd. All rights reserved. Movidius Limited is a company registered in Ireland with offices at 19 Mountjoy Square, Dublin 1, Ireland and registered number 408970.

The contents of this document are provided in connection with Movidius Ltd. ("Movidius") products. Movidius makes no representations or warranties with respect to the accuracy or completeness of the contents of this publication and reserves the right to make changes to specifications and product descriptions at any time without notice. The information contained herein may be of a preliminary or advanced nature and is subject to change without notice. No license, whether express, implied, arising by estoppel or otherwise, to any intellectual property rights is granted by this publication. Except as set forth in Movidius' Standard Terms and Conditions of Sale, Movidius assumes no liability whatsoever, and disclaims any express or implied warranty, relating to its products including, but not limited to, the implied warranty of merchantability, fitness for a particular purpose or infringement of any intellectual property right.

Movidius' products are not designed, intended, authorized or warranted for use as components in systems intended for surgical implant into the body, or in other applications intended to support or sustain life, or in any other application in which the failure of Movidius' product could create a situation where personal injury, death, or severe property or environmental damage may occur. Movidius reserves the right to discontinue or make changes to its products at any time without notice.

Movidius, the Movidius Logo and Movidius' product and product feature names are trademarks and/or registered trademarks of Movidius.

Other product names used in this publication are for identification purposes only and may be trademarks of their respective companies.

Reproduction of this document, or parts thereof, in any form, without the express written permission of Movidius is strictly prohibited.

Feedback

To contribute feedback, please contact Movidius Ltd. Include the following information:

- Myriad 3D, 3D White Paper, Movidius Mobile 3D - White Paper - 30101D