



UNIVERSITAT DE
BARCELONA

Departament de Biblioteconomia,
Documentació i Comunicació
Audiovisual

Facultat de Biblioteconomia i
Documentació

Treball de Grau en Comunicació Audiovisual

Curs 2016-2017

EL VÍDEO 360º MONOSCÓPICO

Conceptos básicos para grabar un vídeo 360º monoscópico

ANDREA DÍAZ AMOR

TUTOR: CARLOS AGUILAR

Barcelona, Junio de 2017

Este obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-SinObraDerivada 3.0 España. <https://creativecommons.org/licenses/by-nc-nd/3.0/es/>



Índex de contenidos

1. Introducción a la realidad virtual.....	3
1.1. El Cubo AIP de Zelter	4
1.2. Qué es un vídeo 360°	5
2. Marco teórico.....	6
3. Desarrollo para la creación de un vídeo 360° monoscópico	9
3.1. Especificaciones técnicas de un vídeo 360° monoscópico.....	9
3.1.1. Proyección equirectangular.....	10
3.1.2. Estándares de resolución y bitrate.....	11
3.2. Calidad de la experiencia de un vídeo 360°.....	13
3.3. Grabar un vídeo 360° monoscópico.....	14
3.3.1. Planificación del <i>rig</i>	14
3.3.1.1. Radio del <i>rig</i> , campo de visión de las cámaras y número de cámaras	15
3.3.1.1.a. El parallax	16
3.3.1.1.b. Distorsión y homografía	19
3.3.1.2. Centro de perspectiva. Punto de <i>no-parallax</i>	22
3.3.1.2.a. Grabar pivotando sobre el punto de <i>no-parallax</i>	23
3.3.1.2.b. Limitaciones del uso del punto de <i>no-parallax</i>	23
3.3.2. <i>Framerate</i> , resolución de los sensores, color	25
4. Análisis para realizar la selección del equipo técnico adecuado	28
4.1. Analizar el entorno	29
4.2. Analizar la acción	30
4.3. Condiciones temporales.....	31
5. Conclusiones.....	32
6. Bibliografía	33
7. Anexo I. Glosario.....	35
8. Anexo II. Figuras	37

Abstract

In the second decade of the XXI century the virtual reality (VR) industry has grown exponentially. The number of production companies and commercial brands that are getting onboard the virtual reality bandwagon is increasing, and they use the latest technology in order to engage the public.

But... How can we get an immersive 360° image? What range of options do we have to obtain them? What is their cost? The desire to get this immersive product is increasing, but the technical needs are not precisely few. We find ourselves in a moment of great development of VR technology. Notwithstanding, adequate tools for creating good VR content barely exist. The lack of specific software, the high cost of production and postproduction support are impediments that seem almost unconquerable in order to popularize the creation of this kind of contents and also their scope. If you work in virtual reality, you will face a new situation everyday which no one has confronted before. The start point of this work consists on the detection of limitations that costs can impose to the production of VR content. From them, standard procedures for making VR videos are identified and the right way to maximize performance is explored. It focuses on how to use cameras in the right way, and get the most out of them in order to get an experience as immersive as possible.

“The best tools currently available for detecting stitching errors are the eyes”

Williams, D. and Burns, P.D.

1. Introducción a la realidad virtual

El concepto de realidad virtual está formado por dos conceptos opuestos. La RAE (2014) considera virtual a aquello que “tiene virtud para producir un efecto, aunque no lo produce de presente, frecuentemente en oposición a efectivo o real” y realidad como la “verdad, lo que ocurre verdaderamente”. Por conjunción semántica, la realidad virtual es aquello que emula la realidad.

A lo largo de los años, la definición de la realidad virtual se adapta a medida que se desarrolla la técnica de captura y consumo. El primero en desarrollar la idea fue Ivan Sutherland en 1965 “[...]hacer que el mundo (virtual) a través de una ventana parezca real, suene real, se sienta real, y responda realísticamente a las acciones del que lo ve” (I. Sutherland, 1965). No será hasta 1987 que Jaron Lanier (VRS, 2016) acuña el término ‘realidad virtual’. A medida que se desarrolla la tecnología, se modifica el concepto de realidad virtual y así, Greenbaum redefine el concepto en la década de los noventa:

La realidad virtual es un mundo alternativo lleno de imágenes generadas por ordenador que son capaces de responder a los movimientos humanos. Estos ambientes simulados se visitan generalmente con la ayuda de un costoso traje de datos que cuenta con gafas de vídeo estereofónicas y guantes de datos de fibra óptica. (Greenbaum, 1992, p.59).

Tan sólo un año después, Cruz-Neira (1993) reflexiona y se centra más en la figura del usuario. Defiende la realidad virtual en referencia a ambientes tridimensionales generados por ordenador que sitúan al espectador en el centro de la acción dándoles contenido inmersivo, interactivo y multisensorial utilizando *tracking* de movimiento¹ y sonido binaural². La NASA (2016) considera la realidad virtual como el uso de la tecnología informática para crear el efecto de un mundo tridimensional interactivo en el cual los objetos tienen un sentido de presencia espacial.

Como puede verse, las definiciones abundan y, aunque existen diferencias entre ellas, todas tienen tres aspectos en común: “la realidad virtual es una experiencia interactiva e inmersiva (con sentimiento de presencia) en un mundo simulado” (T. Mazuryk y M. Gevautz, 1996).

¹ *Tracking de movimiento*: característica técnica que debe tener un casco de realidad virtual para poder rastrear el movimiento de la cabeza del usuario y moverlo a través del espacio virtual que se le presenta.

² Sonido binaural: audio grabado mediante el uso de dos micrófonos situados entre sí a una distancia semejante a la de una cabeza de persona adulta con el fin de crear para el oyente una sensación de sonido 3D como si estuviera en la estancia que se le presente. A diferencia del estéreo, gracias a la distancia entre los micrófonos, existe un retardo del surround dando un sonido final más cercano a lo que una persona podría escuchar en esa situación.

1.1. El Cubo AIP de Zelter

En 1992, Zelter estudia los tres conceptos básicos que hacen de la realidad virtual una experiencia distinta a lo que cualquier otra tecnología digital puede ofrecer: la autonomía del usuario, la interacción que puede realizar con el espacio ficticio y la sensación de presencia del usuario en el entorno que se le presenta (Zelter, 1992).

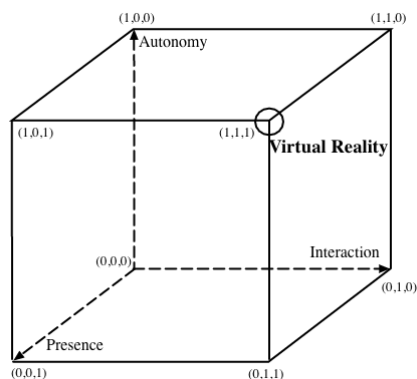


Figura 1. El Cubo AIP de Zelter

- Autonomía.
La autonomía mide la capacidad del usuario para actuar y reaccionar ante el entorno ficticio en el que se encuentra.
- Interacción.
La interacción mide la capacidad del entorno de reaccionar ante las acciones del usuario. El entorno generado digitalmente tiene unos límites. Cuanto menores sean los límites, mayor será la experiencia de inmersión del usuario.
- Presencia
La presencia es la sensación del usuario de estar en el entorno digital que se le presenta.

Una experiencia de realidad virtual debe aglutinar estos tres conceptos para conseguir la inmersión del usuario. El usuario debe ser autónomo, moviéndose libremente a través del entorno virtual. Tiene que poder interactuar con el entorno que se le presenta y el entorno debe reaccionar a sus acciones. Para que la experiencia sea grata, el usuario debe sentir que se encuentra realmente en el mundo virtual en el que es sumergido.

Para que un usuario pueda visualizar este contenido de realidad virtual y conseguir los tres puntos del cubo de Zelter, son necesarias -al menos- unas gafas de realidad virtual (HMD - *head-mounted display*) que habitúan a estar conectadas a unos cascos y otros periféricos que permiten al usuario navegar a través de la experiencia. Los HMD utilizan la técnica del *tracking* de movimiento para permitir al usuario moverse dentro de la esfera sobre la que se proyecta el entorno virtual, pivotando sobre el

punto (0,0,0) de la misma. La clave es conseguir que el usuario sienta realmente que se encuentra dentro del entorno virtual que se le está presentando (eMarketer, 2016).

Gracias a su calidad de presencia e inmersión, la realidad virtual abre la puerta a todo un mundo de experiencias para el usuario. Si algo es demasiado caro, peligroso o sencillamente imposible de practicar, la realidad virtual permite al usuario obtener en primera persona la vivencia de la experiencia. Por ello, se ha desarrollado en una amplia variedad de campos como el entretenimiento, el arte, la ciencia, el deporte y la arquitectura³.

1.2. Qué es un vídeo 360°

La realidad virtual comienza generando contenido creado exclusivamente por composición digital. No será hasta la década de 2010 que se empiezan a grabar vídeos 360°. El motivo principal es que hasta este momento existía una carencia de hardware y de software para la producción y edición de vídeos 360° grabados con cámaras físicas. No obstante, existía un gran interés en desarrollar la técnica para grabar este tipo de contenido, pues se consideraba que la calidad de la imagen sería más realista (y por tanto más inmersiva) si era grabada que si era generada por ordenador. Por tanto, en cuanto hubo la tecnología que lo permitió, se empezó a generar contenido.

Según K. Korrel (2016), un vídeo 360° es “un vídeo proyectado en una esfera alrededor del espectador”. Llegados a este punto, se pueden diferenciar 3 tipos de vídeos 360°:

- Creados por composición digital (CGI).
- Grabados físicamente con el uso de cámaras.
- Los vídeos que conjugan ambas técnicas.

A su vez, cada uno de ellos puede ser -como en el caso de un vídeo convencional- monoscópico o estereoscópico. Este trabajo se centra en la grabación de vídeos 360° monoscópicos. Sin embargo, ya sea mono o estéreo, nos encontramos ante un vídeo plano⁴, por tanto ¿un vídeo 360° cumple con los 3 vértices del cubo de Zelter?

Al no proporcionar al usuario un entorno a través del cual puede moverse con libertad en los 3 ejes, se pierde el concepto de autonomía del usuario. Ergo, ¿se puede

³ Virtual Reality for Education pone a disposición toda una serie de proyectos en los que se ha utilizado la realidad virtual en diferentes ámbitos educativos. Disponible en: <https://virtualrealityforeducation.com/> [consultado: 1 junio 2017]

⁴ Aunque sea un vídeo estereoscópico, realmente el vídeo final es plano, es decir, sólo existen dos dimensiones (alto y ancho) a diferencia de un entorno virtual en el cual también existe profundidad.

considerar un vídeo 360° realidad virtual? La eterna pregunta y perenne discusión. Los expertos del sector no se ponen de acuerdo en si debe ser considerado realidad virtual o no, por lo que no se puede afirmar o negar con firmeza. Lo que sí está claro es que el vídeo 360° mantiene aspectos en común con la VR:

- La preocupación de la VR, así como la de los vídeos 360°, es conseguir que el usuario disfrute de una experiencia inmersiva y tenga sensación de presencia en el entorno que se le presenta.
- Tanto la VR como los vídeos 360° están pensados para ser consumidos con un HMD, con el fin de conseguir una mejor experiencia inmersiva. No obstante, los vídeos 360° también pueden ser consumidos a través de reproductores que permiten al usuario girar en el entorno virtual sin necesidad de utilizar un HMD, como ocurre con el *player* VR de Youtube, Facebook o el propio de GoPro (GoProVR Player).
- Ambos permiten la visión de contenido de 360°.

La conexión entre estos dos conceptos parece difícilmente quebrantable, y es que sin los avances realizados en realidad virtual, ¿qué posibilidad habría para los vídeos 360°?

2. Marco teórico

Intentar rastrear los orígenes de la realidad virtual es como tratar de localizar la fuente de un río, pues es producida por el flujo acumulado de muchas corrientes de ideas, alimentada por muchos manantiales de inspiración. (J. Woolley, 1993, p.37).

Si bien las definiciones de realidad virtual, como hemos visto, han ido configurando su marco conceptual, no es menos cierto que el origen de la técnica para la creación y reproducción de realidad virtual es ciertamente difuso. Teniendo en cuenta que una de las principales preocupaciones a la hora de crear contenido inmersivo para realidad virtual es crear la ilusión de estar en otro lugar, se puede hablar de las pinturas panorámicas del s.XIX como el germen de lo que ocurriría cien años después (VRS, 2016). Estos cuadros eran tan grandes que comprendían todo el campo de visión humano, sin dejar ver al espectador nada más que la pintura. A finales del x.XIX se empezó a investigar la fotografía estereoscópica, con el fin de crear sensación de presencia con la falsa tridimensionalidad de los objetos de la imagen. A mediados del s.XX aparecen los primeros simuladores, que pretendían emular la realidad a través de imágenes y del movimiento del soporte.

El primer intento que puede equipararse al concepto actual de realidad virtual data de la década de 1950, no obstante, no fue patentado hasta 1962. Se llamó Sensorama, fue creado por el cineasta Morton Heilig y era una cabina de uso individual que estimulaba todos los sentidos. Pudiendo disfrutar de una pantalla estereoscópica en color, disponía de sonido binaural, una silla vibratoria, un generador de olores y ventilación. Se le considera un intento porque a pesar de crear el sentido de presencia, no era posible la interacción con el mundo simulado, uno de los vértices del cubo AIP (T. Mazuryk y M. Gevautz, 1996).

Es en 1965 cuando Ivan Sutherland⁵ propone la solución última⁶ para la realidad virtual con la creación del concepto The Ultimate Display, un mundo virtual estereoscópico que permite la interacción de los usuarios con los objetos del entorno simulado a tiempo real (VRS, 2016). Tardó 3 años en volverse tangible. Junto con su alumno Bob Sproull construye el primer hardware de realidad virtual de la historia, que recibió el nombre de The Sword of Damocles debido a su extraña apariencia.

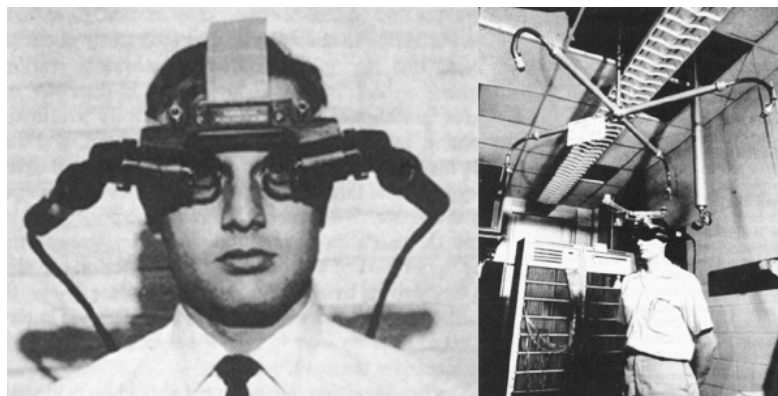


Figura 2. The Sword of Damocles - Ivan Sutherland

A pesar de ser un gran paso para la historia de la tecnología, Sutherland dejó el hardware de lado para centrarse en el software y The Sword of Damocles quedó en el olvido. Era grande y aparatoso, necesitaba una habitación entera para él sólo y era tan pesado que no se podía mover. Los gráficos eran tan primitivos que sacaban al usuario de la inmersión que debe suponer la realidad virtual (McLellan, 1992). No obstante, Sutherland asienta las bases de la realidad virtual que despuntará de nuevo a finales de los ochenta.

⁵ Ivan Sutherland. Informático, ingeniero eléctrico, pionero de Internet, profesor en la Universidad de Harvard (1966) y en la Universidad de Utah (1968) y ganador de múltiples premios como el Computerworld Smithsonian Award (1996) y el Premio Turing, de la Association for Computing Machinery (1988). (Historian, 2015). Disponible en: <http://www.i-programmer.info/history/people/329-ivan-sutherland.html> [consultado: 7 mayo 2017]

⁶ Este término no debe dar lugar a confusión. Es denominado como *la solución última* por el nombre que recibió el concepto (The Ultimate Display) y porque asienta las bases de lo que Lanier llamará en 1987 'realidad virtual'.

Por fin, en 1987 Jaron Lanier acuña el término 'realidad virtual' y así el campo de investigación pasa a tener un nombre. Además, fue el fundador del laboratorio VLP Research (Visual Programming Languages), una de las primeras compañías en desarrollar y vender productos de realidad virtual. En él desarrolla una extensa gama de artes de realidad virtual, como el HMD EyePhone o el Data Suite, un traje de cuerpo entero con sensores para rastrear el movimiento de brazos, piernas y tronco y reproducirlo dentro de la experiencia inmersiva. (VRS, 2016).

La realidad virtual despegó durante la época de 1990. Comienza a ser accesible al público, se abren salas de recreativos especializadas. En 1993 Sega anuncia su primer HMD. Sega VR era un prototipo con *tracking* de movimiento y sonido estéreo. La compañía de videojuegos pretendía sacar el producto a la venta a precio de usuario. Sin embargo, se quedó en la fase de prototipo por complejidades técnicas (Segaretto, 2017).

Le siguió Nintendo dos años después lanzando la Virtual Boy, siendo la primera consola portátil que muestra gráficos en 3D. A pesar de valer solamente 180 dólares, fue un fracaso comercial, pues la tecnología era tosca. Los gráficos eran bicolor, el software fallaba y las gafas no se podían utilizar en una posición cómoda. La realidad virtual volvió a caer en el olvido, la tecnología no estaba preparada para dar el salto, los resultados no conseguían una experiencia inmersiva y su desarrollo era extremadamente caro (VRS, 2016).

A principios del s.XXI se investiga en campos tecnológicos que facilitarán el resurgir de la realidad virtual: aparecen los smartphones, se mejoran las capacidades gráficas en el diseño digital, la industria de los videojuegos avanza en el desarrollo de cámaras con sensores de profundidad y de controladores de movimiento (VRS, 2016).

Hasta este momento, el contenido para realidad virtual era exclusivamente generado través de composición digital. En la segunda década del s.XXI, grandes empresas de tecnología deciden apostar por la realidad virtual y se comienza a desarrollar la tecnología necesaria para poder grabar vídeos 360° con cámaras físicas.

En 2014, Facebook gasta 2 billones de dólares para comprar Oculus Rift (S. Barnes, 2016). En 2015, GoPro decide comprar Kolor, una de las mayores compañías dedicadas a la creación de software para vídeo 360° (O'Kane, 2015). En 2016 HTC y Valve se unen y lanzan su HMD HTC Vive, el cual permite caminar y tiene controladores para interactuar con los objetos virtuales. Llegando a vender 15.000

unidades en los primeros 10 minutos de venta (J. Kastrenakes, 2016). Las compañías de videojuegos no se quedan atrás, especialmente Playstation que lanza en 2016 su HMD PlaystationVR, que permite cierta autonomía al usuario a la hora de moverse. Las redes sociales comienzan a permitir subir contenido en 360°: Facebook, Youtube, Vimeo. Aparecen software de edición dedicados exclusivamente a vídeos esféricos, como PTGui, Kolor, Mettler. Desde el Q1 2014 al Q2 2015, se invirtieron más de 1 billón de dólares en proyectos de realidad virtual y aumentada (eMarketer, 2016). Este auge de la realidad virtual se debe principalmente a dos motivos: la bajada del coste de producción y la subida de la calidad del mundo virtual (S. Barnes, 2106).

Para poder mantener este mercado, es necesario crear contenido del que disfrutar y comienzan a desarrollarse todo tipo de cámaras que permiten grabar vídeos de 360°. El 2014 Google saca al mercado su cámara de realidad virtual Google Jump y su HMD de bajo coste Google Cardboard. Samsung no se queda atrás y lanza su gama Gear, que abarca relojería inteligente, HMD y cámara 360°. Nokia presenta su cámara OZO. Cada vez más compañías deciden dar el salto al desarrollo de realidad virtual, el número de cámaras aumenta exponencialmente: LG Cam 360, Ricoh Theta S, PixPro SP360 4K, Giroptic 360 cam, 360fly 4K, Bublcam... La lista se hace más larga. El mercado está repleto de cámaras para la filmación de vídeos en 360° con diferentes estructuras, características y costes. Todas ellas distintas y cada una con sus limitaciones.

Este trabajo pretende explicar los aspectos básicos de las cámaras 360° para saber qué cámara o estructura de cámaras es la más adecuada en función de las características del rodaje.

3. Desarrollo para la creación de un vídeo 360° monoscópico

A la hora de crear un video 360° habrá que tener presente las especificaciones técnicas del mismo así como del soporte que reproducirá la obra, para poder adecuar el equipo técnico a las mismas.

3.1. Especificaciones técnicas de un vídeo 360° monoscópico

Un vídeo 360° es un vídeo proyectado en una esfera alrededor del espectador (K. Korrel, 2016). Gracias a esta técnica, el espectador tiene la sensación de estar en el lugar de la acción y, por primera vez, tiene la libertad de observar la parte de la escena que desea. Esta es la versión oficial, la que llega al usuario y puede disfrutar utilizando un HMD. No obstante, algunas webs como Youtube o Facebook ya disponen de *players* que permiten la visualización de este contenido sin necesidad de un casco de

realidad virtual, por lo que se pierde parte de la experiencia inmersiva, fundamental para conseguir los tres vértices del cubo AIP.

La visión que tiene el usuario del vídeo no es la misma que se utiliza para editarlo. Un vídeo 360° puede tener diferentes proyecciones. El espectador disfruta de una proyección esférica, mientras que la edición se realiza mayormente en *quirectangular* (proyección esférica equidistante).

3.1.1. Proyección quirectangular

La proyección *quirectangular* corresponde a la proyección plana de una esfera, como ocurriría en la relación de un globo terráqueo con un mapamundi. Si lo comparamos con un vídeo, la fila de píxeles central horizontal del *quirectangular* corresponde con el ecuador de la esfera. La imagen se va deformando hasta llegar a los polos, de tal manera que todos los píxeles de la fila superior (e inferior) se transforman en un único punto, el polo norte (y el polo sur) cuando pasa a la proyección esférica. La proporción estándar del *quirectangular* es de 2:1 (H. Lee et al., 2010).

Como se puede ver en la Figura 3, se contempla la esfera completa, es decir, 360° horizontales y 180° verticales.

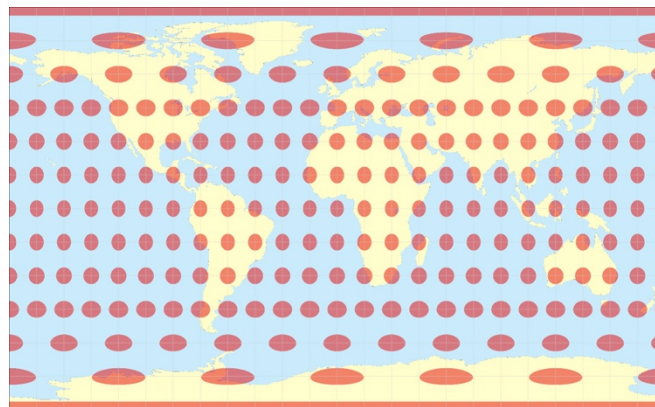


Figura 3. Proyección quirectangular. Esta imagen pone en relación un globo terráqueo con su proyección quirectangular y la deformación que sufren los puntos de la esfera a mitad que se acercan a los polos.

Es habitual ver vídeos que se denominan 360°, pero que no representan la esfera completa sin llegar a dar los 180° verticales. De hecho, no son pocas las cámaras 360° que no proporcionan la esfera completa y gozan de prestigio, como Jump o Ladybug3.



Figura 4. Fotograma realizado por una cámara Ladybug 3.

En este fotograma se observa una banda negra en la parte inferior de la imagen. Esto ocurre porque esta cámara no tiene capacidad para grabar los 180° verticales, dando sólo información de 145° (Ptgrey, 2012). No obstante, esta cámara facilita el trabajo de composición con otras características que se desarrollan en el apartado *Cómo grabar un vídeo 360°*, como la sincronía de sensores y el tratamiento del *parallax*.

3.1.2. Estándares de resolución y bitrate

La resolución de un vídeo 360° viene dictada por el reproductor en el que se realiza su visualización. Lo que sí es común en todos los reproductores es la lectura de la proyección *quirectangular* y el ratio 2:1. Algunos *players* reproducen sin problemas vídeos con proporción 16:9, aunque es posible que el reproductor cause errores como un mal cierre de esfera en la parte posterior, como se puede ver en la parte central de la Figura 5.



Figura 5. Ejemplo de mal cierre de esfera a causa de las proporciones. En este caso, el editor del vídeo trató de pasar de 16:9 a 2:1 directamente, lo que generó esa línea vertical que se observa en el centro de la imagen producido por un mal cierre de esfera.

Al igual que cuando un vídeo convencional de ratio 16:9 es reproducido en una pantalla de 4:3 sin utilizar la técnica de *letterboxing*⁷ y se pierde información en los laterales, los reproductores de vídeos 360° pueden dar errores semejante. En este caso, puede aparecer un fallo de unión en la parte posterior de la esfera donde se genera un corte seco que va de arriba a abajo en la zona de unión de la parte derecha e izquierda del equirectangular. Por tanto es recomendable la proporción 2:1 siempre que sea posible.

Las características básica de calidad de un HMD son bastante iguales que en las pantallas tradicionales. La resolución, el rango dinámico, la reproducción del color y la frecuencia de refresco suponen la calidad básica del sistema de visualización. (V.T. Peltoketo, 2016, p.7)

Los estándares de resolución de los vídeos 360° están basados en las resoluciones del vídeo tradicional⁸. Supongamos que se quiere realizar un vídeo a gran resolución, por lo que lo habitual sería pensar en un vídeo a 4K (4096x2048 píxeles). No obstante, esto no quiere decir que el usuario vaya a disfrutar de la calidad de un vídeo a 4K. Cuando un vídeo es reproducido por un *player* 360°, lo que se está haciendo es proyectar el vídeo en una esfera y poner una cámara virtual en el punto (0,0,0) de la misma. Por tanto, no se está viendo la resolución completa del vídeo. Si el campo de visión de la cámara virtual del *player* es de 120°, el usuario verá el vídeo a una resolución de 1,3K. Si se quiere disfrutar de una resolución real de 4K, entonces el vídeo debe estar a 12K (12288 x 6144 píxeles) y encontrar un *player* que sea capaz de reproducirlo (M. Kintner, 2015).

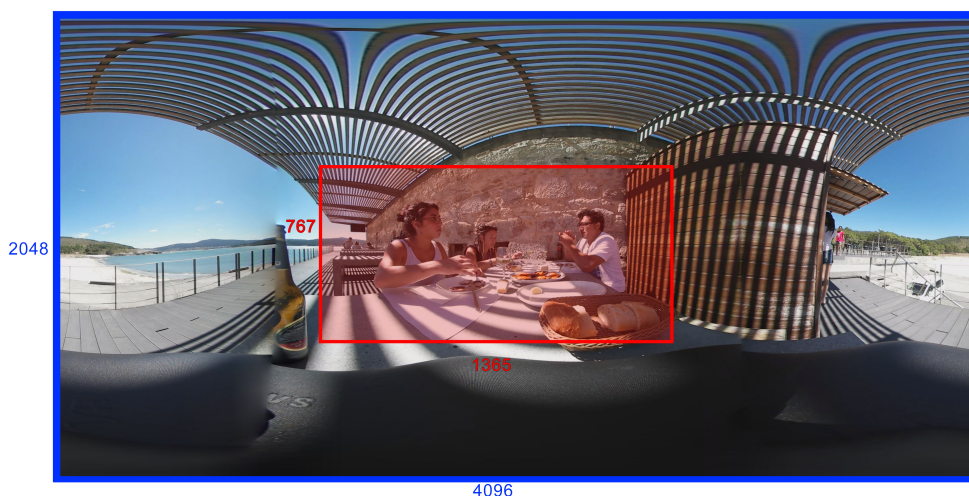


Figura 7. Resolución total frente a resolución de visión.

⁷ *Letterboxing*: técnica que consiste en la introducción bandas negras arriba y abajo de un vídeo con proporción 16:9 para que al reproducirlo en una pantalla con proporción 4:3, la imagen no se vea recortada ni deformada.

⁸ En el Anexo II. Figuras se encuentra una tabla con los estándares de resolución de los vídeos 360° [Figura 6]

Los vídeos 360° requieren una tasa de bits mayor que los vídeos convencionales debido a la gran resolución necesaria para soportar todo el campo de visión (M. Budagaci et al., 2015). Para poder comprender bien cómo la resolución y la tasa de bits afecta a un vídeo 360° en comparación con un vídeo convencional hay que tener en cuenta, sobre todo, la cantidad de información que se está procesando en cada uno de los vídeos. Un vídeo convencional reproduce una parte del entorno que rodea la cámara, sin embargo, un vídeo 360° recoge toda la información que la rodea. Por tanto, se está reproduciendo más información con una resolución semejante, por lo que para conseguir una menor pérdida de definición en la imagen, es aconsejable subir la tasa de bits tanto como el reproductor lo permita (M. Kintner, 2015).

No obstante, no todos los *players* soportan ni una resolución ni un bitrate alto. Por ejemplo, el *player* de Samsung MilkVR recomienda vídeos monoscópicos con una resolución mínima de 3840x1920 (2:1) y recomienda una tasa de *bitrate* mínima de 40Mbit/s, con un máximo de 60Mbit/s (Samsungvr, 2017). MilkVR permite unas altas especificaciones porque los móviles de Samsung que trabajan con él están pensados para poder disfrutar de vídeos 360°. No obstante, si se utiliza un *smarthphone* que no fue diseñado para la reproducción de vídeo 360°, la resolución máxima es de 1920x960 con una tasa de bits variable entre 6Mbit/s y 15Mbit/s.

Todos estos estándares cambian cuando se trata de un vídeo estereoscópico. El proceso de grabación y edición de un vídeo 360° estereoscópico se diferencia notablemente del monoscópico, pero para poder comprenderlo se necesitaría un estudio propio.

3.2. Calidad de la experiencia de un vídeo 360°

La aspiración de la realidad virtual es presentar al espectador un entorno lo más real posible para conseguir una buena experiencia inmersiva, dándole la sensación de estar en el lugar de la acción. El uso de los HMD es fundamental para conseguir este objetivo, pero que el vídeo tenga una buena calidad es vital para engañar al cerebro y hacer creer al espectador que se encuentra en el entorno representado. Cuando se habla de *calidad del vídeo 360°*, no se refiere únicamente al resultado final compuesto por la resolución, color, tratamiento de sonido y *bitrate*.

La calidad de la experiencia vendrá determinada por el nivel de inmersión y el sentimiento de presencia que proporcionan el hardware y el contenido (M. Mazuryk y M. Gervautz, 1996). V. T. Peltoketo⁹ en *Presence capture cameras* (2016) califica el

⁹ Ingeniero desarrollador de la cámara OZO (Nokia).

stitching como el elemento crítico para definir la calidad de un vídeo 360°. El *stitching* es la unión entre la grabación de múltiples cámaras para conseguir un único vídeo. Este elemento ya estaba presente hace más de cien años cuando se empezaron a hacer imágenes panorámicas (E. Anderson, 2016).

Cuando el usuario se pone el HMD, se está engañando al cerebro y éste cree que lo que ve es real. No obstante, un ente extraño en el entorno sacará al espectador de la experiencia. Por tanto, si el *stitching* se percibe, se pierde la sensación de inmersión. El objetivo principal será unir las imágenes para conseguir un único vídeo que dé sensación de continuidad visual, es decir, que no se perciba el cosido entre las imágenes. Éste será el eje vertebrador de todo el trabajo de grabación y postproducción del vídeo 360°.

Para conseguirlo, lo ideal sería grabar un vídeo con una cámara, una sola lente y un único sensor que permitiera un campo de visión de 360°x180°. Por desgracia, aún no existe tecnología que lo permita, por lo que hay que ingeniárselas para conseguir un buen *stitching*. Esto dependerá fundamentalmente del *rig* escogido para grabar el vídeo.

3.3. Grabar un vídeo 360° monoscópico

Grabar un vídeo 360° es más complejo de lo que a cualquiera le gustaría, y es que al grabar todo el entorno y utilizar varias cámaras a la vez es fácil que ocurra alguna desgracia. Por ello, la elección del equipo técnico será fundamental para conseguir salir del rodaje con éxito.

3.3.1. Planificación del *rig*

Un *rig* es el conjunto y disposición de las cámaras que se utiliza en el rodaje, aunque también puede ser una única cámara 360° con varias lentes y sensores. Ambos se pueden describir mediante tres parámetros: el radio del *rig*, el *field of view* de las cámaras y el número de cámaras (R. Anderson, 2016). Estos serán los parámetros principales a la hora de escoger un *rig*. No obstante, existen otras tres variables que serán esenciales en la elección: el *framerate*, la resolución de los sensores y el posible tratamiento del color que permita la cámara.



Figura 8. Rig F360 Explorer con 6 cámaras GoPro Hero3+.

No existe un *rig* ni una cámara 360° mejor que otra, sino que depende de las circunstancias del rodaje (más información en el apartado *Análisis para realizar la*

selección del equipo técnico adecuado). El entorno, las distancias, el plano estático o en movimiento, las condiciones lumínicas hacen que para cada ocasión se requieran unas cámaras u otras. El soporte que vaya a reproducir el vídeo también condiciona la elección de la cámara, ya que, aunque un *rig* nos pueda asegurar una mayor resolución que otro, si el *player* escogido no es capaz de reproducir el vídeo, habrá que comprimir el vídeo y se perderá la calidad de origen.

3.3.1.1. Radio del *rig*, campo de visión de las cámaras y número de cámaras

En el mercado se pueden encontrar cámaras y *rigs* con diferentes números de lentes desde 2 hasta más de 16, y cada una de ellas brinda unas comodidades que no dan las otras.

El método utilizado para realizar *stitching* se basa en coser los diferentes vídeos utilizando las áreas comunes entre ellos, por lo que es elemental que exista superposición entre las imágenes de cada cámara. (K. Korrel, 2016, p.3).

La lógica dice que si sólo se utilizan dos lentes, entonces sólo habrá un *stitching*, por tanto cabría esperar que el resultado fuera mejor, pues no habría tantos puntos críticos en los que poder fijarse. Sin embargo, no siempre aparece reflejado de esa manera en el vídeo final.

Si se escoge una cámara de dos lentes, se necesita un campo de visión alto para poder llegar a completar la esfera. Por ejemplo, si escogemos grabar con Kodak SP360 4K¹⁰ cada una de las lentes tiene un campo de visión de 235°. Sumando los grados de cada una hay 470° de información, lo que supone 110 grados de información repetida.

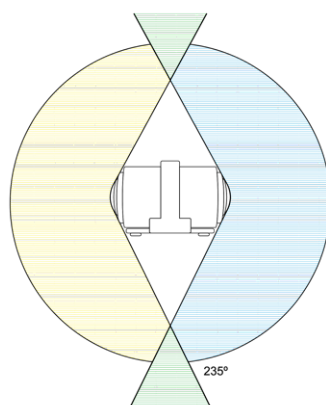


Figura 9. Rig de Kodak SP360 4K. Las áreas amarillas y azules corresponden al campo de visión de cada cámara. Las áreas verdes son la unión de ambas áreas, es decir, las zonas coincidentes que servirán para unir los vídeos y crear la esfera.

¹⁰ Especificaciones de la cámara Kodak SP360 4K.

https://kodakpixpro.com/docs/specsheets/actioncam/sp360_4k/sp360_4k-specs-web.pdf

No obstante, estos grados son aproximados ya que existe una distancia entre las cámaras que supondrá una disminución de la superposición entre imágenes. Por lo general, el beneficio de utilizar un número de cámaras bajo es que la distancia entre ellas es la mínima posible. Por eso todas las cámaras 360° de dos lentes son muy pequeñas: Samsung Gear 360, Kodak SP360 4K, Ricoh Theta S, Nikon KeyMission 4K. Cuanto menor sea la distancia entre las cámaras, menor será el *parallax* entre las imágenes.

3.3.1.1.a. El *parallax*

El *parallax* es la diferencia de alineación de un objeto al mirarlo desde dos puntos de vista distintos. Se puede comprobar la desviación angular mirando un objeto con el ojo izquierdo y con el derecho por separado.

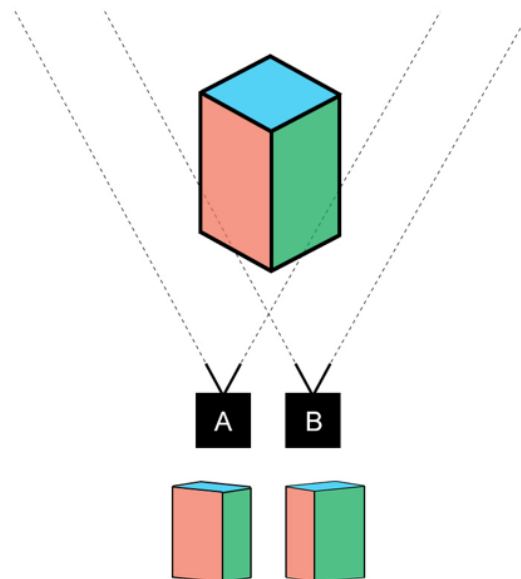


Figura 10. Esquema explicación de *parallax*.

La Figura 10 muestra dos cámaras que graban el mismo objeto desde puntos de vista diferentes. La diferencia en la geometría del objeto grabado por cada cámara es el *parallax*. A la hora de crear un vídeo 360°, las imágenes grabadas por las diferentes cámaras se estiran y contraen con el fin de corregir el *parallax* y hacer coincidentes la geometría de los vídeos. No obstante, no es una corrección fácil de realizar.

El *parallax* en un vídeo 360° crea fallos de *stitching* por los puntos de vista diferentes de cada cámara. Para conseguir un *stitching* perfecto, las cámaras deberían estar en el mismo sitio, lo que es complejo de conseguir (ver apartado *Centro de perspectiva. Punto de no-parallax*), por lo que el *parallax* se convertirá en el peor enemigo a la hora de grabar un vídeo 360°.

Jeremy Camps¹¹ (2016) explica que es imposible eliminar el *parallax* por completo, pero que existen estrategias para hacerlo menos notable:

- La distancia entre lentes debe ser la mínima posible.
- Cuanto más cercanos estén los objetos de la unión de las cámaras, mayor será el *parallax* y más notable será el *stitching*. Es aconsejable no poner objetos cerca de la cámara en las zonas de unión.
- Cuanto menor sea el número de lentes, menor será el *rig* y se reducirá el número de *stitchings*.

R. Anderson¹² (2016) puntualiza que reduciendo el número de cámaras se puede hacer un *rig* más pequeño, pero se incrementa la distancia del punto más cercano para realizar el *stitching* aumentando el tamaño de los puntos ciegos.

Al existir cierta distancia entre las cámaras, se generan puntos ciegos en la esfera correspondientes a los grados que quedan sin grabar. En la Figura 11 se puede ver un esquema de los puntos ciegos del *rig* de Kodak SP360 4K, donde se generan áreas que no son captados por los sensores, por tanto es aconsejable que ningún objeto ocupe lugar en estas zonas. En la Figura 12, el esquema corresponde a la cámara de Google Jump. Al existir más de una cámara en cualquier *rig*, siempre se crearán puntos ciegos.

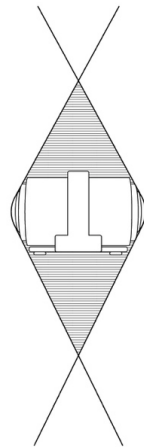


Figura 11. Rig de Kodak SP360 4K. Las áreas grises corresponden a los puntos ciegos de la grabación.

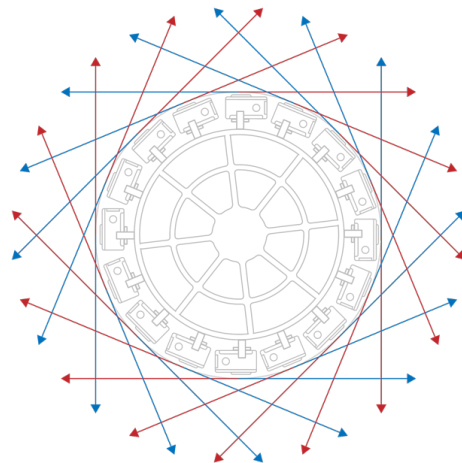


Figura 12. Esquema del campode visión de las cámaras de Jump y sus casi inexistentes puntos ciegos en el eje horizontal.

En el caso de la cámara de Google Jump, al tratarse de cámaras con lentes en toda la periferia y una distancia pequeña entre ellas, parece que los puntos ciegos se

¹¹ Jefe de soporte de la cámara Orah.

¹² Ingeniero perteneciente al equipo de creación de la cámara Jump.

transforman en algo anecdótico y que se puede hacer el *stitching* sin muchas restricciones -aunque a la hora de la postproducción se demuestra que es menos baladí de lo que podría parecer-.

No obstante, se pueden encontrar cámaras en el mercado con puntos ciegos de gran tamaño, como es el caso de la OZO¹³. Que no se malinterprete, OZO es una gran cámara para grabar vídeo 360°: permite la estereoscopía, dispone de software propio con análisis de geometría, graba audio omnidireccional, dispone de sincronía de sensores (*genlock*)¹⁴, da una salida de vídeo a 4k y 30 frames. Sin embargo, como toda cámara 360° tiene sus pros y sus contras. En el caso de la OZO ocurre que no dispone de lentes en la partes posterior. Sin embargo, sus 8 lentes de 195° generan una esfera completa, aunque en la parte posterior existe un punto ciego de entre 2 y 3 metros¹⁵.

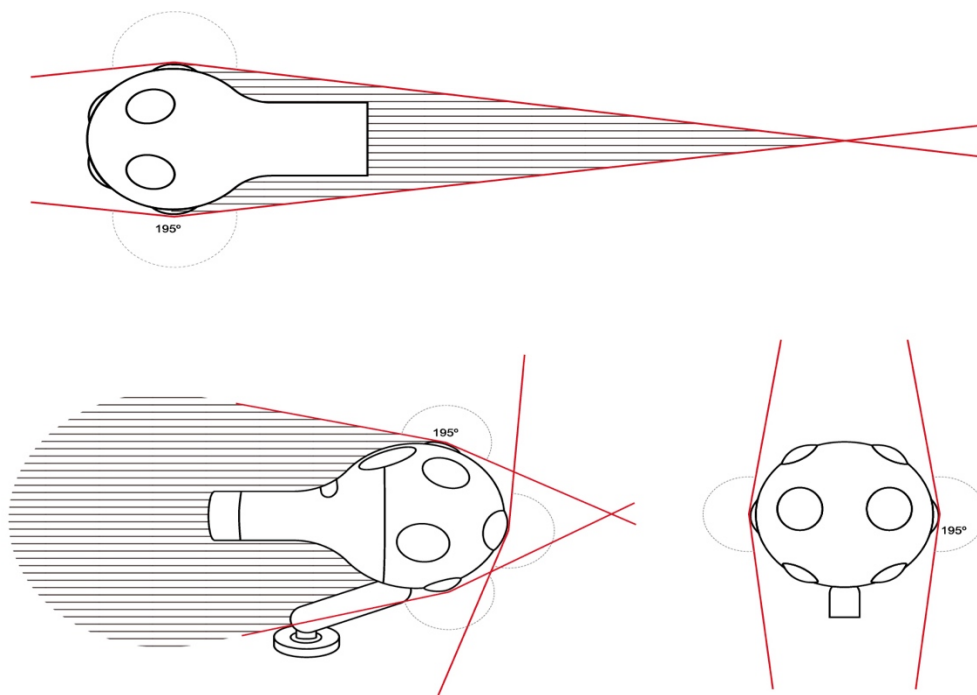


Figura 13. Análisis de campos de visión y puntos ciegos de la OZO. La imagen superior muestra el campo de visión de las cámaras traseras laterales, que son las que permiten el cierre de esfera en la parte posterior creando un gran punto ciego. Las demás lentes graban sin problema el resto de la esfera creando puntos ciegos anecdóticos.

Por tanto, siempre que se grabe un vídeo 360° hay que tener en cuenta las áreas de acción. A pesar de que se pueda hacer una aproximación de estas *áreas prohibidas* mediante el cálculo de grados por lente, no serán las únicas restricciones para definir las.

¹³ Cámara 360 de Nokia. <https://ozo.nokia.com/eu/nokia-ozo-specs/>

¹⁴ *Genlock*: característica técnica de los sensores de una cámara 360° y algunos *rigs* que permite la sincronización perfecta del material grabado en cada uno de los sensores.

¹⁵ Datos conseguido en una masterclass de operador de cámara OZO.

3.3.1.1.b. Distorsión y homografía

El *parallax* y la distorsión de lente creadas por un amplio campo de visión¹⁶ generan desigualdades en la geometría de los objetos que se encuentran en los bordes de la grabación. Por tanto, aunque se esté grabando la misma información por ambas cámaras y existan puntos coincidentes (G. Xu, 2015), la geometría se ve tan distorsionada que conseguir un buen *stitching* es una tarea imposible.



Figura 14. Frame grabado con Samsung Gear 360.

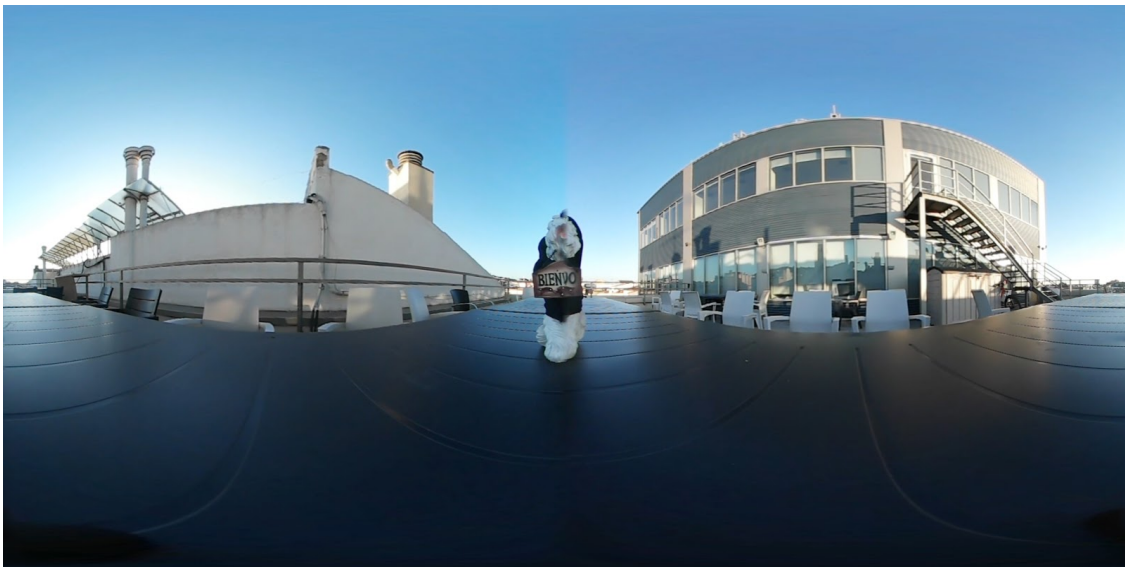


Figura 15. *Stitching* automático con Gear 360 Action Director.

Las figuras 14 y 15 representan el mismo frame. El *stitching* está generado por el propio software de Samsung (Gear 360 Action Director), que realiza un análisis

¹⁶ En *A new algorithm to correct fish-eye- and strong wide-angle-lens-distortion from single images* (C. Brauer-Burchardt y K. Voss, 2001) se explica la deformación de lente al hacer uso de amplios campos de visión y el algoritmo para corregirlo.

geométrico y une las cámaras en función de la distancia guiándose por los objetos más complejos y cercanos. No obstante, el *stitching* no es perfecto. Esto ocurre por dos motivos:



Figura 16. Zoom al *stitching* de las dos imágenes.

i. La figura se encuentra a 25cm de la cámara. A causa del recorte del sensor¹⁷, el punto ciego aumenta en las partes centrales de unión, perdiendo información del cartel que lleva el perro colgando. Nunca se podrá unir con éxito estas imágenes pues no se dispone de toda la información necesaria en esa zona.

ii. A medida que se aleja del recorte del sensor, se acerca al borde de la lente proporcionando más información de la escena. En este caso, existen puntos coincidentes de la cara del perro en las dos imágenes. No obstante, esta información se encuentra distorsionada hacia lados opuestos en ambas lentes. El software no une perfectamente estas imágenes por dos motivos:

- La deformación es tan grande que sería imposible cuadrar los puntos, pues habría que forzar demasiado las imágenes.
- La gran compresión de la información en el borde de la lente genera una pérdida de calidad a la hora de estirla en la proyección *equirectangular*, por tanto el software la utiliza para realizar el análisis y crear la unión, pero la desprecia a la hora de generar la imagen final.

Cuando se transforman diferentes imágenes para conseguir una correspondencia entre los puntos coincidentes de tal forma que los puntos y las rectas corresponden la una con la otra, se habla de homografía. Por desgracia, la homografía no entiende de *parallax* y cuando se juntan ambos conceptos lo habitual es que se creen artefactos como imágenes fantasmas o estructuras rotas (F. Zhang y F. Liu, 2014). Por tanto, a la hora de decidir la línea del *stitching* habrá que tener picardía y redireccionarla por donde menos se note.

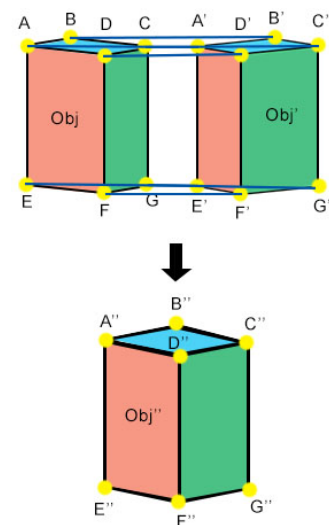


Figura 17. Homografía

¹⁷ En el caso de la Samsung Gear, el sensor no es suficientemente grande para el campo de visión de la lente, por lo que la imagen queda recortada por los bordes perdiendo información.

El efecto que genera el *parallax* a la hora de unir dos imágenes se puede comparar con el enfoque de una cámara. El enfoque se realiza a una distancia específica y, a partir de ella, el entorno se ve más o menos borroso en función de la distancia al punto enfocado. A mayor distancia, más borroso se verá. Cuando se realiza un *stitching* ocurre lo mismo.

Intentar realizar una homografía perfecta con un *rig* 360° es imposible por dos motivos: el *parallax* entre imágenes y la deformación de los puntos de unión coincidentes entre las imágenes. Cuando con un software de edición se generan puntos de control para unir imágenes, se le está obligando a deformarlas para que encajen. Esta deformación se basa en el algoritmo del programa, que hará un análisis geométrico y estirará o comprimirá la imagen en función de los puntos seleccionados. Sin embargo, la relación de deformación entre puntos que se encuentran a diferentes distancias del *rig* no es la misma. Como ocurre con el enfoque de una cámara, sólo se puede enfocar a una distancia, por tanto, sólo se podrá unir perfectamente las imágenes a una distancia concreta. Cuanto más lejos estén los objetos de la distancia de *stitching* escogida, peor será la unión creando artefactos como duplicados o imágenes fantasmas.

La Figura 18 muestra la imagen recogida por tres de las 6 lentes de la cámara Omni. Hay cierta información que se repite en las tres imágenes, por lo que la unión parece posible. En este caso, para unir las cámaras se decidió utilizar la opción de *stitching* automático que brinda Autopano Video¹⁸. Sin embargo, el resultado no es convincente. El software ha decidido coger como puntos de unión las piedras que se encuentran en la parte izquierda de 18(d) y apenas encontró puntos coincidente entre 18(b) y 18(c). Se genera una imagen doble por:

- El software no fue capaz de forzar a 18(b) y 18(c) a unirse, por lo que tienen pocos puntos contra los que luchar con respecto a la fuerza que hace 18(a) en ellos.
- Las casas del fondo se encuentran lejos de los puntos escogidos para realizar el *stitching*, por lo que no se puede conseguir una buena homografía.

Lo ideal es tener el control suficiente en la escena para poder unir todas las imágenes del *rig* a una distancia semejante. Si se quiere unir X con Y a una distancia de 10

¹⁸ Software de Kolor (comprado por GoPro) que permite realizar automática y manualmente un *stitching*.

metros y, a la vez, Y con Z a una distancia de 100 metros, Y tiene que forzar tanto por ambos lados que finalmente no puede unir bien con ninguna de las imágenes.

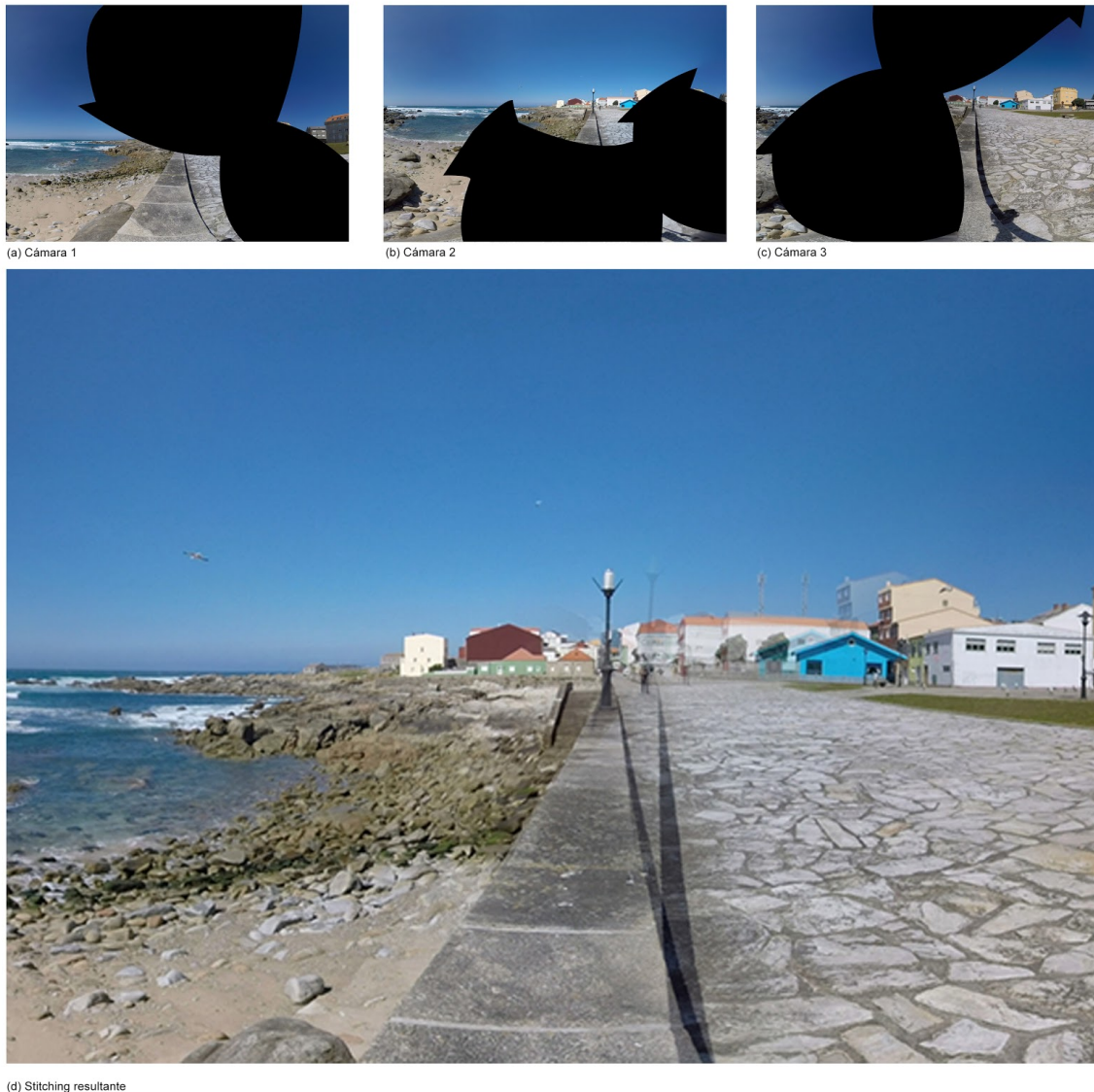


Figura 18. En la parte superior (a, b, c), la imagen grabada por las cámaras por separado. Abajo, el stitching resultante (d) generando una imagen suplicada.

El principal culpable de no conseguir una homografía perfecta es el *parallax* que, según Camps (2016) es imposible de eliminar. Es cierto que existen maneras de burlarlo, pero ¿cuál es la mejor solución?

3.3.1.2. Centro de perspectiva. Punto de *no-parallax*

La mejor manera de solucionar los problemas que conlleva el *parallax* es encontrar el centro de perspectiva, también conocido como punto de *no-parallax* o *entrance pupil*. Esta técnica consiste en utilizar una única cámara que pivota sobre su punto de *no-parallax*. Al girarla, los objetos que están en primer plano y en el fondo se encuentran perfectamente alineados (R. Littlefield, 2006). Este procedimiento no sólo asegura un mejor resultado en el *stitching* gracias a que elude el *parallax*, sino que permite utilizar

cámaras de mejor calidad que las propias cámaras de 360°, que suelen ser pequeñas cámaras de acción con *settings* limitados y sensores muy pequeños. A pesar de ello, existe un retahíla de limitaciones a la hora de grabar con esta técnica. Para poder comprender las limitaciones, primero hay que tener claro cómo se debe grabar.

3.3.1.2.a. Grabar pivotando sobre el punto de *no-parallax*

Lo primero es decidir qué cámara y con qué objetivo se va a utilizar. En este caso es fundamental escoger correctamente el campo de visión de la cámara. Como ya se dijo anteriormente, a mayor campo de visión, menor número de cámaras será requerido o, en este caso, menor número de giros y por tanto, menor número de zonas críticas de unión. No obstante, un mayor campo de visión supone una mayor deformación en el borde de la imagen, por lo que sigue habiendo *parallax* y la homografía sigue siendo imposible.

Una vez escogido el equipo fotográfico, hay que utilizar preferiblemente una rótula panorámica para poder conseguir el punto sobre el que pivotar. En este momento hay que tener claro cuántos grados se debe girar para conseguir áreas de unión. Ya se dispone de tablas con datos calculados en función del campo de visión y el sensor [Ver *Figura 19. Anexo II*]. No obstante, es posible que se escoja una combinación que aún esté sin calcular. En este caso, habrá que tener en cuenta que es necesario disponer de áreas solapadas entre las diferentes imágenes. “Para realizar una buena panorámica, la superposición debe ser del 20% de imagen” (J. J. González, 2012).

Hay que tener presente que es posible que sólo se consigan 360° horizontales, quedando sin grabar la zona superior e inferior, sobre todo si se escoge un campo de visión pequeño. Por tanto, una vez hecho el horizonte habrá que grabar la zona superior e inferior pivotando sobre el punto de *no-parallax*. Aunque el campo de visión escogido abarque los 180° verticales, es aconsejable grabar también esas zonas, pues si no se unirán en función de varias cámaras que recogen los mismos puntos y el *stitching* será complicado. Este método conlleva nuevas limitaciones a la hora de grabar.

3.3.1.2.b. Limitaciones del uso del punto de *no-parallax*

Una de las principales limitaciones que supone esta técnica son las áreas de grabación. Al grabar con una única cámara, hay que acotar la acción dentro de este espacio. Estas áreas serán más pequeñas en función del *field of view* escogido. Como se dijo anteriormente, lo mejor es que sea lo menor posible para tener menos deformación en el borde la imagen. Sin embargo, a menor campo de visión, menor es

el campo de acción. Por tanto, hay que planear con antelación las áreas de acción y la dirección de los actores.

Al grabar cada dirección en un momento diferente, resulta fácil encontrar problemas con la luz y las sombras, sobre todo si se graba en exteriores. Por un lado, existe la posibilidad de cambios de luz en el entorno, acentuando la unión de las imágenes a causa del cambio de color. Por otro lado, el cambio de las sombras o la posible prolongación de éstas fuera de las áreas de acción crean *stitchings* indeseados e imposibles de corregir. Lo mismo ocurre con la cámara superior si está grabando, por ejemplo, un cielo con nubes que se mueven. A mayores, si hay personajes situados en diferentes áreas que interactúan entre ellos, hay que tener presente la sincronía entre sus diálogos y acciones. Esta técnica supone mucho tiempo de preparación, por lo que no es aconsejable utilizar esta técnica si se dispone de poco tiempo para el rodaje.

La última gran desventaja de esta técnica es que prácticamente obliga a grabar plano fijo. Es cierto que se puede grabar en movimiento utilizando un *travelling* o una grúa motorizada, pero esto eleva el coste y no asegura una buena sincronización entre las imágenes: es fácil que haya vibraciones que varían en cada pasada y que aparezcan problemas de *rolling shutter*¹⁹.

Hasta aquí quedan resumidos los aspectos físicos básicos que interfieren a la hora de conseguir un buen *stitching*. Lo primero será decidir el equipo de grabación óptimo para la escena que se quiere grabar. Esta decisión supondrá escoger el número de cámaras para poder completar la esfera. Al no grabar con una única lente y sensor, se generan puntos ciegos de los que no se tendrá información y áreas demasiado deformadas como para conseguir una buena homografía, por lo que es aconsejable no acercarse a estas zonas críticas. Aparece el gran enemigo *parallax*, el cual se puede burlar en cierta medida utilizando una única cámara y pivotando sobre su centro de perspectiva. No obstante, esta técnica conlleva a cierta limitación en cuanto a áreas, sincronización y movimientos.

Controlados estos aspectos físicos básicos, ¿qué *settings* de la cámara se deben tener en cuenta?

¹⁹ El *rolling shutter* supone un gran problema a la hora de hacer el *stitching*. Cuando un objeto (o la propia cámara) se mueve muy rápido, la imagen se registra verticalmente de tal manera que la última fila de píxeles no se graba en el mismo momento en el que es registrada la primera. Esto produce distorsión en la imagen. (cont. Anexo I. Glosario)

3.3.2. *Framerate*, resolución de los sensores, color

Elaborar un vídeo 360° se base en la unión de diferentes imágenes. Esta unión nunca será satisfactoria si el *parallax* entre las imágenes es amplio. Aunque no exista *parallax* y la homografía geoméricamente hablando sea perfecta, existen más factores que pueden acentuar el *stitching*. Se pueden diferenciar dos maneras de trabajar en función de la cámara o *rig* seleccionado.

Existen cámaras 360° que disponen de un software interno que conecta todos los sensores, facilitando el uso de los *settings*. Por ejemplo, modificar el *shutter*, el *framerate*, la resolución, la relación de aspecto y el tratamiento de color de todas las cámaras a la vez. Esto facilita el trabajo y disminuye el tiempo de preparación del set de rodaje. Sin embargo, es habitual que sea preferible, por características del entorno o la acción, grabar con un *rig* que no disponga de software interno, por tanto, habrá que modificar los *settings* de las cámaras una a una. Por ejemplo, si se quiere grabar un vídeo 360° donde la cámara corra riesgo de caerse al agua. En este caso, actualmente no existe una cámara 360° ni *rig* que disponga de software interno y que proteja a la cámara en caso de ser sumergida. Por tanto, hay que escoger un *rig* que disponga de protección para el agua (F360 Explorer, por ejemplo), pero cuyo uso de cámaras es individual. Lo ideal es que todas las cámaras tengan los mismos *settings* para conseguir uniformidad en la imagen final.

Observando el flujo de trabajo de un programa para realizar *stitching*, se observa que el primer paso es sincronizar las diferentes imágenes que componen la esfera.

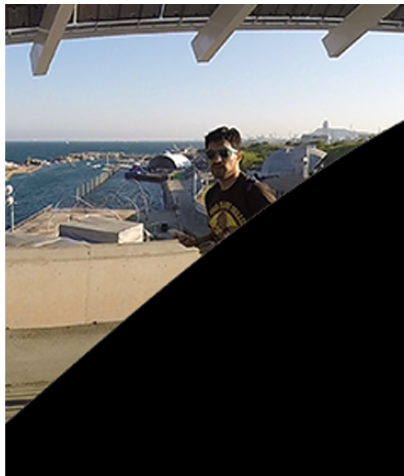


Figura 20. *Flujo de trabajo para realizar un stitching* (K. Korrel, 2016)

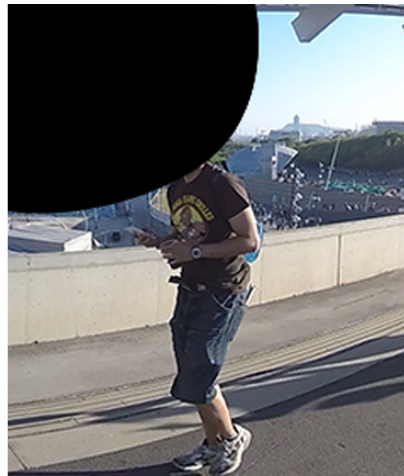
Una buena sincronización será imposible si el *framerate* varía entre las cámaras utilizadas. En el caso de la realidad virtual, lo ideal es tener el mayor número de *frames* por segundo posible por dos motivos:

- Si se disfruta del contenido con un HMD, a mayor número de frames por segundo mayor es la sensación de fluidez de la imagen y, por tanto, mayor sensación de realidad proporciona. Por otro lado, cuanto mayor sea el *framerate*, más aproximada será la sincronía entre las diferentes imágenes (K. Korrel, 2016).
- No conseguir una buena sincronización entre las cámaras supone estropear el *stitching* por completo, aunque el *parallax* sea nulo. Los fallos de *stitching*

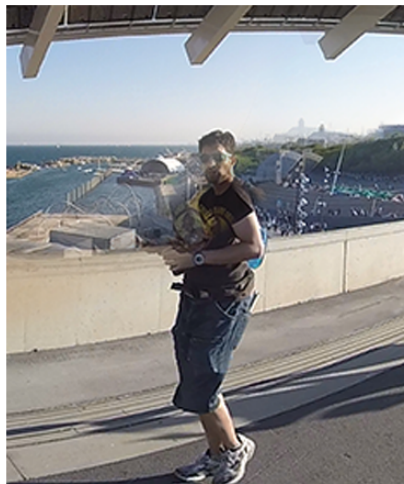
producidos por una mala sincronización se detectan en los planos en movimiento o en el paso de las acciones de una cámara a otra.



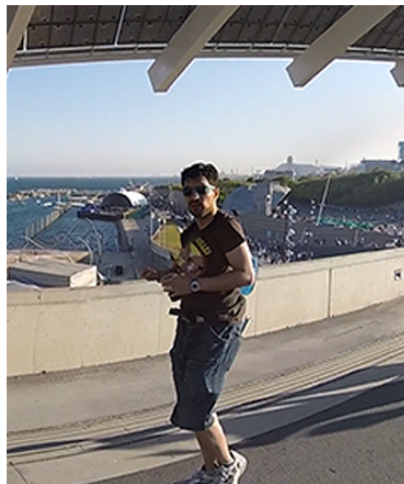
(a) Cámara izquierda



(b) Cámara derecha



(c) *Stitching* con ambas cámaras a la vez sin sincronía



(d) *Stitching* corte seco con sincronía



(e) *Stitching* con corte seco sin sincronía



(f) *Stitching* señalado en rojo sin sincronía

Figura 21. Cómo afecta el fallo de sincronía al *stitching*.

La Figura 21 presenta un problema en la sincronización de dos cámaras. El chico de gafas atraviesa el *stitching* caminando. La imagen 21(a) muestra lo grabado por la cámara de la izquierda, y la 21(b) lo grabado por la cámara de la derecha. Sin

embargo, existe un fallo de sincronización, la cámara de la izquierda está 4 frames adelantada a la de la izquierda (los brutos están a 50fps).

Una vez localizados los puntos de control, se genera la imagen 21(c), es decir, el *stitching* que nos permite ver las dos imágenes a la vez, superpuestas. En este caso hay un desdoble claramente visible en la cabeza del chico²⁰, pues se están uniendo dos imágenes que fueron registradas en momentos diferentes.

La imagen 21(e) muestra la unión de las cámaras con un corte seco, generando una línea de *stitching* claramente visible, señalado en color rojo en la imagen 21(f). Parece que la cabeza se encuentra desplazada hacia delante con respecto a su tronco. Sin embargo, en la imagen 21(d) la cabeza encaja sin problema con el resto del cuerpo. En este caso sólo se hizo corrección de sincronización, consiguiendo un buen stitching sin tener que modificar los puntos de unión.

Hay que tener presente que los fallos de sincronización son más visibles en un vídeo que en una fotografía, pues es un problema dinámico. De hecho, un fallo de sincronía puede ser fácilmente solucionable cuando se trata de una fotografía, como en el caso de la Figura 21. Sin embargo, corregir un problema de sincronización en un vídeo en la fase de posproducción se convierte en toda una odisea.

Si el *rig* escogido no dispone de un sistema de *genlock*, hay que sincronizar las diferentes imágenes manualmente. Ya sea utilizando un programa de *stitching*, de sincronía o de modo manual. La sincronización se puede realizar de dos maneras diferentes: usando audio, como palmadas o una claqueta; o con cambios en la imagen, ya sean de movimiento o de luz.

Aunque se consiga una sincronización perfecta al *frame*, es posible que se encuentren fallos de sincronía. Esto se debe a que se ha producido un cambio en la imagen que es más rápido que la velocidad del obturador, por tanto, aunque tenga una buena sincronía de *frame*, se podrá encontrar fallos de *subframe*. En este caso, tanto el *frame* como el *subframe* se considera una unidad temporal y no física. Es decir, si se graba un vídeo a 25 fps, cada *frame* dura 4 milisegundos. Si en el entorno ocurre una acción que dura menos de 4 milisegundos, entonces hay un fallo de sincronía al

²⁰ También existe una duplicidad en los edificios del fondo. No obstante, este fallo de *stitching* no por culpa de la sincronización, ya que las cámaras no se movían. Este desdoble se produce porque los puntos de unión escogidos formaban parte del hombre, que se encuentra muy lejos de los edificios. Por tanto, la homografía a la distancia de los edificios no es posible.

subframe, pues la acción no puede ser grabada a la vez por los diferentes sensores. Esto es habitual que ocurra cuando la cámara se mueve con rapidez o cuando hay luces estroboscópicas y flases en la escena. Por tanto, Es preferible utilizar una cámara que disponga de *genlock* y evitar así fallos de *subframe*.

La Figura 22 es un ejemplo de fallo de sincronía la *subframe*. En este caso, la cámara estaba fijada a la cabeza de un conductor de moto acuática. Los movimientos de la moto resultaron ser tan bruscos y rápidos que, a pesar de estar grabado a 59,97 fps, fueron más rápidos que la velocidad de los obturadores, generando así fallos en la unión de las cámaras.

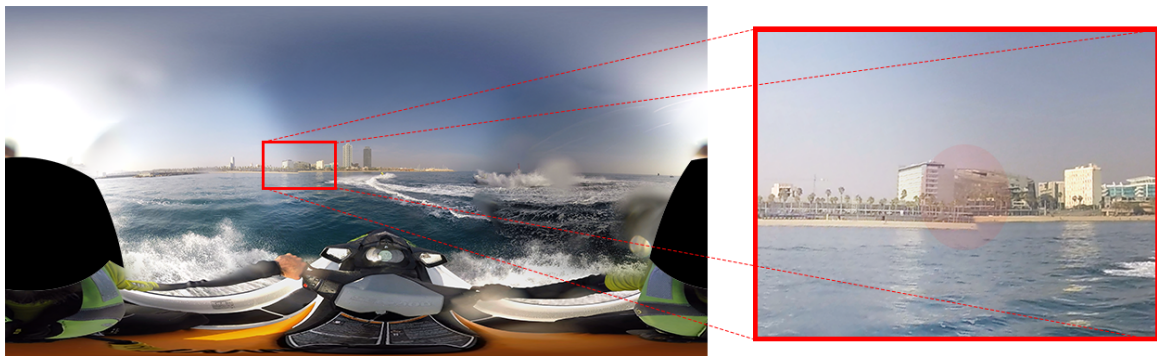


Figura 22. Ejemplo de fallo de sincronía al subframe.

A la hora de realizar el *stitching* es importante saber identificar si los fallos de unión entre cámaras ocurren por una mala elección de los puntos de control o por un fallo de sincronía. A medida que se practica, el ojo se acostumbra a identificar el origen del error, pero al principio resulta complicado reconocer el motivo del fallo.

Una vez sincronizados los vídeos y realizado el *stitching* a través de puntos de control consiguiendo uniformidad geométrica, es posible que se vean fallos de unión a causa de la diferencia de color entre las diferentes imágenes. Lo ideal es que todos los sensores y todas las lentes tengan las mismas características. De igual manera, todas los vídeos deben “tener la misma resolución, mismo ruido, iguales niveles de negro y saturación” (V.T. Peltoketo, 2016).

4. Análisis para realizar la selección del equipo técnico adecuado

Realizar una buena elección de la cámara 360° es fundamental para conseguir un buen resultado en el vídeo y que la unión entre cámaras sea lo menos perceptible posible. Un vídeo 360° no se puede grabar con una única cámara que disponga de una única lente y un sensor y genere una esfera completa de 360°x180°. Actualmente, la cámara perfecta para grabar vídeos 360° no existe, y quién sabe si en algún futuro existirá. Se pueden encontrar decenas de cámaras 360° en el mercado, y siempre es

posible crear un *rig* propio. Aún así, no se puede decir que una cámara sobresalga por encima de otra, pues la elección del equipo viene dictada por las características del rodaje.

Al igual que en el cine tradicional, se trata de elegir las herramientas adecuadas para el trabajo. [...] Hay que centrarse en el sistema de cámaras correcto, no en el perfecto o el mejor. Porque lo que es “mejor” cambiará dependiendo de su aplicación. Hay que escoger el equipo en función de las necesidades del rodaje. (J. Nicholas, 2017).

Las necesidades del rodaje se definen en función del entorno, la acción y las condiciones temporales. Ninguno de los tres conceptos predomina por encima de los otros dos, los tres son igual de importantes y debemos tenerlos constantemente en cuenta. Sin embargo, hay ocasiones en las que una de las condiciones es tan fuerte que obliga a descartar ciertas cámaras o técnicas. Por ejemplo, si el rodaje es en un sitio público donde no hay control sobre la acción, se descarta la posibilidad de realizar el rodaje utilizando el punto de *no-parallax*.

4.1. Analizar el entorno

Al grabar un vídeo 360° se filma todo el entorno que rodea a la cámara, por eso es elemental examinarlo a fondo para comprender su geometría. Para ello, se debe tener en cuenta como mínimo:

- **Elección de la posición del trípode.** Escoger bien la posición del trípode es fundamental para conseguir un buen *stitching*. Es cierto que a veces por exigencias del guión, el trípode no se sitúa en el punto que es técnicamente el más óptimo, pero en la medida de lo posible debe situarse en un lugar donde se puedan controlar las líneas de unión. Para ello, debemos fijarnos en los tres siguientes puntos.
- **Distancia de la cámara al punto más cercano y al más lejano.** Lo ideal es que la distancia entre el punto más cercano y el más lejano sea la mínima posible. Deben ser puntos que se encuentran en las líneas de unión de las cámaras. Es decir, no importa que exista un punto muy alejado de la cámara en medio una de las imágenes recogida por uno de los sensores, si ese punto no se va a utilizar para unir las cámaras. Recordemos que para conseguir una buena homografía, lo ideal es que los puntos de unión estén a la misma distancia de la cámara.
- **Condiciones lumínicas.** Al tratarse de un vídeo de 360° lo habitual es que no se pueda iluminar la escena con focos, pues se verían. Por tanto, hay que trabajar

con las condiciones lumínicas naturales del entorno. Si el espacio tiene poca luz, es preciso un juego de cámaras que capte bien los negros y con un gran rango dinámico, por lo que se descartan las cámaras de acción con sensores pequeños, pues generan mucho ruido y no captan suficiente información en ambientes oscuros. Las condiciones lumínicas también afectan a la posición de la cámara. Por ejemplo, si se va a utilizar una cámara en la que los *settings* de color no se pueden tocar manualmente, es recomendable no poner la cámara al sol si la acción principal ocurre a la sombra, pues no se podrá ver.

- **Posibilidad de camuflar la unión de las cámaras con el entorno.** Una vez comprendida la geometría del espacio, debemos fijarnos en lugares por donde se puede guiar la línea de *stitching*. Por ejemplo: columnas para *stitchings* verticales, áreas sin textura y monocromáticas, superficies horizontales... Aunque la línea de *stitching* depende del número de cámaras y la posición de la misma, es importante tener claro qué zonas de la geometría del entorno se pueden aprovechar para camuflar la unión.

4.2. Analizar la acción

En los vídeos 360° cambian la narrativa y el lenguaje audiovisual. Ahora, el fuera de campo lo escoge el usuario cuando decide mirar para un lado y no para al otro con su HMD. Aunque es cierto que se puede influir al espectador para que mire a un sitio concreto, nunca se puede asegurar al 100%. Ergo, el *stitching* debe estar pulido en su totalidad (360°x180°). Para comprender cómo la acción puede afectar al *stitching*, hay que preguntarse:

- **¿Se trata de una acción controlada?** En caso negativo se descarta automáticamente el uso del punto de *no-parallax*. En caso de filmar una acción controlada, hay que sacarle partido a ello. Ciertamente se puede grabar usando el punto de no-parallax, pero quizás por condiciones temporales es mejor filmar con un *rig* de cámaras DSLR o Freedom. Hay que aprovechar las acciones controladas para evitar que los actores pasen por zonas de *stitching*.
- **¿Cuál es la distancia mínima de la acción a la cámara?** Para evitar problemas de *stitching*, es recomendable que la acción no ocurra muy cerca de cámara. No obstante, si el guión lo exige, lo ideal es utilizar lentes que permitan un amplio campo de visión para que esa acción no se vea cortada por un *stitching*.
- **¿Se mueve la cámara?** La pesadilla de todo editor de vídeo 360° es un plano en movimiento. La postproducción de estos vídeos es bastante compleja y escojas la cámara que escojas siempre parecerá una mala opción. Sin

embargo, utilizando una cámara con *genlock* se asegura que no habrá problemas de sincronización, lo que resulta un alivio en la postproducción.

4.3. Condiciones temporales

Ya se ha analizado el espacio y la acción y resulta que la mejor opción para tener un buen *stitching* es pivotar una cámara DSLR sobre su punto de *no-parallax*. Sin embargo, no hay suficiente tiempo para realizar todos los tiros de cámara que se necesitan para completar la esfera, por tanto se rechaza esta opción. El tiempo también puede influir en la vida de la batería de la cámara. Si se tiene que grabar continuamente durante 1 hora, muchas cámaras 360° son descartadas porque su batería no dura lo suficiente y alimentarlas supone que se vería el cable en el vídeo.

Se puede encontrar la cámara OZO como una de las mejores valoradas en la industria: salida de vídeo 4K a 30 *frames*, dispone de software propio con análisis de geometría, permite la estereoscopía, dispone de sincronía de sensores, graba audio omnidireccional. Sin embargo, tiene un punto ciego de 3 metros en la parte posterior y pesa 4,2 kg y mide 264 x 170 x 238mm. Por tanto, no es la mejor opción para grabar dentro de estancias pequeñas o, por ejemplo, para realizar un *point of view*.

Supongamos que queremos grabar un *point of view* de una persona que está dando una vuelta en bicicleta. La OZO, a pesar de ser una buena cámara, quedaría descartada por su peso, medida y porque el su *software* de edición no permite modificar el horizonte, por tanto, la imagen no se podría estabilizar a no ser que se pasara por otro software una vez exportado. Otra elección podría ser la Omni, pues pesa menos, también dispone de sincronía de sensores y permite grabar a 60 *frames*. No obstante, sigue siendo muy aparatosa y sería difícil situarla en la cabeza del actor para conseguir un buen *point of view*. Por tanto, ¿qué opciones quedan?

La mejor opción sería una cámara pequeña, con el menor número de lentes posible y con sincronía de sensores. Buscando un equipo con estas características, posiblemente se acabaría escogiendo la Samsung Gear 360 o la Orbit360 4K, cámaras que a pesar de tener unas especificaciones teóricamente “peores” que la OZO o la Omni, por las condiciones del rodaje se convierten en la mejor elección.

5. Conclusiones

A causa del auge que están teniendo los vídeo 360°, surge la necesidad de compilar en este trabajo los aspectos básicos para conseguir comprender qué es y cómo grabar un vídeo 360° para conseguir el mejor resultado posible. En primera instancia, hay que ceptar que no existe el vídeo 360° perfecto, pues a día de hoy no se dispone de una cámara 360° constituida por una única lente y un sensor siendo capaz de generar una esfera completa de 360°x180°. Al haber más de una lente, en todos los vídeos habrá al menos un *stitching*, por eso el equipo técnico usado será decisivo para conseguir una vídeo 360° de calidad.

La decisión dependerá de las condiciones del rodaje. Por lo tanto, gran parte del trabajo del operador de cámara es analizar el entorno, la acción y las condiciones lumínicas para saber qué cámara será mejor en cada ocasión. Todo esto, teniendo en cuenta que el *stitching* estará siempre presente, pues en todo vídeo 360° hay *parallax* entre las diferentes imágenes, aún usando el punto de *no-parallax*, pues la homografía será imposible debido al gran campo de visión de las lentes.

Este trabajo recoge los conceptos básicos a tener en cuenta para disimular el *stitching*: la importancia de la geometría del entorno, la distancia entre las cámaras, la deformación de la imagen en función del campo de visión, la elección de los settings del equipo técnico, la repercusiones de no conseguir una buena sincronización entre los sensores y el temido *parallax*.

Tras estas páginas de análisis, sólo se puede afirmar que conseguir un resultado perfecto en un vídeo 360° resulta imposible si no se pasa por una fase de postproducción. No obstante, teniendo en cuenta los conceptos expuestos el trabajo se facilitará.

6. Bibliografía

- Anderson, R., Gallup, D., Barron, J. T., Kontkanen, J., Snavely, N., Hernández, C., ... & Seitz, S. M. (2016). Jump: virtual reality video. *ACM Transactions on Graphics (TOG)*, 35(6), 198.
- Barnes, J. (2016). Understanding Virtual Reality in Marketing: *Nature, Implications and Potential*. London: King's College London
- Budagavi, M., Furton, J., Jin, G., Saxena, A., Wilkinson, J., & Dickerson, A. (2015, September). 360 degrees video coding using region adaptive smoothing. In *Image Processing (ICIP), 2015 IEEE International Conference on* (pp. 750-754). IEEE.
- Camps, J. (2016). What is parallax? How does it impact 360 videos output? *Support Orah Video Stitch* [blog] 7 julio. Disponible en: <http://support.video-stitch.com/hc/en-us/community/posts/211167728-What-is-parallax-How-does-it-impact-360-video-output-> [consulta: 19 marzo 2017]
- Cruz-Neira, C. (1993). Virtual reality overview, overview of virtual reality systems. *Course Notes 23 of SIGGRAPH'93*.
- eMarketer (2016) Virtual Reality Is an Immersive Medium for Marketers: Marketers should start experimenting sooner rather than later. Disponible en: <http://www.emarketer.com/Article/Virtual-Reality-Immersive-Medium-Marketers/1013526> [consulta: 1 abril 2017]
- Greenbaum, P. (1992). The lawnmower man. *Film and video*, 9(3), 58-62.
- J. Kastrenakes (2016). HTC sold 15,000 Vive units in its first 10 minutes on sale. *The Verge*. Disponible en: <http://www.theverge.com/2016/3/1/11143628/htc-vive-15000-sold-in-first-10-minutes> [consulta: 3 abril 2017]
- J. Nicholas (2017). Virtual reality Pop. *The Perfect VR/360° Camera*. Disponible en: <https://virtualrealitypop.com/the-perfect-vr-360-camera-8513d25aa8c7> [consulta: 7 mayo 2017]
- J. Woolley (1993). Virtual worlds: A journey in hype and hyperreality. Oxford, Inglaterra: Blackwell
- Kintner, M. (2015). 4K VR 360 video: What is it and how can i produce it. *360 rize*. Disponible en: <https://www.360rize.com/2015/02/4k-vr-360-video-what-is-it-and-how-can-i-produce-it/> [consulta: 1 junio 2017]
- Korrel, K. (2016). Video stitching for virtual reality environments. *Science*.
- Lee, H., Tateyama, Y., & Ogi, T. (2010, October). Realistic visual environment for immersive projection display system. In *Virtual Systems and Multimedia (VSMM), 2010 16th International Conference on* (pp. 128-132). IEEE.

Mazuryk, T., & Gervautz, M. (1996). Virtual reality-history, applications, technology and future.

McLellan, H. (1996). Virtual realities. En D.H. Jonassen *Handbook of research for educational communications and technology* (pp. 457-487). Taylor y Francis.

NASA (2016) Virtual Reality: Definition and Requirements. Disponible en: <http://www.nas.nasa.gov/Software/VWT/vr.html> [consulta: 1 abril 2017]

Nokia, OZO (2016) Technical specifications. *Nokia*. Disponible en: <https://ozo.nokia.com/eu/nokia-ozo-specs/> [consulta: 19 marzo 2017]

O'Kane, S. (2015). GoPro just bought a virtual reality company. *The Verge*. Disponible en: <http://freedom360.us/gopro-bought-kolor/> [consulta: 1 abril 2017]

Peltoketo, V. T. (2016, April). Presence capture cameras-a new challenge to the image quality. In *SPIE Photonics Europe* (pp. 98960Q-98960Q). International Society for Optics and Photonics.

PTGrey (2012) *Ladybug3 12 MP FireWire 1394b Spherical Digital Video Camera, Black*. Disponible en: <https://www.ptgrey.com/ladybug3-12-mp-firewire-1394b-spherical-digital-video-camera-black> [consulta: 7 mayo 2017]

RAE (2014). Diccionario de la lengua española. Disponible en: <http://dle.rae.es/?id=buDJhh3> [consulta: 1 abril 2017]

Segaretro (2017). SegaVR. Disponible en: http://segaretro.org/Sega_VR [consulta: 3 abril 2017]

Sutherland, I. E. (1965). The ultimate display. *Multimedia: From Wagner to virtual reality*.

VRS (2016). Virtual Reality Society. Disponible en: <https://www.vrs.org.uk/> [consulta: 29 marzo 2017]

Williams, D., & Burns, P. D. (2013, January). Image Stitching: Exploring Practices, Software, and Performance. In *Archiving Conference* (Vol. 2013, No. 1, pp. 126-131). Society for Imaging Science and Technology.

Woolley, B. (1993). *Virtual worlds: A journey in hype and hyperreality*. Benjamin Woolley.

Lin, M., Xu, G., Ren, X., & Xu, K. (2015, June). Cylindrical panoramic image stitching method based on multi-cameras. In *Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), 2015 IEEE International Conference on* (pp. 1091-1096). IEEE.

Zhang, F., & Liu, F. (2014). Parallax-tolerant image stitching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3262-3269).

7. Anexo I. Glosario

Genlock: característica técnica de los sensores de una cámara 360° y algunos *rigs* que permite la sincronización perfecta del material grabado en cada uno de los sensores.

HMD (head-mounted display): gafas de realidad virtual que habitúan a estar conectadas a unos cascos y otros periféricos que permiten al usuario navegar a través del mundo virtual. Los HMD utilizan la técnica del *tracking* de movimiento para permitir al usuario moverse dentro del entorno que se le presenta.

Homografía: correspondencia geométrica entre los puntos coincidentes de diferentes imágenes de tal forma que los puntos y las rectas corresponden la una con la otra.

Letterboxing: técnica que consiste en la introducción bandas negras arriba y abajo de un vídeo con proporción 16:9 para que al reproducirlo en una pantalla con proporción 4:3, la imagen no se vea recortada ni deformada.

Parallax: es la diferencia de alineación de un objeto al mirarlo desde dos puntos de vista distintos. Se puede comprobar la desviación angular mirando un objeto con el ojo izquierdo y con el derecho por separado.

Proyección *equirectangular*: también conocida como proyección cilíndrica equidistante, es proyección plana de una esfera

Rig: conjunto y disposición de las cámaras que se utiliza en el rodaje, aunque también puede ser una única cámara VR con varias lentes y sensores. Ambos se pueden describir mediante tres parámetros: el radio del *rig*, el *field of view* de las cámaras y el número de cámaras

Sonido binaural: audio grabado mediante el uso de dos micrófonos situados entre sí a una distancia semejante a la de una cabeza de persona adulta con el fin de crear para el oyente una sensación de sonido 3D como si estuviera en la estancia que es presentado. A diferencia del estéreo, gracias a la distancia entre los micrófonos, existe un retardo del surround dando un sonido final más cercano a la que una persona podría escuchar en esa situación.

Stitching: la unión entre la grabación de múltiples cámaras para conseguir un único video.

Tracking de movimiento: característica técnica que debe tener un casco de realidad virtual para poder rastrear el movimiento de la cabeza del usuario y moverlo a través del espacio virtual que se le presenta.

Rolling shutter: efecto creado en una imagen Cuando un objeto (o la propia cámara) se mueve muy rápido, la imagen se registra verticalmente de tal manera que la última fila de píxeles no se graba en el mismo momento en el que es registrada la primera. Esto produce distorsión en la imagen. Supone un gran problema a la hora de realizar el stitching porque a cada cámara le afectará el *rolling shutter* de una manera distinta, por lo que tiende a hacer más visible la línea de unión entre cámaras.

8. Anexo II. Figuras

Figura 6. Estándares de resolución de los videos 360.

Frame Aspect Ratio	Resolución (Pixel cuadrado 1.0)
12K	12288x6133
8K	8192x4096
6K	6144x3072
5K	5120x2560
4K	4096x2048
4K HD	3840x1920
2K	2048x1024
1080p	1920x960
720p	1280x640

Fuente: Unravel. Disponible en: http://unravel.com.au/files/Unravel_Resolution_CheatSheet.pdf

Figura 13. Tabla de estándares para realizar panoramas utilizando el punto no parallax. Imágenes extraídas de Panotools.

Focal Length	(Micro-) Four Thirds (crop factor 2.0)			
	Portrait		Landscape	
	pics/360°	rotate	pics/360°	rotate
8mm fisheye	4	90°	3	120°
7mm rectilinear	5	72°	4	90°
9mm rectilinear	6	60°	5	72°
11mm rectilinear	7	51,5°	6	60°
12mm rectilinear	8	45°	6	60°
14mm rectilinear	9	40°	7	51,5°
17mm rectilinear	11	33°	8	45°
18mm rectilinear	11	33°	9	40°
25mm rectilinear	15	24°	12	30°
35mm rectilinear	21	17°	16	22,5°
40mm rectilinear	24	15°	18	20°

Fuente: Panot

Focal Length Lens Type	crop factor 1.5/1.6				Full Frame			
	Portrait		Landscape		Portrait		Landscape	
	pics/360°	rotate	pics/360°	rotate	pics/360°	rotate	pics/360°	rotate
8mm fisheye	4	90°			3	120°	3	120°
10.5mm fisheye	6	60°			3	120°		
16mm fisheye					6	60°	4	90°
10mm rectilinear	6	60°	4	90°				
14mm rectilinear	8	45°	6	60°				
18mm rectilinear	10	36°	8	45°				
24mm rectilinear	12	30°	8	45°	8	45°	6	60°
28mm rectilinear	15	24°	10	36°	12-8	30°-40°	8-6	40°-60°
35mm rectilinear	18	20°	12	30°	18-12	20°-30°	12-8	30°-40°
50mm rectilinear	24	15°	18	20°	18	20°	12	30°
70mm rectilinear	36	10°	24	15°	24	15°	15	24°

ools. Disponible en http://wiki.panotools.org/Entrance_Pupil_Database [consulta: 1 junio 2017]