matematicalia
revista digital de divulgación matemática

TECNOLOGÍA

Volumen 7, núm. 4 (diciembre 2011)
10 páginas

http://www.matematicalia.net
ISSN: 1699-7700

Recibido: 09 de enero de 2012
Publicado: 06 de febrero de 2012

Tecnología y matemáticas en el cine y la televisión en tres dimensiones

Fernando Pérez Nava
e-mail: fdoperez@ull.es

Manuel Rodríguez Valido
e-mail: mrvalido@ull.es

Eduardo Magdaleno
e-mail: emagcas@ull.es
página web: http://webpages.ull.es/users/emagcas

Grupo de Sensores Inteligentes
Universidad de La Laguna

Resumen

Desde la introducción del cine y la televisión, se ha realizado un gran esfuerzo tecnológico para perfeccionar la experiencia visual de los espectadores. Las mejoras en la calidad de imagen y sonido han contribuido a una experiencia cada vez más cercana a nuestro mecanismo de percepción. Se cree que el cine y la televisión tridimensionales (3DTV), que permite a los espectadores ver el contenido en tres dimensiones, es el siguiente paso en la evolución de este campo, puesto que proporciona una forma más natural de percepción visual. Aunque ha habido varios intentos de introducir tanto el cine como la televisión en 3D, su aceptación generalizada se considera cada vez más factible gracias a los avances en el campo de los sistemas de captura 3D, el procesamiento de imágenes y los dispositivos de visualización 3D. En este artículo se muestra una revisión de los conceptos teóricos y tecnológicos del cine y la televisión en 3D. Para una mejor comprensión de dichos conceptos se construye de forma práctica un sistema de 3DTV utilizando materiales asequibles y fácilmente disponibles.

1. Introducción

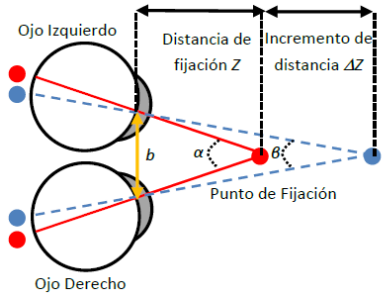
Es posible que el lector de este artículo haya visto recientemente una película en 3D sentado en la butaca de un cine. Le será familiar entonces esa sensación de objetos que se presentan justo enfrente del espectador dando la sensación de que se pueden tocar, o habrá sentido la necesidad de ladear la cabeza para evitar golpearse con algún elemento de la película. El cine en 3D añade una dosis de realismo a la historia que se desarrolla ante nuestros ojos, puesto que ésta es nuestra forma natural de visión. Esto se debe a que la evolución provocó que la mayoría de los depredadores tengan sus dos ojos mirando hacia delante, lo que les permite percibir la profundidad, ayudándoles a calcular las distancias al abalanzarse sobre su presa.

Así como las películas en color se impusieron a las películas en blanco y negro, puesto que nuestro sistema de percepción visual es capaz de apreciar el color, es razonable suponer que las películas en 3D se impondrán sobre el formato visual 2D, mayoritario en la actualidad. De hecho, la idea del cine en 3D no es nueva en absoluto y sus orígenes se sitúan a comienzos del siglo XIX [1]. Su desarrollo tuvo un impulso notable a mediados del siglo XX, pero las dificultades tecnológicas hicieron que el efecto 3D fuera de mediana calidad. El desarrollo a finales del pasado siglo de nuevos sistemas de captura digital en 3D, junto a la aparición de nuevas técnicas de procesado digital de imagen, ha dado como resultado una sensación 3D de alta calidad, generando un renacimiento de esta tecnología. Aunque los costos de producción de una película en 3D suponen aproximadamente un 20% de incremento sobre los de una película convencional, los ingresos obtenidos compensan con creces el incremento en la inversión. Las películas 3D más recientes han generado el triple de ingresos que las

copias en 2D (es por esto que algunos productores de cine llaman también al 3D como "3 veces Dólares"). Para su asentamiento definitivo, el fenómeno 3D deberá, no obstante, superar una serie de retos económicos, tecnológicos y artísticos. En este trabajo abordaremos varios de los aspectos tecnológicos del cine y la televisión en 3D desde un punto teórico y práctico. Intentaremos acercar esta tecnología mostrando cómo es posible construir un sistema de televisión en 3D a partir de materiales asequibles y fácilmente disponibles.

1.1. Evolución del cine y la 3DTV

La percepción 3D estereoscópica fue explicada por primera vez por Charles Wheatstone en 1838 [1]. Detalló que, debido a que cada ojo ve el mundo desde posiciones horizontales ligeramente diferentes, la imagen que se forma en cada ojo es distinta de la otra. De esta forma, los objetos que se sitúan a diferentes distancias de los ojos proyectan imágenes que difieren en su ángulo (*disparidad angular*, η). Esto proporciona la información necesaria para calcular su distancia como se puede ver en la Figura 1. Wheatstone mostró que para generar la sensación 3D basta con que cada ojo reciba cada una de estas imágenes por separado, e inventó el estereoscopio. Los hermanos Lumière se basaron en este invento para mostrar imágenes 3D en movimiento en la feria de París de 1903. La primera película completa en 3D se presentó en Los Ángeles en 1922, y en 1928 John Logie Baird aplicó los principios de la estereoscopía a un dispositivo experimental de 3DTV. A pesar de estos desarrollos iniciales y de distintos esfuerzos para asentar la tecnología 3D, especialmente en la industria cinematográfica de los años 50, el éxito comercial no se produjo debido a las deficiencias técnicas que provocaban una insuficiente calidad del efecto 3D. De hecho, las únicas aplicaciones 3D con cierto éxito se situaron fuera del campo multimedia, en ciertos nichos de mercado como los simuladores y dispositivos de visualización. La única excepción en el campo multimedia lo constituye el IMAX 3D, que ha tenido un cierto éxito comercial desde su lanzamiento en 1986 y que desde entonces promueve un número limitado de producciones 3D cada año. La situación comenzó a cambiar de forma gradual a principio de los años 90 del siglo pasado con la transición de los dispositivos de captura y transmisión analógicas a servicios digitales. Finalmente, las películas 3D han obtenido un gran éxito comercial a principios de los 2000, culminando con el éxito sin precedentes de la película *Avatar* a finales de 2009. Además, recientemente varias emisoras de televisión, como Canal+, Sky, ESPN, Fox Deportes o la BBC, comenzaron a emitir en vivo algunos eventos deportivos en 3D.



$$\text{Para } \alpha, \beta \approx 0, \alpha \approx \frac{b}{Z}, \beta \approx \frac{b}{Z+\Delta Z} \text{ y si } \Delta Z \approx 0,$$
$$\eta = \alpha - \beta \approx \frac{b}{Z^2}\Delta Z$$

Figura 1. Expresión de la disparidad angular η para un incremento ΔZ desde la distancia de fijación Z.

1.2. La cadena de procesamiento 3D

Desde el momento en que se adquiere la información en 3D hasta que es visualizada por el espectador, es necesario que pase por una serie de fases. Como se muestra en la Figura 2, un sistema de vídeo 3D consiste en las etapas de captura de contenidos 3D, representación del contenido 3D, compresión, codificación y transmisión de datos, decompresión, procesamiento de la señal y visualización de los datos 3D [2]. El objetivo de la etapa de captura de contenidos es la obtención de contenidos 3D. Hay tres formas principales de generación de contenidos: la utilización de una cámara 2D cuya imagen es procesada para obtener una representación 3D, la utilización de una cámara de profundidades que captura un vídeo 2D junto con la profundidad de cada elemento de

la imagen, o la utilización de un conjunto de N cámaras que capturan la escena desde diversos puntos de vista. Los datos obtenidos en la etapa de captura se procesan y transforman para obtener una representación 3D en función de la aplicación que se requiera. El tamaño de esta representación es mucho mayor que la del cine o la televisión tradicionales. Por ejemplo, una representación basada en el vídeo y la profundidad multiplica por 2 los datos obtenidos, mientras que la utilización de N cámaras la multiplica por N. La utilización de los métodos de codificación tradicionales sobre cada cámara individual genera un elevado volumen de datos a transmitir y es ineficiente, puesto que los datos son altamente redundantes; por tanto, es necesaria la utilización de nuevos métodos de codificación. Finalmente, los datos son decodificados y procesados para los diferentes tipos de dispositivos de visualización 3D.

Figura 2. La cadena de procesamiento 3D.

1.3. Dispositivos de captura 3D

Los dispositivos de captura más habituales de un sistema de vídeo 3D (Figura 3) suelen ser un par de cámaras 2D, y una cámara 2D junto con una cámara de profundidades o un conjunto de N cámaras 2D. En caso de ser requerida, la información de profundidad puede ser obtenida mediante una cámara de profundidades o estimada a partir de las cámaras 2D. En la actualidad la mayor parte de los contenidos 3D se generan utilizando un par de cámaras o una cámara 2D dotada de una óptica especial. También es posible usar los contenidos obtenidos a partir de una única cámara 2D para obtener contenidos 3D mediante técnicas de conversión 2D-3D. Aunque estrictamente el efecto 3D se puede conseguir a partir de un par de cámaras, la utilización de un conjunto de N cámaras ofrece como ventaja la obtención de contenidos 3D desde diversos puntos de vista. Esto permite que el usuario pueda elegir el punto de vista para ver la escena, o que varios usuarios puedan ver la escena desde distintos puntos de vista de manera simultánea. La utilización de varias cámaras no está exenta de problemas, puesto que es necesario que todas sean consistentes tanto en la respuesta del color como en el sincronismo. Una alternativa a la utilización de N cámaras la constituyen las *cámaras plenópticas*, en las que una óptica especial permite que una cámara 2D se comporte como N cámaras independientes.

Figura 3. Dispositivos de captura 3D. Izquierda: par de cámaras 2D. Centro: cámara 3D a partir de una cámara 2D con una óptica especial. Derecha: sistema multicámara.

1.4. Transmisión y codificación de la señal 3D

La información a transmitir con N cámaras en alta definición (HD) requiere N veces el máximo ancho de banda que la HDTV. Por tanto, es necesaria la utilización de esquemas de codificación eficientes para transmitir toda la información [3]. Diversos organismos como el ISO-MPEG y el ITU-VCEG han impulsado el proceso de estandarización para los medios digitales, incluyendo los relacionados con el vídeo 3D. El sistema básico de captura 3D con dos cámaras ya era soportado por el formato MPEG-2 desde mediados de los 90. Sin embargo, la inexistencia de un mercado significativo de vídeo 3D limitó su uso. Una primera solución simple para la 3DTV y compatible con la televisión 2D fue la utilización de un esquema que dividía por dos la resolución de las imágenes de las dos cámaras y las

combinaba en una nueva imagen, que era codificada y transmitida como si fuera una imagen 2D. El inconveniente de este sistema era la obtención de imágenes 3D con la mitad de la resolución que las cámaras originales. Otras soluciones más avanzadas utilizan las redundancias entre las cámaras y proporcionan la forma de obtener imágenes 3D sin pérdida de resolución. La codificación multivista para N cámaras es una extensión reciente de la codificación H.264/AVC, que proporciona actualmente la forma más eficiente de codificar la información suministrada por N cámaras. Finalmente, la transmisión conjunta de vídeo 2D y profundidad se realiza utilizando un estándar conocido como MPEG-C Part3.

1.5 Visualización de la señal 3D

El último paso para mostrar el contenido 3D al espectador lo constituye el proceso de visualización [4]. El vídeo 3D es decodificado y postprocesado para el dispositivo de visualización 3D. Todos los dispositivos utilizan la idea de Wheatstone de enviar imágenes distintas a cada ojo.

Existe una amplia variedad de dispositivos de visualización (Figura 4). Los más utilizados necesitan elementos auxiliares (gafas) para obtener la sensación 3D. El primer tipo de gafas utilizado fueron las *gafas anaglifos* (rojo/azul). Los colores actúan como filtros que permiten enviar información distinta a cada ojo. Su principal inconveniente es la mala reproducción del color y un efecto 3D de mediana calidad. Para evitar estos problemas surgieron las *gafas polarizadas*. Su funcionamiento se basa en mostrar sobre la pantalla dos imágenes superpuestas a través de diferentes filtros polarizados ortogonales. Las gafas contienen también un par de filtros polarizados con la misma orientación que la pantalla, que bloquean en cada ojo la luz polarizada en la dirección ortogonal. Sus principales problemas son la pérdida de luminosidad de la imagen 3D y su precio, superior al de las gafas anaglifos. Las gafas polarizadas se utilizan principalmente en los cines, aunque también en algunas pantallas de televisión. Una última alternativa la constituyen las *gafas de obturadores activos*. Cada lente del ojo contiene un cristal líquido transparente que tiene la propiedad de oscurecerse cuando se le aplica una corriente eléctrica. Las gafas se controlan con un transmisor que envía una señal a las gafas indicando qué ojo oscurecer, dependiendo de que la imagen que se muestra en la pantalla sea la correspondiente al ojo derecho o al izquierdo. Éste es el tipo de gafas que se utiliza más habitualmente en los televisores 3D. Tiene el inconveniente de que las gafas son caras, necesitan recargar su batería, y la imagen 3D sufre una pérdida de luminosidad.

Otra alternativa para la visualización 3D la constituyen los *dispositivos autoestereoscópicos*, en los que no es necesaria la utilización de gafas. En la actualidad existen dos tipos principales: el de barrera y el lenticular. Los dispositivos *de barrera* colocan una barrera sobre la pantalla conteniendo unas aberturas que dirigen cada imagen a su ojo correspondiente. Su principal inconveniente es la pérdida de luminosidad de la imagen. En los dispositivos *lenticulares* se sitúa sobre la pantalla una hoja lenticular formada por una serie de lentes semicilíndricas, que son las encargadas de enviar la imagen correspondiente a cada ojo. Los dispositivos lenticulares conservan la luminosidad de la imagen. Tanto los dispositivos de barrera como los lenticulares se pueden emplear para mostrar N vistas; en este caso, la resolución de la imagen 3D se divide por N. Ambos tipos tienen dificultades para ser empleados en grandes pantallas.

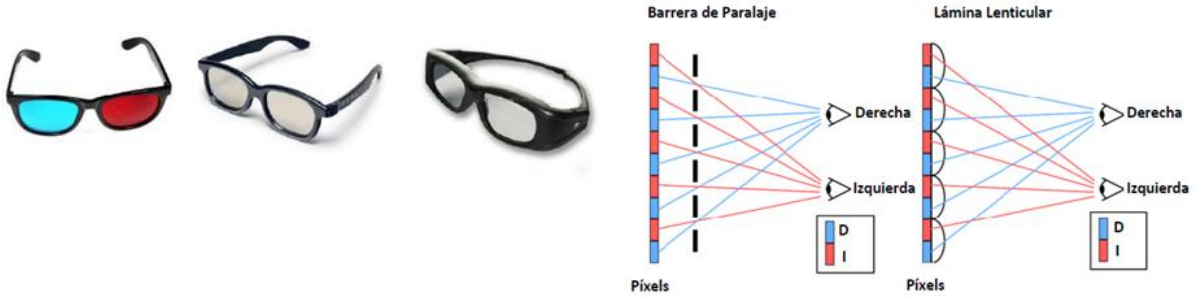


Figura 4. De izquierda a derecha: gafas anaglifos, polarizadas y activas, y dispositivos autoestereoscópicos.

2. Renderizado basado en una imagen de profundidades

La forma más simple de representar una escena 3D es utilizar dos imágenes diferentes, una para el ojo derecho y otra para el ojo izquierdo. Otra alternativa a esta representación consiste en la utilización de una imagen junto con la profundidad de cada uno de sus elementos. A partir de esta representación se pueden generar una o más vistas virtuales de la escena 3D, que se mostrarán en un televisor 3D por medio de la técnica conocida como *Renderizado Basado en Imagen de Profundidades* (RBIP) [1]. El RBIP tiene varias ventajas sobre la aproximación clásica basada en dos vistas: permite ajustar de forma simple el contenido 3D a diversos dispositivos de visualización 3D; permite utilizar de forma simple el contenido generado en 2D mediante técnicas de conversión 2D-3D; y permite ajustar el efecto estereoscópico a las preferencias personales del usuario.

Para explicar el proceso de renderizado es necesario mostrar el modelo de la *cámara estenopeica* (Figura 5). Al proyectar un punto $\boldsymbol{M}$ en el mundo sobre el plano de la imagen, se genera un punto $\boldsymbol{m}$ sobre dicho plano. Las coordenadas de ambos puntos se relacionan a través de los parámetros de la cámara mediante una transformación en dos pasos. La orientación de la cámara y su posición determina la relación existente entre el sistema de coordenadas de la cámara y el sistema de coordenadas del mundo. Esta relación se describe mediante una matriz de rotación $\mathbf{R}$ y un vector de traslación $\boldsymbol{t}$ (ó $-\mathbf{R}\boldsymbol{C}$, donde $\boldsymbol{C}$ representa las coordenadas del mundo de la cámara) que se denominan *parámetros extrínsecos* de la cámara.

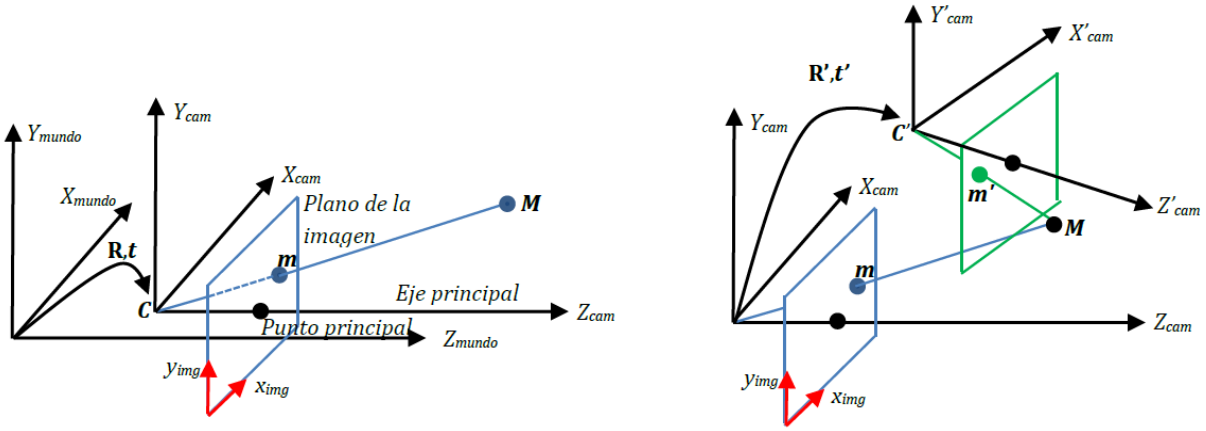


Figura 5. Izquierda: geometría de la cámara estenopeica. Derecha: geometría del RBIP.

La relación entre las coordenadas de la cámara y las coordenadas de la imagen queda determinada por varios parámetros denominados *intrínsecos*: $\alpha_x = fm_x$, $\alpha_y = fm_y$, donde f representa la focal de la cámara (distancia entre el origen de la cámara y el plano de la imagen) y donde m_x y m_y son factores de escala que transforman las unidades en píxeles a distancias, el parámetro γ representa un coeficiente de asimetría entre los ejes x_{img} e y_{img} que suele ser cero, y los parámetros u_0 y v_0 representan el punto principal, que idealmente debería estar situado en el centro de la imagen. La proyección de un punto en 3D al plano de la imagen queda, por tanto, determinada por los parámetros extrínsecos de la cámara $\mathbf{R}$ (matriz de 3x3) y $\boldsymbol{t}$ (vector de 3x1) y los parámetros intrínsecos de la cámara, que se agrupan en una matriz $\mathbf{K}$ de tamaño 3x3. Utilizando coordenadas homogéneas podemos escribir:

$$Z_{cam}\boldsymbol{m}_{img} = \mathbf{K}(\mathbf{I}|\mathbf{0})\boldsymbol{M}_{cam}, \quad \boldsymbol{M}_{cam} = \begin{pmatrix} \mathbf{R} & \boldsymbol{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \boldsymbol{M}_{mundo}, \quad \boldsymbol{t} = -\mathbf{R}\boldsymbol{C},$$

$$\boldsymbol{M}_{mundo} = \begin{pmatrix} X_{mundo} \\ Y_{mundo} \\ Z_{mundo} \\ 1 \end{pmatrix}, \quad \boldsymbol{M}_{cam} = \begin{pmatrix} X_{cam} \\ Y_{cam} \\ Z_{cam} \\ 1 \end{pmatrix}, \quad \boldsymbol{m}_{img} = \begin{pmatrix} x_{img} \\ y_{img} \\ 1 \end{pmatrix}, \quad \mathbf{K} = \begin{pmatrix} \alpha_x & \gamma & u_0 \\ 0 & \alpha_y & v_0 \\ 0 & 0 & 1 \end{pmatrix}. \qquad (1^{a\ b\ c})$$

El proceso para generar imágenes virtuales es simple si disponemos de una cámara de color para la que poseemos la profundidad (coordenada Z) de cada píxel. Sin pérdida de generalidad supongamos, por simplificar, que dicha cámara se encuentra en el origen de coordenadas del mundo; entonces $\mathbf{R}$ es la matriz identidad $\mathbf{I}$ y $\boldsymbol{t}$ es el vector $\mathbf{0}$. Utilizando (1) tenemos que un elemento de la imagen $\boldsymbol{m}$ con coordenadas homogéneas $\boldsymbol{m}_{img}$ se reproyecta en el sistema de coordenadas del mundo como

$$\boldsymbol{M} = Z_{cam}\mathbf{K}^{-1}\boldsymbol{m}_{img}, \qquad (2)$$

donde Z_{cam} es la profundidad correspondiente a $\boldsymbol{m}_{img}$ que proporciona la cámara de distancias. Si ahora queremos calcular la imagen virtual $\boldsymbol{m'}$ del punto $\boldsymbol{m}$ con una cámara trasladada mediante un vector $\boldsymbol{t}'$, que se encuentra rotada con respecto a la cámara original mediante una matriz $\mathbf{R'}$ y cuyos parámetros intrínsecos se encuentran descritos por una matriz $\mathbf{K}'$, tenemos, sustituyendo (2) en (1):

$$Z'_{cam}\boldsymbol{m}'_{\boldsymbol{img}} = Z_{cam}\mathbf{K}'\mathbf{R}'\mathbf{K}^{-1}\boldsymbol{m}_{img} + \mathbf{K}'\boldsymbol{t}'. \qquad (3^{a\ b\ c\ d\ e\ f})$$

Por tanto, si se conocen todos los valores de profundidad de la cámara original, se puede recuperar la imagen que obtendría la cámara virtual aplicando (3) a todos los valores de la imagen original.

3. Un sistema de televisión 3D utilizando Kinect

La televisión 3D se encuentra actualmente en fase de expansión; esto ha provocado que las principales empresas tecnológicas hayan hecho grandes inversiones en investigación y desarrollo. Sin embargo, y a un nivel básico, es posible construir un sistema de 3DTV utilizando materiales asequibles, matemáticas básicas y conocimientos elementales de programación. En esta sección detallaremos paso a paso cómo hacerlo. Para ello utilizaremos el sensor de videojuegos Kinect como dispositivo de captura 3D y utilizaremos la técnica del RBIP para generar N vistas virtuales utilizando (3) y, de esa forma, mostrar las imágenes 3D en un monitor 3D autoestereoscópico que construiremos nosotros mismos. Nuestro objetivo es ilustrar de forma práctica parte de los contenidos presentados en las secciones anteriores.

3.1. Dispositivo de captura: sensor Kinect

El sensor Kinect es un dispositivo, diseñado originalmente para la consola de juegos Xbox 360, que es capaz de generar simultáneamente una señal de vídeo en color de la escena junto a la profundidad de todos los elementos que aparecen en ésta. Para ello, el sensor está equipado con dos cámaras (Figura 6): una cámara infrarroja (IR) que se utiliza para la detección de profundidad, y una cámara en color. La detección de profundidades se realiza mediante el principio de la luz estructurada. Un laser IR proyecta un conjunto de puntos a la escena cuya posición se captura mediante la cámara IR; estudiando la deformación de estos puntos se puede recuperar la distancia de los elementos de la escena. Las características técnicas del sensor son las siguientes: resolución de la cámara IR de 640x480 píxeles, resolución de la cámara de color de 1600x1200 píxeles, máxima frecuencia de refresco de 60 imágenes/segundo, rango de recuperación de distancias de 0.8-3.5m, una resolución espacial de 3mm a 2m de distancia y una resolución Z en profundidad de 1cm a 2m de distancia. Desde su lanzamiento se reconoció el potencial del sensor para realizar tareas distintas de los juegos de ordenador. El español Héctor Martín fue el primero en desarrollar un controlador independiente para el sensor. A este desarrollo le siguieron varios entornos de programación de código abierto, y en la actualidad se pueden encontrar dos sistemas dominantes: OpenKinect [5] y OpenNI [6]. OpenKinect es un proyecto de código abierto basado en la realización de ingeniería inversa sobre el sensor. OpenNI es un sistema de programación de código abierto que se conecta con el sensor mediante un entorno de código cerrado proporcionado por PrimeSense, que es la empresa que desarrolló la tecnología en la que se basa Kinect. El precio del sensor es del orden de unos 150 euros (2011).

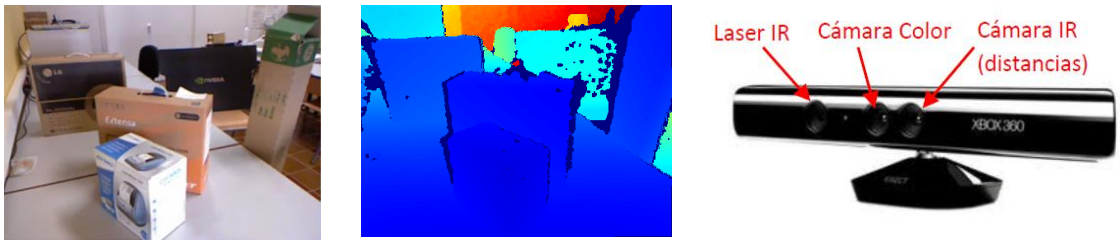


Figura 6. De izquierda a derecha: imagen en color de Kinect, imagen de distancias, imagen del sensor.

3.2. Calibración

Dado que utilizaremos la técnica del RBIP para generar las N vistas que necesitará el monitor 3D, necesitamos que cada píxel de la imagen de color tenga asociada una distancia. Para ello podemos utilizar la ecuación (3) y asignar al punto $\boldsymbol{m'}$ de la imagen de color la distancia medida para el punto $\boldsymbol{m}$ en la imagen de profundidades. No obstante, para poder utilizar (3) necesitamos determinar todos los parámetros desconocidos de la ecuación: las matrices intrínsecas $\mathbf{K}$ y $\mathbf{K'}$ de la cámara de distancias y color de Kinect, junto con la matriz de rotación $\mathbf{R'}$ y la traslación $\boldsymbol{t'}$ entre ambas cámaras. A la estimación de estos parámetros se le llama *proceso de calibración*. El proceso de calibración de cámaras es un campo bien estudiado y para el que existen varias soluciones de software libre. En la librería OpenKinect el proceso de calibración se lleva a cabo tomando varias imágenes de un plano sobre el que se ha colocado un patrón similar a un tablero de ajedrez. A partir de las esquinas de los cuadrados del tablero se obtiene una estimación de la matriz $\mathbf{K'}$ con los parámetros intrínsecos de la cámara de color. Utilizando las esquinas se obtiene también una estimación de la matriz $\mathbf{K}$ con los parámetros intrínsecos de la cámara de profundidad. La localización de las esquinas tanto en la imagen de color como en la de profundidad permite estimar la orientación relativa entre ambas, determinada por $\mathbf{R'}$ y $\boldsymbol{t'}$. Este proceso puede eliminarse utilizando el entorno OpenNI, ya que éste proporciona los parámetros del sensor establecidos en el proceso de fabricación. Los parámetros intrínsecos que calculan ambos entornos de programación son más complejos que los descritos en (1), pues incluyen las posibles deformaciones de las lentes en ambas cámaras y la transformación de las disparidades que mide el sensor Kinect a distancias Z.

3.3. Alineación de la imagen de profundidad y de color

Una vez aplicado el proceso de calibración de la sección 3.2, usando (3) se asigna a cada píxel de la imagen de color su valor de profundidad tomado de la cámara de profundidades. Sin embargo, el proceso se complica debido a que la imagen de profundidad tiene una cantidad significativa de píxeles sin distancia. Esto es debido a diversas causas, tales como objetos fuera del rango de distancias del sensor, superficies reflectantes u oclusiones (Figura 7), y provoca que la imagen de color tenga píxeles para los que se desconoce su distancia. Para completar la imagen en color con todas las profundidades se emplean los píxeles para los que se conoce tanto el color como la profundidad, y se utiliza el principio intuitivo de que en regiones de la imagen donde hay poca variación de color, la variación de profundidad debe ser también pequeña. De forma analítica, para un píxel $\boldsymbol{x}$ sin profundidad se estima $Z(\boldsymbol{x})$ como:

$$Z(\boldsymbol{x}) = \sum_{\boldsymbol{y}\in E(\boldsymbol{x})} s(\boldsymbol{x},\boldsymbol{y}) Z(\boldsymbol{y}), s(\boldsymbol{x},\boldsymbol{y}) = \frac{w(\boldsymbol{x},\boldsymbol{y})}{\sum_{\boldsymbol{y'}\in E(\boldsymbol{x})} w(\boldsymbol{x},\boldsymbol{y'})}, w(\boldsymbol{x},\boldsymbol{y}) = \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|^2}{2\sigma_d^2} - \frac{\|\boldsymbol{c}(\boldsymbol{x})-\boldsymbol{c}(\boldsymbol{y})\|^2}{2\sigma_c^2}\right),$$

donde $E(\boldsymbol{x})$ es un conjunto de puntos de la imagen de color alrededor de $\boldsymbol{x}$ para los que se conoce su profundidad. Cada uno de estos puntos $\boldsymbol{y}$ colabora en la formación de $Z(\boldsymbol{x})$ en función de su cercanía geométrica $\|\boldsymbol{x}-\boldsymbol{y}\|^2$ en la imagen a $\boldsymbol{x}$ y de la cercanía $\|\boldsymbol{c}(\boldsymbol{x})-\boldsymbol{c}(\boldsymbol{y})\|^2$ de su color $\boldsymbol{c}(\boldsymbol{y})$ al color de $\boldsymbol{x}$ denotado por $\boldsymbol{c}(\boldsymbol{x})$. Este tipo de filtrado se conoce como *cross-bilateral filtering*, y su código de programación se puede encontrar en [7].

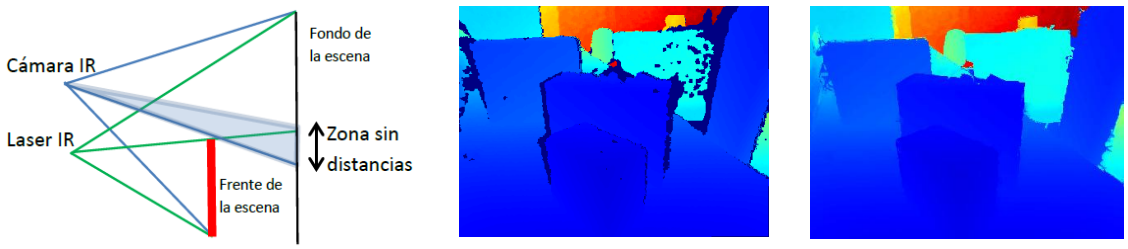


Figura 7. De izquierda a derecha: región sin distancia en la cámara de profundidad debida a oclusiones, imagen de profundidad con regiones sin distancia (en azul oscuro), e imagen de profundidad corregida.

3.4 Generación de vistas

Una vez obtenida la imagen en color con profundidad, generaremos N vistas virtuales para el monitor 3D autoestereoscópico. Para ello utilizaremos la disposición de paralaje cero (Figura 8). En esta disposición la ecuación de RIBD (3) se simplifica, puesto que se verifican las relaciones en (4) y un punto de coordenadas (x_{img}, y_{img}) con profundidad Z_{cam} se proyecta en la imagen virtual de acuerdo a (5):

$$\boldsymbol{K}' = \boldsymbol{K} + \begin{pmatrix} 0 & 0 & h \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \boldsymbol{R} = \boldsymbol{I}, \boldsymbol{t} = \begin{pmatrix} t_x \\ 0 \\ 0 \end{pmatrix}, h = -t_x \frac{\alpha_x}{Z_{cam}}, \tag{4}$$

$$\left(x'_{img}, y'_{img}\right) = \left(x_{img}, y_{img}\right) + \left(t_x \alpha_x \left(\frac{1}{Z_{cam}} - \frac{1}{Z_c}\right), 0\right). \tag{5}$$

Entonces, variando $t_x = m\Delta x,\ m = -r, \ldots, s,\ r + s + 1 = N$, se obtienen todas las vistas virtuales.

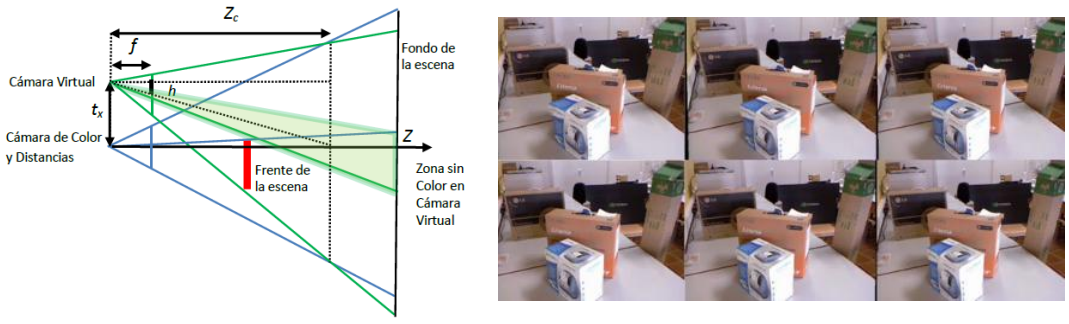


Figura 8[a] [b]. Izquierda: geometría de la generación de vistas y oclusiones. Derecha: ejemplo práctico.

Un problema que aparece a la hora de generar las vistas virtuales es el de las oclusiones: la cámara virtual ve una zona de la escena que no aparece en la cámara con color y distancias (Figura 8). Esto provoca que aparezcan regiones en las imágenes virtuales donde no hay información de color. Estas oclusiones son más complejas que las de la sección 3.3, puesto que en estas regiones no hay información de color ni de distancia. Existen diversas técnicas [1] para resolver este problema, desde propagar el color del fondo de la imagen hasta procedimientos más sofisticados que cortan y pegan pequeñas regiones del fondo de la imagen.

3.5 Generación de información para la pantalla lenticular

Como se señaló en la sección 1.5, los monitores autoestereoscópicos lenticulares se basan en colocar un conjunto de lentes cilíndricas sobre un monitor de forma que la información 3D se dirija de forma separada a los ojos de los espectadores que se encuentran enfrente de éste. Para realizar un sistema básico de visualización 3D hemos utilizado el monitor LCD de un microrordenador ASUS EeePC 901 y lo hemos cubierto con una lámina de lenticular Asus Eee PC 901 i-Art 3d lens sheet, cuyo precio es de 99 dólares (2011).

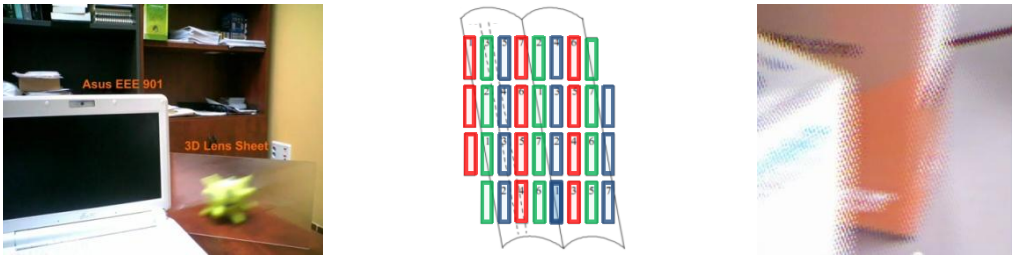


Figura 9[a b c]. Izquierda: monitor LCD y lámina de microlentes. Centro: posición relativa de las microlentes sobre el monitor [8]. Derecha: detalle de la imagen preparada para el monitor 3D.

Las microlentes se encuentran orientadas con respecto al monitor LCD con un ángulo α (Figura 9). Para visualizar la imagen se debe detallar cómo colocar las N vistas calculadas en la sección 3.4 sobre el monitor. Para ello renumeraremos las N vistas como $0,\ldots,N-1$ y descompondremos cada píxel del monitor LCD en 3 subpíxeles de colores Rojo, Verde y Azul (Figura 9). Rellenaremos cada subpíxel con coordenadas (u,v) del monitor con el subpíxel (u,v) de la vista $n(u,v)$ mediante la fórmula [8]

$$n(u,v) = mod\big(u + u_{desp} - 3v\tan(\alpha), N_{vistas/lente}\big)/N_{vistas/lente} \times N,$$

con $u_{desp,}$ α, $N_{vistas/lente,}$ N parámetros que dependen de la lámina lenticular y el monitor LCD. Un detalle de esta composición se puede ver en la Figura 9. El resultado completo del procesamiento puede visualizarse en http://www.youtube.com/watch?v=wBZTM-fn3o0.

4. Conclusiones

La tecnología de visualización 3D es tan antigua como el cine. Su interés radica en que la visión en 3D es nuestra forma natural de percibir el mundo. La introducción de la digitalización de vídeo y sonido a finales del siglo pasado ha proporcionado un nuevo impulso a esta tecnología que puede provocar su definitivo asentamiento. En la actualidad, la posibilidad de ver una película en 3D, tanto en el cine como en el hogar, está al alcance del público en general. En este trabajo se ha intentado aunar el marco general de los medios 3D y sus conceptos teóricos con la aplicación práctica mediante el desarrollo de un sistema de captura y visualización 3D de bajo coste. No obstante, para que el 3D se imponga definitivamente debe todavía simplificar el proceso de captura 3D, mejorar el proceso de compresión y continuar el desarrollo de dispositivos de visualización 3D que no necesiten de gafas. Estamos convencidos que del trabajo conjunto de investigadores en diversas disciplinas, entre ellas las matemáticas, saldrá la solución definitiva que nos proporcionará una experiencia visual natural y cercana a la de la vida real.

Referencias

[1] [a b c d] O. Schreer, P. Kauff , T. Sikora (eds.): *3D Videocommunication: Algorithms, concepts and real-time systems in human centered communication*. John Wiley and Sons, 2010.

[2] ^ D. Minoli: *3DTV Content Capture, Encoding and Transmission: Building the Transport Infrastructure for Commercial Services*. John Wiley and Sons, 2010.

[3] ^ A. Smolic, K. Mueller, N. Stefanoski, J. Ostermann, A. Gotchev, G. B. Akar, G. Triantafyllidis, A. Koz: Coding algorithms for 3DTV – a survey. *IEEE Transactions on Circuits and Systems for Video Technology* 17, no. 11 (2007), 1606-1621.

[4] ^ P. Benzie, J. Watson, P. Surman, I. Rakkolainen, K. Hopf, H. Urey, V. Sainov, C. von Kopylow: A survey of 3DTV displays: Techniques and technologies. *IEEE Transactions on Circuits and Systems for Video Technology* 17, no. 11 (2007), 1647-1658.

[5] ^ *Openkinect*, http://www.openkinect.org.

[6] ^ *OpenNI*, http://www.openni.org.

[7] ^ S. Paris, P. Kornprobst, J. Tumblin, F. Durand: *A Gentle Introduction to Bilateral Filtering*, http://people.csail.mit.edu/sparis/bf_course.

[8] a b C. van Berkel: Image preparation for 3D-LCD. *Proc. SPIE 3639* (1999), 84-89.

Sobre los autores

Fernando Pérez Nava es profesor titular de la Universidad de La Laguna desde 2004. Se licenció en Matemáticas con la especialidad de Estadística en 1989 y se doctoró en 2001 por la Universidad de La Laguna, donde ha desempeñado varios cargos de gestión. Es autor de más de 30 publicaciones científicas relacionadas con la visión por ordenador y el reconocimiento de patrones. Ha participado en varios proyectos nacionales y europeos y colaborado con diversas empresas del sector privado, donde estuvo trabajando hasta incorporarse a la universidad. Es coautor de un libro sobre visión artificial y una patente sobre cámaras 3D. Sus intereses incluyen los sistemas de captura de campos de luz y el diseño de sensores inteligentes.

Manuel Rodríguez Valido es licenciado en Físicas y doctor por la Universidad de la Laguna. Actualmente es profesor titular de Universidad en el área de Tecnología Electrónica. Imparte docencia en la Escuela Técnica Superior de Ingeniería Informática y en el Master en Ingeniería Electrónica de esta universidad, en las materias *Electrónica digital* y *Diseño electrónico*. Ha desempeñado diversos cargos de gestión, y actualmente es el director del citado Master. Cuenta con diversas publicaciones en las áreas de sensores inteligentes, procesado y codificación de imágenes 3D, diseño de sistemas en FPGAs y redes de sensores inalámbricas.

Eduardo Magdaleno obtuvo la licenciatura en Ciencias Físicas (especialidad Astrofísica) por la Universidad de La Laguna en 1999. En 2002 se licenció en Ingeniería Electrónica, y en 2009 obtuvo el grado de doctor en Tecnología Electrónica por la misma Universidad. Desde 2002 ha impartido docencia de *Electrónica digital* y FPGAs en las titulaciones de Náutica, Ingeniería en Automática y Electrónica Industrial, Ingeniería Electrónica y Física. Su investigación incluye el desarrollo de algoritmos hardware eficientes en FPGA, sensores inteligentes y visión 3D.