

MÁSTER EN COMPUTACIÓN GRÁFICA Y SIMULACIÓN

2018

Trabajo de Final de Máster

Investigación, evaluación e implementación de
métodos que simulen seis grados de libertad
en fotos y vídeo para Realidad Virtual

Autor: Gregorio Iniesta Ovejero
Tutor: Diego Bezares Sánchez

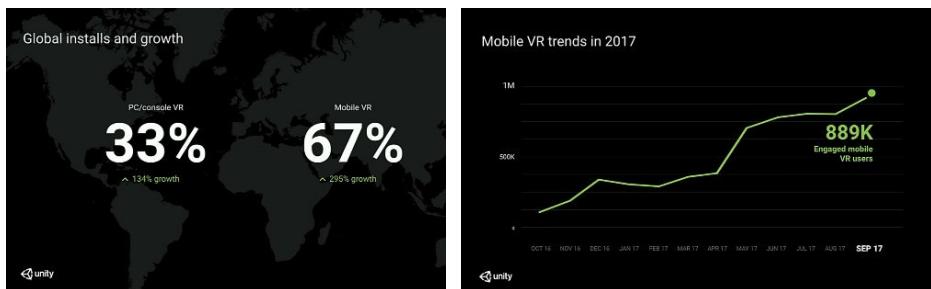
Índice general

1. Resumen	1
2. Introducción	3
2.1. Grados de libertad	3
2.2. Foto y vídeo en realidad virtual	4
2.3. Qué aporta este proyecto	5
3. Planteamiento del problema	7
4. Objetivos	9
5. Estado del Arte	11
5.1. Marco teórico	11
5.1.1. Fotografía y vídeo	11
5.1.2. Mapas de profundidad	15
5.1.3. Fotogrametría	17
5.2. Project Sidewinder	18
5.3. Nube de puntos	19
5.4. Cámara y herramientas de realidad virtual de OTOY y Facebook	20
5.5. Facebook: Seis Grados de libertad con Fotogrametría	20
5.6. "Welcome to Lightfields"	23
6. Desarrollo	25
7. Resultados	27
8. Conclusiones	29
Bibliografía	31

1. Resumen

2. Introducción

La realidad virtual es una tecnología que desde hace unos años ha estado intentando hacerse un hueco en la industria del entretenimiento. Unity expuso estadísticas de consumo de realidad virtual durante el pasado año, proporcionando cifras como un 134 % de crecimiento en sobremesa y consola y un 295 % de crecimiento en las instalaciones de realidad virtual en móvil lo que hace casi un millón de usuarios activos en aplicaciones hechas con Unity a finales de 2017. Por último las estadísticas de uso en multimedia superan el 50 % del tiempo de uso llegando a más del 80 % según Facebook [9].



2.1. Grados de libertad

Un factor clave a tener en cuenta cuando hablamos de realidad virtual son los grados de libertad, que definen la capacidad de movimiento a la hora de interactuar. Generalmente se habla de tres y seis grados de libertad. En el caso de tres grados de libertad, hace referencia a los giros sobre el

eje principal (viraje o *yaw*, inclinación o *pitch* y cabeceo o *roll*), mientras que cuando se amplia a seis grados de libertad hace referencia al desplazamiento en los tres ejes. Esto aplicado a las gafas de realidad virtual informa de la capacidad del dispositivo de captar los giros de la cabeza en todas las direcciones (tres grados de libertad) o si además capta el desplazamiento por la sala (seis grados de libertad).

El mercado de dispositivos de realidad virtual se está enfocando hacia seis grados de libertad. La alta gama ya dispone de ellos desde el principio mientras la gama media y baja ya está adaptándose como demuestra Oculus Santa Cruz o Vive Focus.

2.2. Foto y vídeo en realidad virtual

El contenido que se genera todavía esta basado en gran parte en técnicas ya conocidas como reproducción de vídeo y fotos cuya máxima adaptación consiste simplemente en poner una imagen ligeramente diferente en cada ojo.

La característica principal de este tipo de contenido es que cada imagen esta tomada desde un punto fijo en el espacio. Esto provoca una problemática que consiste en que el usuario únicamente tiene tres grados de libertad a la hora de visualizarlo en unas gafas de realidad virtual. Además de esta restricción debido a las técnicas de grabación que se usan, el cabeceo o *roll* provoca ver imágenes duplicadas y puede provocar incomodidad o incluso mareo.

Debido a las estadísticas de consumo del contenido multimedia grandes empresas como Google, Facebook y Disney entre otras, están trabajando en mejorar los sistemas de visualización de vídeo y fotos mediante vídeo volumétrico, campos de luz o fotogrametría.

2.3. Qué aporta este proyecto

Este trabajo toma la demo “Welcome to lightfields” como referencia pero creando un proyecto de código libre. Trata de conseguir una sensación real de tres dimensiones habilitando seis grados de libertad en un espacio reducido de desplazamiento a partir de un vídeo estereoscópico plano en dos dimensiones y su mapa de profundidad. Para ello se hará un paralaje en tres dimensiones píxel a píxel en función del mapa de profundidad y la posición del usuario.

Durante el desarrollo del proyecto se llevan a cabo pruebas con diferentes técnicas y se evalúa la viabilidad en diferentes dispositivos (plataformas móviles y de sobremesa), el realismo del resultado así como la escalabilidad de los métodos.

3. Planteamiento del problema

La fotografía y el vídeo para realidad virtual actualmente tienen muchas limitaciones tanto para producirlo como para visualizarlo. Una de estas limitaciones y del cual este proyecto se ocupa es la falta de libertad movimiento.

La foto y el vídeo sólo tienen tres grados de libertad y de ellos cabeceo o *roll* (inclinar la cabeza sobre los hombros) no funciona como cabría esperar y puede provocar desde mareos hasta ver las imágenes duplicadas. La captura de imágenes tanto reales como virtuales se hace con cámaras que irremediablemente están en un punto concreto del espacio y eso en principio limita el movimiento del usuario.

En mayo de 2017, Google dio una charla aportando que cerca del 50 % del tiempo pasado en Daydream se centra en experiencias de vídeo. Un año después Facebook en el F8 [9], hablando sobre el estudio para el diseño de Oculus Go, puso de manifiesto que el 99 % de los usuarios consumen vídeo y que el 83 % de tiempo utilizado se destina a multimedia llegando a la conclusión que es uno de los casos de uso principales.

Por ello, este trabajo se centra en mejorar la visualización del vídeo con un sistema que permita al usuario desplazarse físicamente dentro de un área reaccionando el vídeo a ese posicionamiento en tiempo real.

4. Objetivos

Este proyecto ha sido creado con el objetivo de investigar, evaluar e implementar software para proporcionar seis grados de libertad en la visualización de contenidos a partir de imágenes planas y su mapa de profundidad aplicado en tecnologías de realidad virtual.

5. Estado del Arte

La realidad virtual actualmente abarca muchas áreas y se puede interactuar con ella de muchas formas. La forma más común y en la cual se va a centrar este documento es en la representación de imágenes mediante un casco o gafas de realidad virtual.

5.1. Marco teórico

5.1.1. Fotografía y vídeo

La fotografía y el vídeo en realidad virtual es un campo muy amplio en el que se pueden encontrar diferentes maneras guardar y reproducir la información.

Estereoscopía/Monoscopía

La monoscopía implica que sólo existe una imagen para ambos ojos y por lo tanto no hay sensación de profundidad.

La estereoscopía es el factor más importante en contenido multimedia para realidad virtual puesto que es el que más favorece la inmersión como se explica en [1]. Consiste en tener dos imágenes que muestran la misma escena desde dos puntos cercanos que representan los ojos (normalmente

a 6,5cm de distancia). Cada una de estas imágenes se reproducen en una de las pantallas de las gafas, de tal manera que el cerebro del usuario se encarga de reconstruir la escena como si fuera realmente tridimensional. La imagen sigue siendo plana y por lo tanto no es posible desplazarse por el entorno capturado.

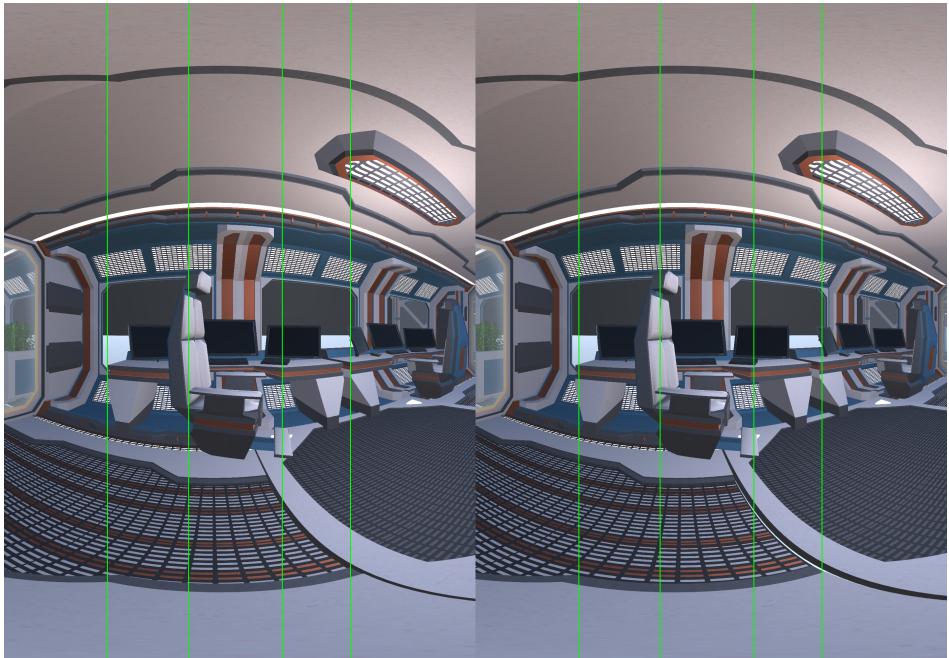


Figura 5.1: Ejemplo de imagen estéreo con líneas verticales para poder comparar el desplazamiento

Es importante tener en cuenta que la separación entre las cámaras definirá la sensación de tamaño. Nuestro cerebro interpreta que si las cámaras están a menos distancia que nuestros ojos, todo parecerá más grande y será más pequeño si ponemos demasiada distancia.

La captura con cámara de imágenes estereoscópicas 360º no es trivial

debido a que se deben hacer las diferentes capturas moviendo las cámaras en una circunferencia teniendo en cuenta el radio de la cabeza. Disney y Google explican como hacerlo en [10] y [13].

Campo de visión

Por lo general se utilizan dos ángulos para el campo de visión.

El más conocido y del que se suele hablar trata 360° de visión, es decir, una foto que genera una esfera alrededor del dispositivo. Es el que proporciona más inmersión. Habitualmente se utilizan varias fotos y se reconstruye la imagen. Este proceso es complicado y si no se hace bien se produce un efecto en el ensamblado que se llama stitching y que se evidencia con unos saltos de color o cortes en objetos. Este problema no es trivial y cuesta eliminarlo como se comenta en [2].

El otro ángulo más utilizado es 180° que proporciona toda la visión frontal pero no hay imagen detrás. Las ventajas que más destacan son que puede ser grabado con cámaras poco especializadas y que se puede concentrar mayor cantidad de píxeles en el frente consiguiendo mayor calidad en la acción destacada.

Proyecciones

Una proyección es una forma de representación de un elemento tridimensional, en este caso una geometría esférica, en un espacio bidimensional.

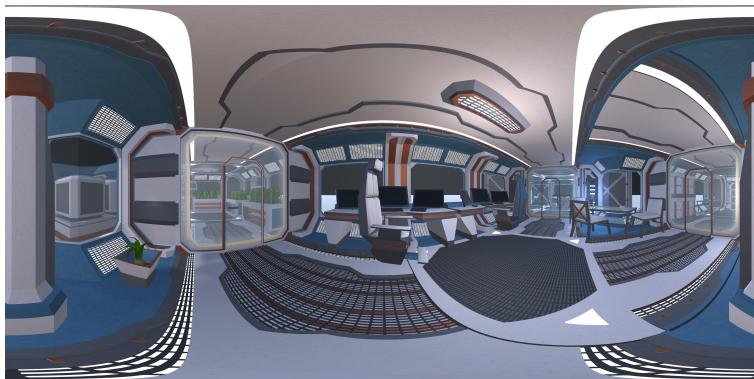
A la hora de guardar o reproducir un contenido multimedia es importante tener en cuenta la proyección a utilizar ya que pueden distorsionar la imagen y saturar regiones poco interesantes, como el cielo, con una gran cantidad de píxeles.

Una de las proyecciones más extendidas y utilizadas en realidad virtual es la equirectangular (5.2a) que coincide con la proyección más utilizada en la actualidad para representar el mundo en dos dimensiones. Esta proyección proporciona demasiada información en los polos que típicamente es el lugar al que menos se suele mirar. Esto hace que gran parte de los píxeles se malgasten. Facebook para mejorar el streaming de video en realidad virtual 360 propone mejoras en [8] que pueden ser utilizadas también en vídeo local como por ejemplo aplicar una distorsión intencionada a la imagen proporcionando más espacio a la parte central de la imagen y compensando esa distorsión en el reproductor.

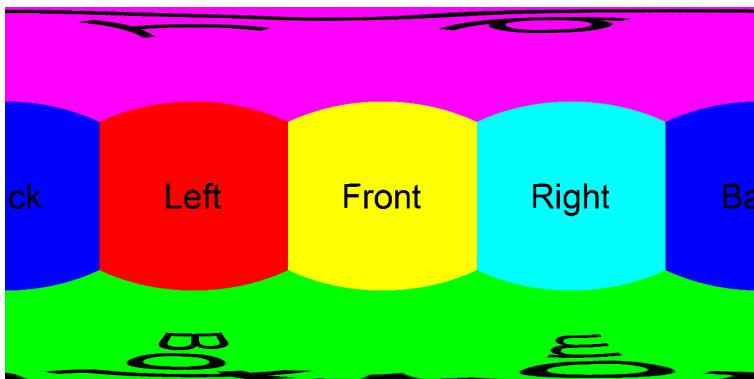
Otra proyección que se utiliza mucho en videojuegos es la cúbica que divide la imagen en seis partes que forman las caras de un cubo. Tiene mayor densidad en las aristas del cubo pero que el porcentaje de píxeles útiles aumenta. Normalmente se le aplica una deformación al cubo dándole curvatura reduciendo la densidad de píxeles para que la distribución sea más uniforme.

Por último mencionar la proyección de barril que se construye como un cilindro. La distribución de píxeles es uniforme en el campo de visión típico. Esta proyección desaprovecha píxeles que se pierden en los huecos que dejan las tapas del cilindro.

Existen una infinidad de proyecciones y cada una tiene una serie de ventajas e inconvenientes.



(a) Ejemplo de proyección equirectangular



(b) Esquema de proyección equirectangular

5.1.2. Mapas de profundidad

Debido a la cantidad de técnicas que utilizan mapas de profundidad o *depthmap* es interesante explicar en qué consisten.

Los mapas de profundidad son imágenes que en cada pixel se encuentra codificada la profundidad de la foto en ese punto. Generalmente se utiliza una escala de grises o de rojos aunque se pueden recurrir a métodos más complejos.[6]

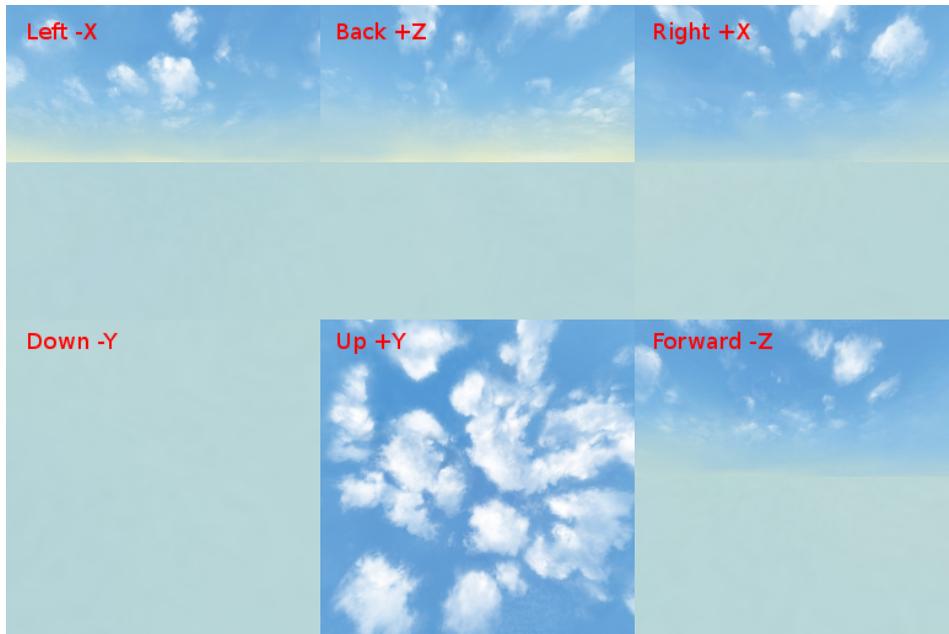


Figura 5.3: Ejemplo de proyección cúbica

En el caso de la escala de grises (5.5), los tonos más oscuros representan elementos en el fondo de la imagen, mientras que los tonos mas claro representan elementos más cercanos.

En el caso de una imagen generada por ordenador, es fácil obtener un buen mapa de profundidad. Sin embargo en el caso de las imágenes captadas por cámaras reales, existe la posibilidad de que la cámara esté preparada o en caso contrario habría que aplicar algoritmos que calculen la profundidad en cada píxel.

Las cámaras que están preparadas para obtener el mapa de profundidad utilizan típicamente la emisión de infrarrojos. Existe una tecnología llamada LIDAR que obtiene mapas de profundidad de alta precisión con un haz



Figura 5.4: Ejemplo de mapa de profundidad 360 en escala de rojos

láser, pero el tiempo que tarda en obtenerlo no lo hace compatible con la grabación de vídeo.

Dentro de los algoritmos que infieren el mapa de profundidad, los más conocidos utilizan imágenes estereoscópicas como el algoritmo BM intenta parear elementos que se encuentren a la misma altura y el algoritmo SGBM que es una variación del anterior añadiendo una ventana de búsqueda para encontrar las correspondencias.

5.1.3. Fotogrametría

La fotogrametría consiste en deducir la ubicación de múltiples puntos en el espacio a partir de una serie de fotografías. Después de eso se reconstruyen los triángulos para generar una malla texturizada que represente de

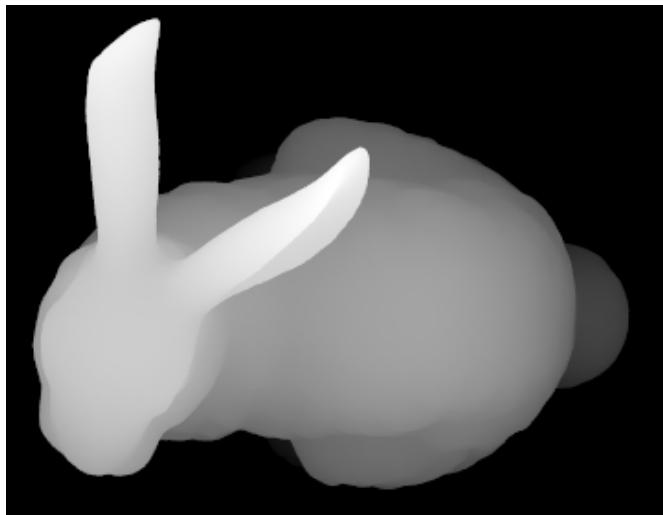


Figura 5.5: Ejemplo de mapa de profundidad cenital en escala de grises

la manera más fiel posible la escena fotografiada. Javier de Matías analiza en profundidad la fotogrametría en [4].

Al igual que en los mapas de profundidad, utilizar imágenes estereoscópicas ayudan a reconstruir la malla con mayor facilidad.

Algunos de los programas más conocidos para generación de mallas a partir de fotos son PIX4D y PhotoScan entre otros.

5.2. Project Sidewinder

Adobe presento en 2017 [3] una demo que utilizaba un *depthmap* de manera muy sencilla para permitir seis grados de libertad dentro de un video

360°. No proporcionan mucha información ya que es una prueba de concepto.

El desplazamiento punto a punto parece correcto pero se ve una distorsión en los bordes probablemente debido al estado temprano del proyecto.

5.3. Nube de puntos

Josh en [6] nos muestra una aplicación de esta técnica creando un punto en el espacio por cada píxel de la foto o el vídeo donde indique el mapa de profundidad.

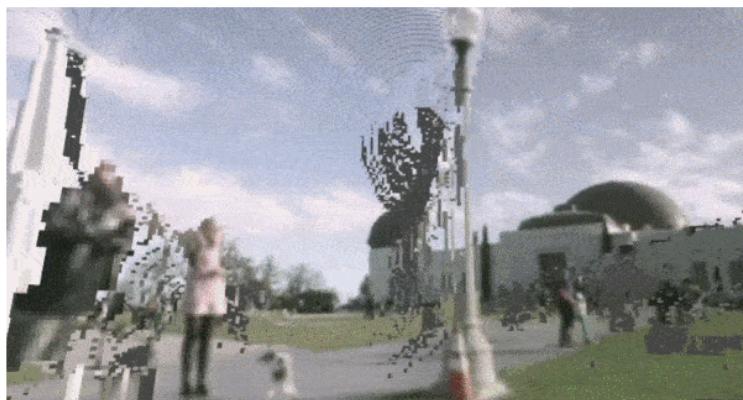


Figura 5.6: Ejemplo de nube de puntos

Esta implementación por contra provoca que aparezcan muchos huecos como se puede ver en 5.5 y generalmente se suele acompañar la implementación con una selección del tamaño del punto como muestra [6] para ver una imagen más sólida.

Otro de los problemas que tiene esta técnica es la cantidad de puntos que deben ser tratados, ya que una resolución *QHD* (2560x1440) requiere cuatro millones de puntos siendo la resolución recomendada actualmente es *4K*. Probablemente el rendimiento sea bajo en equipos poco potentes y en dispositivos móviles.

5.4. Cámara y herramientas de realidad virtual de OTOY y Facebook

En el F8 de 2017 en Los Ángeles *OTOY* y *Facebook* presentaron una colaboración para producir vídeo 360º volumétrico e interactivo a un precio asequible [12]. No hay noticias de 2018 por lo que puede que esté abandonado.

La colaboración consistía en una cámara 360º especializada y una serie de herramientas para procesar el contenido y visualizarlo. El procedimiento implica subir el contenido a la nube de *OTOY* para procesarlo y así reconstruir la escena como una malla tridimensional.

La calidad presentada en las demos era buena con poca distorsión aunque el desplazamiento que presentaban era pequeño.

5.5. Facebook: Seis Grados de libertad con Fotogrametría

Una de las formas de conseguir seis grados de libertad es reconstruir la escena mediante fotogrametría como muestra *Facebook* en [7] para pro-

cesar la imagen y obtener una malla que pueda ser mostrada usando técnicas convencionales.

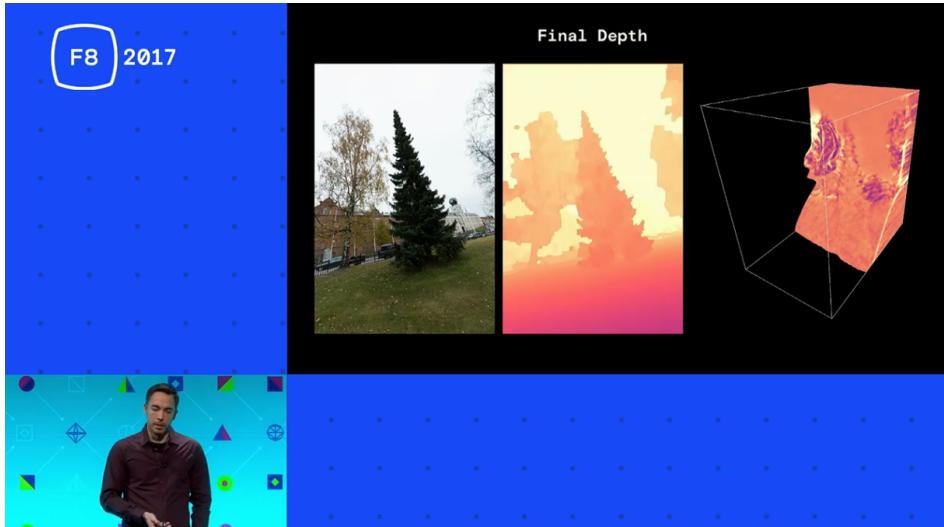
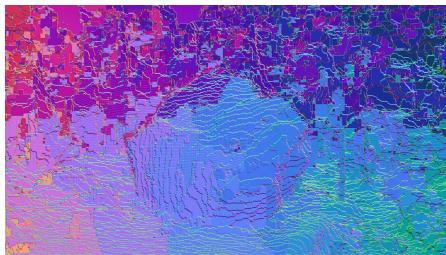


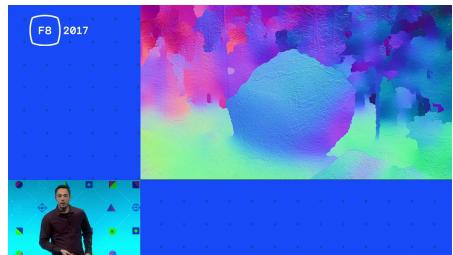
Figura 5.7: Conferencia mostrando un mapa de profundidad de límite inferior

Este procedimiento utiliza una técnica nueva que han llamado “mapa de profundidad de límite inferior” (5.5) que la profundidad de cada pixel debe ser estrictamente mayor que en el depthmap normal. Crear este mapa de profundidad le ayuda al ensamblado de las imágenes para crear la imagen final 360°. Además mezclando el mapa de profundidad con otros algoritmos son capaces de obtener un mapa de normales bastante preciso. 5.8b

Finalmente crean la malla a partir de una nube de puntos. Esta malla como veremos en otras técnicas, tiene agujeros detrás de los objetos que no pueden ser llenados por falta de información. En este caso optan por difuminar de manera sutil las zonas desconocidas dando un buen resultado.



(a) Conferencia mostrando los artefactos que obtienen



(b) Conferencia mostrando los artefactos arreglados

Todo esto lo aprovechan para poder generar un entorno tridimensional con el que poder interactuar y lo ejemplifican jugando con la iluminación 5.9a o incluso inundando la escena 5.9b.



(a) Conferencia mostrando la malla iluminada



(b) Conferencia mostrando un escenario inundándose

Una de las limitaciones que tiene este método es que está diseñado para fotografía en 360° y no se menciona en ningún momento al vídeo 360° por lo que se puede deducir que no está preparado. Por otro lado este tipo de procesado de imágenes requiere una cantidad grande de tiempo.

5.6. “Welcome to Lightfields”

Una compañía llamada Lytro creo un sistema que llamo campos de luz o Lightfields [11] por su similitud con el concepto físico. Más tarde Google se interesó por la compañía comprando algunas de sus patentes e incorporando empleados a su plantilla.

Google continuó el proyecto [5] y construyó un soporte que permite hacer fotografías de una escena desde puntos situados en una esfera.



Figura 5.10: Prototipo haciendo una captura del interior de una cabina.

A partir de todas las fotos recuperadas se calcula la imagen correspondiente en función de la posición del usuario haciendo una interpolación entre diferentes imágenes, eso hace que se recupere una imagen muy fiel a lo que se vería en la realidad.

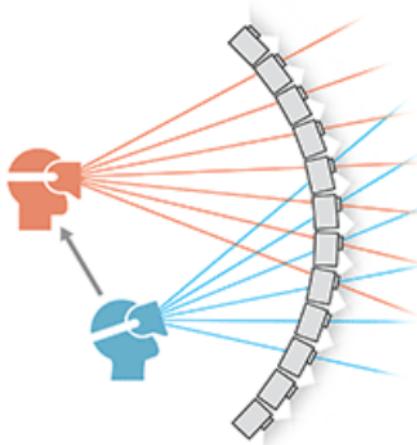


Figura 5.11: Esquema de funcionamiento de los lightfields.

Además esto tiene una implicación que no había sido contemplada hasta ahora que son las superficies especulares. Estas superficies en el resto de técnicas se podían ver con reflejo o no, pero nunca respondería a la posición del usuario. La interpolación de imágenes hace que los espejos reflejen algo diferente en cada posición de la cabeza.

Esta técnica para fotografía es probablemente la mejor en cuanto a calidad de imagen pero sin embargo no se puede utilizar en vídeo debido al tiempo que se tarda en capturar un sólo fotograma. Tampoco permite alterar la luz como se mostraba en la demo de Facebook.

6. Desarrollo

7. Resultados

aasd asd asdafew ferg dfg sg

8. Conclusiones

aasd asd asdafew ferg dfg sg

Bibliografía

- [1] Diego Bezares. Charla gamelab 2016 . estereoscopía en realidad virtual. r.v v.r vr,. URL <https://www.youtube.com/watch?v=70gqUSPTriE>.
- [2] Diego Bezares. Video vr en daydream y plataformas de r.v móviles, . URL <https://www.youtube.com/watch?v=y2mkVQ57-90&t=634s>.
- [3] Adobe Creative Cloud. Projectsidewinder: Adobe max 2017 (sneak peeks) | adobe creative cloud. URL <https://www.youtube.com/watch?v=HSXMs2wnNc4>.
- [4] Javier de Matías Bejarano. *TÍTULO: TÉCNICAS DE FOTOGRAFÍA Y VISIÓN POR COMPUTADOR PARA EL MODELADO 3D DE ESTRUCTURAS GEOMORFOLÓGICAS DINÁMICAS*. PhD thesis, Universidad de Extremadura, 2013.
- [5] Paul Debevec. Experimenting with light fields. URL <https://www.blog.google/products/google-ar-vr/experimenting-light-fields/>.
- [6] Josh Gladstone. Use depth maps to create 6 dof in unity. URL <https://www.immersiveshooter.com/2018/01/23/use-depth-maps-create-6dof-unity/>.
- [7] Facebook Inc. Casual 3d capture, . URL <https://developers.facebook.com/videos/f8-2017/casual-3d-capture/>.
- [8] Facebook Inc. The evolution of dynamic streaming, . URL <https://developers.facebook.com/videos/f8-2017/the-evolution-of-dynamic-streaming/>.

- [9] Facebook Inc. Oculus go: Designing for media and entertainment, . URL <https://developers.facebook.com/videos/f8-2018/oculus-go-designing-for-media-and-entertainment/>.
- [10] Google Inc. Rendering omni-directional stereo content, . URL <https://developers.google.com/vr/jump/rendering-ods-content.pdf>.
- [11] Ben Lang. Lytro is positioning its light field tech as vr's master capture format. URL <https://www.roadtovr.com/lytro-is-positioning-its-light-field-tech-as-vrs-master-capture-format/>.
- [12] OTOY. Otoy and facebook release revolutionary 6dof video vr camera pipeline and tools. URL <https://home.otoy.com/otoy-facebook-release-revolutionary-6dof-video-vr-camera-pipeline-tools/>.
- [13] Christopher Schroers, Jean-Charles Bazin, and Alexander Sorkine-Hornung. An omnistereoscopic video pipeline for capture and display of real-world vr. *ACM Trans. Graph*, page 13, August 2018. doi: 10.1109/TVCG.2018.2794071. URL <http://richardt.name/publications/parallax360/>.