Successfully Creating Stereoscopic-3D for Mobile Devices

By Veera Manikandan Raju, Engineering Manager, Natural User Interface Group, Texas Instruments

Stereoscopic-3D (S3D) is quickly emerging as a prime technology across various markets, and is proving to be a hot trend that adds an additional dimension of reality to existing 2D videos, games, movies and images. With the advent of 3D TVs hitting store shelves, consumers are now getting acquainted to large screen, realistic S3D effects in home entertainment. Today, S3D experiences are migrating from the large screen to mobile devices, providing realistic and glasses-free, personalized viewing experiences on the go.

Overall, S3D video and imaging use cases can be categorized in two ways – S3D content creation and S3D viewing – both of which have a unique set of challenges in mobile design and development. The following article will discuss how next-generation ARM®IP-based architectures, such as the OMAP™ processors from Texas Instruments Incorporated (TI), addresses design challenges, and will share perspectives on how to successfully establish S3D experiences in the mobile world.

What exactly is "S3D"?

It's important to first understand how S3D experiences are created. In general, S3D essentially adds an extra dimension to a viewing scene, using left and right image pairs via two different cameras. In games, for example, S3D rendering refers to a virtual camera position distanced apart, while in S3D videos/images, contents are created using two different sensors which are physically spaced apart.

The human brain is able to differentiate depth perception when both views (left and right seen through the eyes) are rendered together. Farther objects in a given scene are seen at a distance, while closer objects are seen as closer in proximity to the viewer. With the correct level of depth adjustments, pairs of stereo images provide the most realistic and natural user experience. Farther objects are given positive disparity, and nearer objects are given negative disparity. To accurately provide such disparity, a reference object to focus on is necessary. This is called a convergence plane. In addition, human eyes see a field of view (FOV) that is dynamically variable in nature based on where the eyes are looking, giving a very flexible S3D viewing experience at will (see Figure 1).

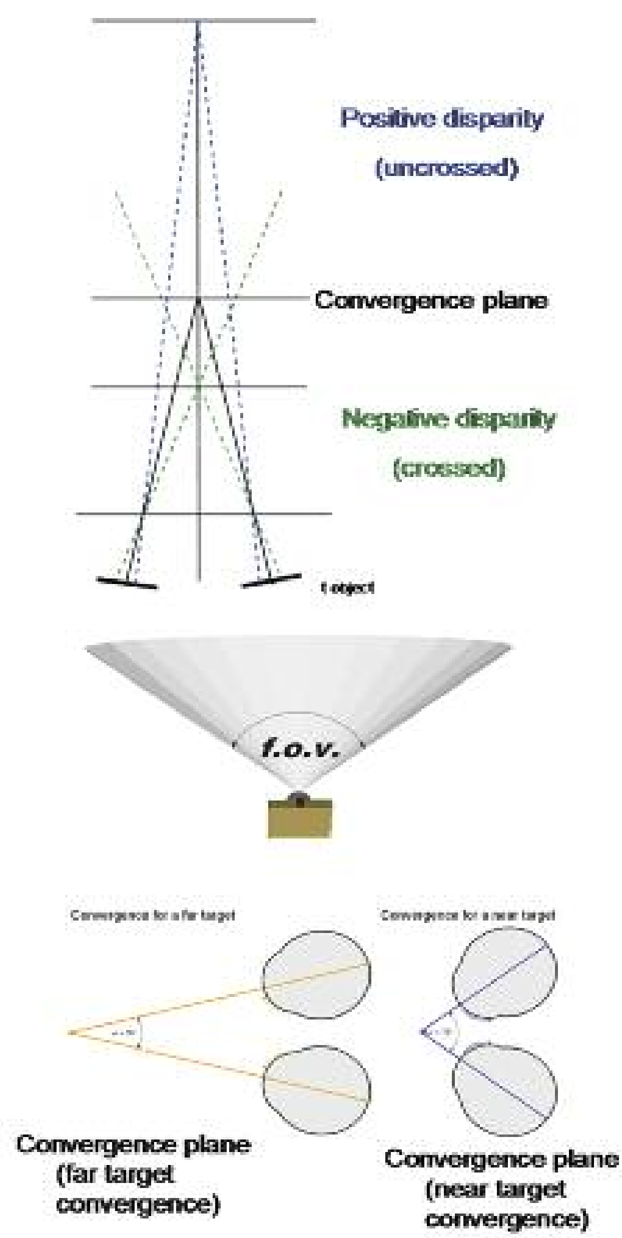

Figure 1

How is S3D content created?

In order to create such a S3D effect, content creation needs to be done with two different camera sensors, and both left and right image pair need to be processed at 60fps (left and right at 30fps independently).

Stereo camera pairs can be positioned in one of two ways when creating S3D images: either in a towing angle, or in a flat angle to achieve the correct FOV. Based on the sensor characteristics, resolution and focal length, a designer will be able to decide how good a recording distance the stereo pair can be used on the target. Positioning of the stereo pair is extremely crucial for getting the right convergence plane. The stereo pair can be positioned at a distance of 65mm (like human eyes) to get a large distance recording experience. In designing a smartphone or other device with similar size attributes, design can consider keeping the positioning at a distance of 35mm, to achieve a personalized recording distance (1m to 3m range).

Such camera pairs, when placed on the gadgets, do not necessarily get mechanically aligned perfectly in translational and rotational directions (see Figure 2). There can be minor misalignment in the millimeters while placing the sensor modules on the form factor device. Such minor variations in physical placements in translational and rotational directions can create large misalignment variations in the image plane. This imposes a huge challenge in terms of calibrating the misalignments up-front and correcting the misalignment on a per frame basis while the content is created. Furthermore, a device's mechanical aspects, even temperature variations and occasional falling of the gadget, can create such misalignment between the stereo pair of sensors. It, therefore, becomes vital to correct such variations in real-time.

What is required on the end device to accurately view S3D content?

Once content is created, it is important to ensure it is viewable on desired devices. System software running in the gadget should be capable of doing the following to provide successful S3D content experiences:

- Combine the stereo image pair and process using the Image Signal Processing (ISP) unit for the correct resolution, distortion corrections, image quality tuning and more.
- Decide the convergence plan at run time using efficient algorithms, and create disparity vectors for the stereo pair at run time to provide pleasing viewer experiences.
- Correct for the misalignments in translational and rotational directions at run time between the stereo image pair, and apply the corrections offsets per frame.
- Synchronize the 3A (Auto Exposure, Auto White Balance, and Auto Focus) between the sensor modules and fine tune the image tuning parameters.

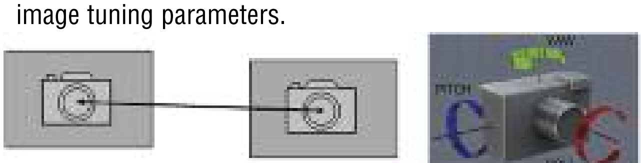

Figure 2

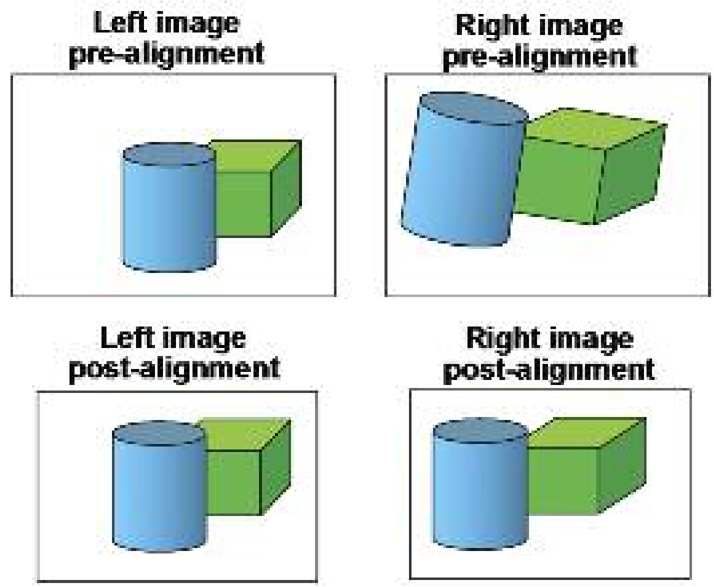

Figure 3

These operations require very sophisticated, real-time hardware acceleration capabilities, which are achieved through combination of next-generation ARM cores and dedicated hardware accelerators that run the sophisticated mathematical computations. The OMAP processor architecture, for example, includes dedicated hardware blocks – controlled through ARM cores – that help these functions perform in an effective fashion.

Now that the hardware is in place, what's next?

Utilizing convergence and misalignment corrections, processed image pairs are passed to the video accelerators of the application processors to encode data in 3D formats. Today's H.264 codec offers an extension to process the S3D information using Supplementary Enhancement Information (SEI), which describes the format and layout of the encoded S3D scene. Emerging standards like Multi-View Codec (MVC) allow designers to encode more than two views for true S3D effect using multiple views. MVC codec correlate the left and right view pair for spatial predictions and motion estimations for effective bit rate savings while encoding. Utilizing the information between the left and right pair for effective bandwidth reduction can improve the system data usage while experiencing an S3D video conference, for example, since users are limited by network bandwidth.

Video encoders and decoders have the S3D awareness based on the content layout. The left and right image can be formatted in multiple ways [side by side, top bottom, interleaved (column/row) and more]. Based on the formatted layout gathered, information is decoded and provided back to the display's sub system for rendering the data in stereoscopic fashion.

Running such video encoders and decoders at twice the frame rate of 2D HD content requires extremely high processing capabilities, which are typically achieved through dedicated hardware accelerators controlled by the ARM® cores. Video and graphics processing mandates the top-notch performance of devices powered by the ARM Cortex™-A9 and Cortex-A15 cores, which are made specifically to handle computational intensive operations.

Generating S3D experiences: What does it take?

Stereoscopic viewing experiences can be generated in multiple forms. Two of the most popular ways to view S3D are through LCD shutter glasses, and auto stereoscopic LCD panels. Shutter glasses achieve S3D experiences by rendering 50 percent of the rendered pictures for the left, and the other 50 percent for the right eye. The technique called "time-sequential multiplexing" then alternately displays left- and right-eye images every time the computer refreshes (draws) the screen.

Turning the left and right glasses using the sync signals generated from the TV creates an S3D effect for the users. It is important to realize that synchronizations need to happen very fast (faster than can be perceived) to ensure that a user thinks he or she is seeing true S3D. This needs immense processing power at the display sub system of application processors especially when you are dealing with high-definition (HD) video.

Termed as “glasses-less 3D," auto-stereoscopic LCD panels display multiple views on the LCD panel. Examples of auto stereoscopic displays include parallax barrier, lenticular and time-sequential LCD panels. Parallax barrier placed in front of the LCD consists of a layer of material with a series of precision slits, allowing each eye to see a different set of pixels, creating a sense of depth through parallax. The viewing angle in parallax barrier LCD is limited to create good quality S3D effect and it also reduces the resolution of the pixel count by half in horizontal direction. For example, half the pixel count is seen by the left eye and the rest by the right eye.

Lenticular displays use two-dimensional arrays of lenslets designed so that when viewed from slightly different angles, a S3D effect is created. Time sequential LCD panels utilize a S3D film, (creating an angular view of light flow through the film), in front of the LCD, controlling the backlights placed on either side of the LCD at 120Hz refresh rate to create 3D viewing experience for the users. 3D film based time sequential panel produce full resolution S3D experience unlike the parallax barrier LCD panels where you lose half resolution.

In mobile devices, auto-stereoscopic panels are becoming popular, and they need extensive display processing capabilities at pixel level to format and create a S3D viewing experiences in real time. The display processing has to be effective at column/row/pixel interleaving for HD resolution stereo pairs at 60fps.

S3D viewing quality poses many challenges and the quality varies with respect to the size of the LCD screen and at which angle the user is viewing the content. It is important that S3D content created solves convergence and misalignment corrections, and creates the appropriate level of disparity in the video. If this is not done effectively, the viewing experiences can irritate human eyes. Research is ongoing in this space with respect to disparity corrections, depth-grading and scene ramping (changing disparity based on the scene pattern changes) to provide positive viewing experiences to end consumers.

Computational power needed to run such content-creation algorithms and pixel-level display processing sub-systems requires application processors to emerge competitively in terms of meeting the needs of S3D HD capabilities. ARM IP-based architectures such as TI’s OMAP™ processors give devices the immense processing power to provide pleasing and natural viewing experiences to users, adding the additional dimension that S3D will be known for in the future. Keep an eye out for S3D-enabled mobile devices, coming your way this year!

END