Image I/O / Creating spatial photos and videos with spatial metadata

Article

Creating spatial photos and videos with spatial metadata

Add spatial metadata to stereo photos and videos to create spatial media for viewing on Apple Vision Pro.

Overview

iPhone 15 Pro and Apple Vision Pro can capture spatial photos and videos, a new type of stereoscopic content that immerses people in a moment and gives them a greater sense of presence in a scene.

Spatial photos and videos are stereo media with additional spatial metadata:

- A *spatial photo* is a multi-image HEIC file containing a left-eye image and a right-eye image, a stereo pair group, plus spatial metadata.

- A *spatial video* is a QuickTime movie with a stereo MV-HEVC video track, plus spatial metadata.

Add spatial metadata to stereo photos or videos created by your app to prompt Apple platforms to consider that media as *spatial* instead of just stereo. Doing so enables presentation of that media in two complementary viewing modes on Apple Vision Pro that provide alternate viewing methods for the same content:

- By default, spatial photos and videos present in a window, with the stereoscopic content inset.

- People can also choose to fully immerse themselves in a spatial photo or video by tapping the Immersive button to present the stereo scene at real-world scale.

Both modes contain visual treatments that can help minimize common causes of stereo viewing discomfort.

Create spatial photos and videos from your own media

While people can use devices like iPhone 15 Pro and Apple Vision Pro to capture spatial media, you can also create and write spatial photos and videos in your own apps. For example:

- A stereo camera app can export camera captures as spatial media.

- A video editing app can import, edit, and export spatial videos.

- A 3D game can export spatial screenshots or gameplay videos.

- A 3D rendering engine can render spatial images or video sequences of a 3D scene.

Spatial photos and videos both use the same spatial metadata, but encode it in different ways. For examples of how to write spatial metadata to stereo photos and videos your app creates, check out Writing spatial photos and Converting side-by-side 3D video to multiview HEVC and spatial video.

Specify camera properties

For visionOS to consider a stereo photo or video as spatial, the media must include three properties of the cameras that captured it:

- The horizontal field of view of each camera

- The baseline of the cameras

- The projection of the captured images

The *horizontal field of view* defines how much of the scene in front of each camera is visible across the width of each image. The horizontal field of view must be the same for both cameras.

The *baseline* defines how far apart the centers of the cameras were (also known as the *interaxial distance*). Baselines that approximate the average distance between human eyes (64 mm) produce content that feels "real-world size" when viewed with immersive presentation. However, a wide range of baselines can produce compelling results. For example, the cameras used to capture spatial videos on iPhone 15 Pro have a baseline of 19.2 mm.

The *projection* defines the relationship between objects in the world and pixels in the image. Spatial photos and videos always use a rectilinear projection.

Choose a horizontal disparity adjustment

Spatial media must also define a *horizontal disparity adjustment* (also known as a *convergence adjustment*). This modifies the perceived stereo depth of the scene by shifting the left- and right-eye images horizontally when you present them in a window in visionOS. As the two images shift horizontally, the perceived 3D scene shifts in depth along the z-axis, with a negative value making content appear closer, and a positive value pushing it farther away.

Horizontal disparity adjustment is expressed as a positive or negative fraction of the image's width. A positive value pushes the left image to the left, and the right image to the right; a negative value pushes them in the opposite direction. Each image is shifted by half of the provided adjustment value, so a positive adjustment of 2 percent moves the left image 1 percent to the left, and the right image 1 percent to the right.

When writing your own spatial media, choose a disparity adjustment value that pushes the closest objects just behind the front of the window, to avoid depth conflicts with other content and UI, while maintaining a sense of depth. The ideal horizontal disparity adjustment value depends on the distance of objects from the camera at capture-time, plus the field of view and aspect ratio of the media. A positive horizontal disparity adjustment of 2 percent typically provides a good default adjustment for many types of stereo media.

Decide when to create spatial media

Determining whether to create spatial media in your app depends on the source content and the experience you want to create.

People can experience spatial media immersively at true scale. Consider adding spatial metadata to stereo content that benefits from the viewer being "in the moment," such as reliving a memory at the same size and scale as the original experience.

Spatial metadata needs to reflect the true camera properties. Ensure you know enough about the capture-time characteristics of the left- and right-eye cameras to provide accurate spatial metadata for them.

Camera properties in spatial videos need to be constant throughout the duration of the video. Don't write spatial metadata to a stereo video if the baseline, horizontal field of view, or recommended disparity adjustment of your stereo video varies over time. For example, a movie or TV show that was captured on multiple stereo cameras with different fields of view might not have consistent spatial metadata across its entire duration, and might not be appropriate for spatial presentation. For this kind of content, prefer `VideoPlayer` or `AVPlayerViewController` for stereo playback presentation.

Spatial media is always rectilinear, typically with a field of view less than 90 degrees. If your content uses an equirectangular, fisheye, or other non-rectilinear projection, don't encode it as spatial media.

See Also

Spatial Photos

`{}`   Writing spatial photos

Create spatial photos for visionOS by packaging a pair of left- and right-eye images as a stereo HEIC file with related spatial metadata.