

VISUALIZACIÓN TRIDIMENSIONAL

INFORME DE PROYECTO DE GRADO

Santiago Zabalza

Tutores:

Dr. Tomás Laurenzo

Ing. Christian Clark

Octubre, 2014

Resumen

Este trabajo de proyecto de grado presenta un estudio sobre las diferentes técnicas de visualización tridimensional y el desarrollo de un prototipo de dispositivo de visualización tridimensional autoestereoscópico sin la necesidad de aditamentos adicionales para el usuario.

La técnica utilizada para la creación del prototipo está basada en una barrera de paralaje y en combinación con seguimiento de la posición del usuario, utilizando una técnica no intrusiva, permite la visualización de la imagen tridimensional y ofrece libertad de movimiento.

El prototipo construido alcanza una calidad de imagen satisfactoria que permite al observador percibir correctamente imágenes tridimensionales. Si bien en ciertos casos la calidad puede estar por debajo a la de algunos dispositivos relevados en el estado del arte, el prototipo da lugar mejoras en base a los trabajos a futuro planteados.

Palabras clave: visualización tridimensional, estereoscopía, computación gráfica, percepción visual.

ÍNDICE GENERAL

1.Introducción.....	1
Objetivos.....	2
Organización del documento.....	2
2.Percepción.....	3
Introducción.....	3
Proceso perceptivo.....	3
Estímulo.....	4
Electricidad.....	4
Experiencia y acción.....	5
Conocimiento.....	5
Introducción a la visión.....	6
Percepción de profundidad.....	7
Señales oculomotoras.....	8
Señales visuales monoculares.....	9
Señales visuales binoculares.....	13
3.Técnicas de representación tridimensional.....	17
Sistemas estereoscópicos basados en accesorios.....	17
Multiplexado por color.....	17
Multiplexado por polarizado.....	19
Multiplexado por tiempo.....	20
Dispositivos de realidad virtual.....	21
Sistemas autoestereoscópicos.....	22
Barrera de paralaje.....	23
Lenticulares.....	26
Dynallax.....	28
HR3d.....	29
Microsoft Wedge.....	32

Evaluación de técnicas.....	37
4.Solución planteada.....	39
Descripción general.....	39
Requerimientos.....	41
Barrera de paralaje dinámica.....	42
Solución matemática.....	42
Hardware.....	48
Software.....	55
Parallax Barrier Library.....	56
Aplicación.....	58
Problemas identificados.....	63
5.Conclusiones y trabajo futuro.....	65
Conclusiones.....	65
Trabajo futuro.....	66
6.Bibliografía.....	69

1. INTRODUCCIÓN

La visualización de imágenes tridimensionales tiene una gran variedad de aplicaciones, como ser dispositivos médicos, videojuegos, publicidad y proyección de películas cinematográficas.

En los últimos años la industria de los medios ha desarrollado un interés creciente en la visualización estereoscópica. La causa de esta atención particular no solamente proviene de los productores de los contenidos sino también de los proveedores de bienes de consumo. Así es que varias empresas han incursionado en la venta de televisores y dispositivos móviles tridimensionales.

Tanto en el cine como en la televisión, para lograr que el espectador pueda apreciar la profundidad de la imagen, es necesario generalmente que éste utilice gafas especiales. Estas gafas utilizan técnicas que, basadas en estereopsis, permiten a los usuarios visualizar imágenes tridimensionales. La estereopsis es un fenómeno que crea la ilusión de profundidad en base al despliegue de dos imágenes sincronizadas y ligeramente diferentes, una para cada ojo [17].

Alternativas de experiencias que no requieren la utilización de accesorios se han venido desarrollando tanto en la academia como la industria [32][46][57][61]. Así aparecen otras opciones como las pantallas lenticulares y las pantallas de paralaje, entre otras; descritas en detalle más adelante en este documento. Estas técnicas no requieren la utilización de aditamentos especiales, como ser gafas, por parte del espectador logrando una interacción más sencilla y natural.

Sin embargo, las técnicas que no requieren de aditamentos tienen cada una sus desventajas. Entre éstas se destacan altos costos de elaboración, funcionamiento desde posiciones predefinidas para el espectador, y baja calidad de imagen. Además, en la actualidad existe poco material y software apropiado reutilizable para el desarrollo de dispositivos que apliquen estas técnicas.

En este proyecto se busca estudiar las diferentes técnicas y crear un prototipo que no requiera la utilización de aditamentos por parte del usuario, aprovechando los avances que se han realizado en el campo de la detección de movimiento y los bajos costos actuales de los monitores con tecnología LCD.

OBJETIVOS

El proyecto consistió en relevar comparativamente las alternativas de visualización existentes, para elegir una y, adaptándola a las condiciones del Laboratorio de Medios del Instituto de Computación de la Facultad de Ingeniería [64] y del mercado local, construir un prototipo de aplicación que la utilice.

Los objetivos específicos fueron:

- Obtener un estado del arte sobre técnicas de visualización tridimensional haciendo énfasis en aquellas que no requieren aditamentos (como gafas).
- Implementar un prototipo de dispositivo de visualización tridimensional que no requiera aditamentos para el usuario junto con el desarrollo del software que permita su utilización.

ORGANIZACIÓN DEL DOCUMENTO

A continuación se detalla la organización del resto del documento.

En el capítulo 2 se presenta una introducción al funcionamiento de la percepción del ser humano haciendo énfasis en particular en la percepción de la profundidad.

En el capítulo 3 se describe el estado del arte de las técnicas de representación tridimensional.

El capítulo 4 presenta la solución planteada para el prototipo de visualización tridimensional. Dentro de este capítulo se detalla la lógica del algoritmo detrás de la solución, el diseño del hardware construido y el desarrollo del software del prototipo.

Finalmente, en el capítulo 5 se presentan las conclusiones del proyecto junto con las posibles mejoras y trabajos a futuro.

2. PERCEPCIÓN

INTRODUCCIÓN

Definir los requerimientos del hardware y las características de las imágenes es el primer paso para construir una pantalla tridimensional. Se necesita un claro entendimiento de cómo una imagen digital estereoscópica es percibida por el ser humano de manera de poder tomar decisiones correctas y con fundamento durante la etapa de diseño. Por ello es necesario hacer un repaso del funcionamiento de la percepción visual en el ser humano, haciendo mayor énfasis en la percepción de la profundidad.

PROCESO PERCEPTIVO

Goldstein [17] define a la percepción como el resultado final de varios procesos complejos. El proceso perceptivo es una secuencia de fases que trabajan en conjunto para determinar la experiencia y la reacción al estímulo del entorno. La Figura 1 muestra un diagrama del proceso dividido en cuatro categorías: estímulo, electricidad, experiencia y acción y, participando en varias etapas del proceso, conocimiento.

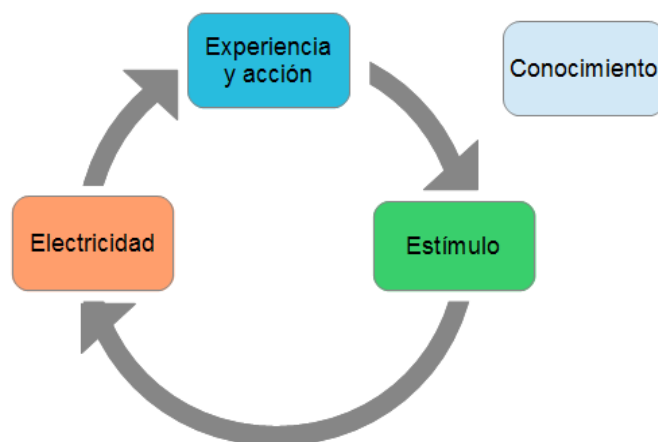


Figura 1. Diagrama del proceso de percepción. [17]

ESTÍMULO

El estímulo se refiere a lo que se encuentra en el entorno, a lo que se le presta atención y estimula los receptores.

Hay dos aspectos del estímulo que se encuentran en el entorno. El primero es la estimulación ambiental, la cual está formada por todas las cosas en el entorno que potencialmente pueden ser percibidas. La cantidad de potenciales estímulos en una escena es muy grande como para poder percibir todo a la vez. Así es que el ser humano concentra su atención en aquello que le resulta especialmente interesante de la escena, como ser un objeto en particular. Al centrar la atención en algún elemento de la escena estamos frente al segundo aspecto del estímulo en el entorno, el estímulo atendido.

Al mirar un objeto en particular directamente, se forma una imagen del mismo sobre los receptores del ojo. Esta etapa de estimulación de los receptores es importante porque el estímulo (el objeto foco de la atención) es transformado a otra forma (una imagen en el ojo del observador).

Dado que el objeto de atención ha sido transformado en una imagen, podemos describir esta imagen como una representación del objeto. La siguiente etapa del proceso de percepción lleva la idea de representación un paso más lejos al transformar la imagen en el ojo en electricidad.

ELECTRICIDAD

Uno de los principios centrales de la percepción es que todo lo que se percibe está basado en señales eléctricas en el sistema nervioso. Estas señales eléctricas son creadas en los receptores, los cuales transforman energía del entorno (como ser la luz que llega a la retina de los ojos) a señales eléctricas en el sistema nervioso. Este proceso de transformación es llamado transducción.

La señal eléctrica es procesada a medida que viaja a través de las capas de neuronas de la retina y eventualmente transmitida por las células ganglionares de la retina al cerebro. La etapa de transmisión es importante ya que si las señales no llegan al cerebro no hay percepción.

Debido al procesamiento neuronal ocurrido durante la transmisión, al llegar al cerebro la señal que representa al estímulo es muy diferente a la señal generada en los receptores. En el sistema nervioso, la representación eléctrica original del estímulo que es creada por los receptores es transformada por el procesamiento a una nueva representación del estímulo en el cerebro.

EXPERIENCIA Y ACCIÓN

En esta etapa del proceso perceptivo las actividades de transducción, transmisión y procesamiento son transformadas a cosas de las que podemos ser conscientes.

La percepción es una experiencia sensorial consciente que ocurre cuando la señal eléctrica producida por las neuronas que representa al estímulo es procesada por el cerebro produciendo la experiencia de ver el objeto de estímulo. Además de percibir, generalmente ocurren otros procesos como reconocer que el objeto que se está viendo es un árbol o acercarse para poder ver más de cerca el árbol. Estos dos pasos adicionales, reconocimiento y acción, son comportamientos que son resultados importantes del proceso perceptivo.

El reconocimiento es la habilidad de ubicar un objeto dentro de una categoría, como puede ser “árbol”, que le da significado. Aunque en un principio percepción y reconocimiento pueden parecer parte de un mismo proceso, estudios [52] han mostrado que son procesos separados.

La acción incluye las actividades motoras como mover la cabeza o los ojos y desplazarse en el entorno. El hecho de que la percepción comúnmente lleve a la acción (como acercarse al objeto de atención o girar la cabeza para mirarlo mejor), significa que la percepción es un proceso de cambio continuo.

Los cambios que ocurren a medida que las personas perciben es el motivo por el cual los elementos de la Figura 1 están ubicados de forma circular. El proceso de percepción se puede describir como una serie de pasos que “comienza” con un estímulo del entorno y “termina” con la percepción, reconocimiento y acción; sin embargo, el proceso en su totalidad es tan dinámico y cambiante que no puede definirse con un punto de comienzo y un punto final.

CONOCIMIENTO

El diagrama del proceso perceptivo de la Figura 1 también incluye un cuarto componente, el conocimiento. El conocimiento corresponde a cualquier información que sea utilizada por una persona durante el proceso perceptivo. Es ubicado por fuera del ciclo del proceso porque puede afectar a varios de los pasos del mismo.

La información utilizada por una persona en una situación puede ser de cosas aprendidas hace muchos años, como ser cuando se aprende como luce un árbol, o conocimiento obtenido de eventos que acaban de ocurrir.

INTRODUCCIÓN A LA VISIÓN

En esta sección se hará una introducción al funcionamiento del sistema visual. La visión comienza cuando la luz visible es reflejada desde los objetos del entorno hacia el ojo.

La luz reflejada por los objetos del entorno ingresa al ojo a través de la pupila y es enfocado por la córnea y el cristalino para formar imágenes nítidas de los objetos sobre la retina, la cual contiene los receptores para la visión (Figura 2).

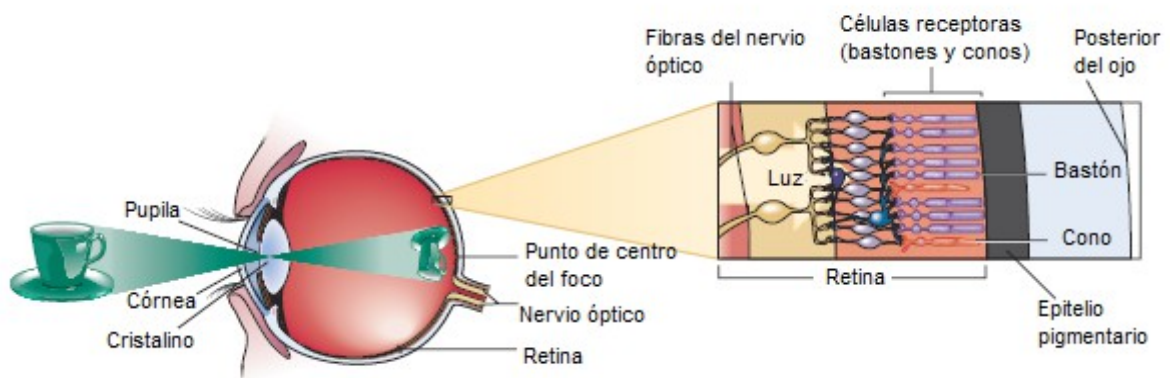


Figura 2. La imagen de la taza es enfocada en la retina, la cual se alinea con el lado posterior del ojo. La imagen aumentada a la derecha muestra los receptores y otras neuronas que forman la retina. [17]

Existen dos tipos de receptores visuales, bastones y conos, los cuales contienen componentes químicos sensibles a la luz llamados pigmentos visuales que reaccionan a la luz y disparan señales eléctricas. Estas señales viajan a través de la red de neuronas que forman la retina. La señal finalmente emerge del ojo por la parte trasera mediante el nervio óptico, que conduce la señal al cerebro.

Luego de reflejada la luz desde los objetos al ojo, necesita ser enfocada en la retina. La córnea, la capa transparente que cubre el frente del ojo, representa cerca del ochenta por ciento de la capacidad de foco del ojo. Sin embargo, la córnea se encuentra en un lugar fijo y no puede ajustar su foco. El cristalino, que provee el veinte por ciento restante de la capacidad de foco, puede cambiar su forma y ajustar el foco del ojo para estímulos ubicados a diferentes distancias.

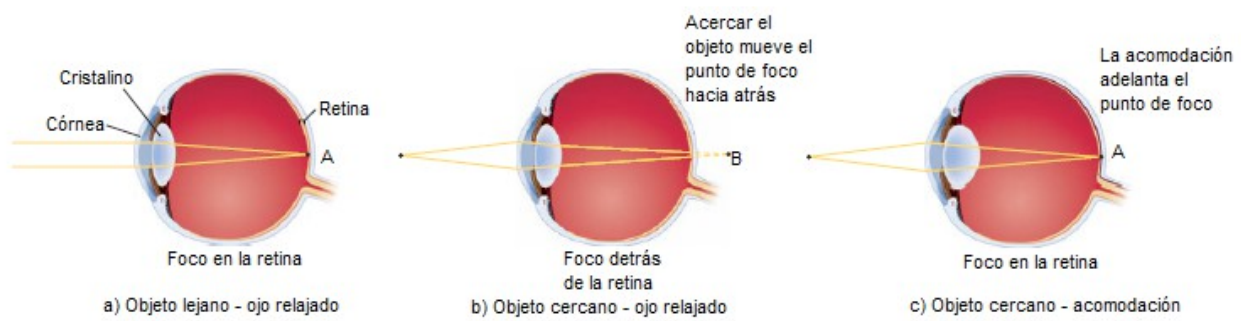


Figura 3. Enfoque de los rayos de luz en el ojo. a) Rayos provenientes de una fuente pequeña de luz a más de 6m. son aproximadamente paralelos. El punto de foco para rayos paralelos es en el punto A en la retina. b) Acercar un objeto al ojo relajado mueve el punto de foco detrás del ojo al punto B, pero la luz se detiene en la cara posterior del ojo. c) La acomodación (ajuste del cristalino) aumenta la capacidad de foco del cristalino y ajusta el punto de foco nuevamente al punto A. [17]

PERCEPCIÓN DE PROFUNDIDAD

La percepción de profundidad es la habilidad visual de percibir el mundo en tres dimensiones. Si pensamos en el ojo como una cámara en la cual la córnea y el cristalino se encargan de enfocar la luz sobre la retina, entonces se puede pensar a la excitación generada sobre los receptores de la retina en un instante como una imagen. Esta imagen si bien está curvada sobre la parte trasera de la cara interna del ojo, es esencialmente una imagen bidimensional [17]. La pregunta que surge entonces es, ¿cómo se obtiene la información de profundidad?

La teoría de las señales de profundidad es una de las principales teorías sobre la percepción de la profundidad [16]. Ésta establece que diferentes fuentes de información, o señales de profundidad, se combinan para dar al observador la disposición tridimensional de una escena [17].

Estas señales visuales se pueden clasificar en tres grupos según Goldstein [17], las señales obtenidas a partir de la habilidad de sentir la posición de nuestros ojos y la tensión de los músculos del ojo (oculomotoras), de una escena vista con un solo ojo (monoculares) y las señales obtenidas de una escena vista con dos ojos (binoculares).

SEÑALES OCULOMOTORAS

Los dos tipos de señales oculomotoras son los ángulos diferentes de inclinación de los ojos y los diferentes grados de tensión en los músculos del cristalino del ojo. Estas señales se conocen como convergencia y acomodación del cristalino respectivamente.

La convergencia es el movimiento de los ojos que ocurre al mirar a objetos cercanos, lo cual genera que la dirección de éstos converja.

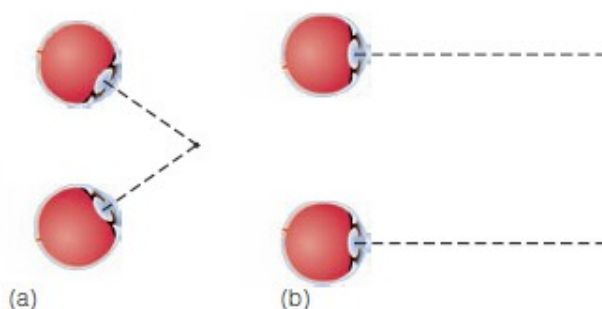


Figura 4. (a) Convergencia de los ojos al mirar a objetos cercanos. (b) Al mirar a objetos lejanos la dirección de los ojos es casi paralela hacia adelante. [17]

La acomodación es el cambio en la forma del cristalino del ojo que ocurre cuando se enfoca a un objeto a diferentes distancias.

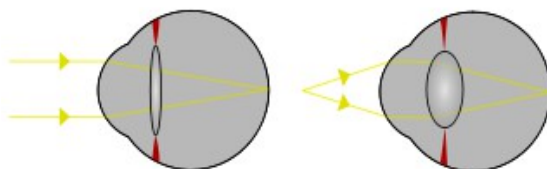


Figura 5. La luz proveniente de objetos a distintas distancias logra enfocarse en la retina por el cambio en la forma del cristalino. [17]

Cada una de estas dos señales oculomotoras solamente provee de un valor sobre la profundidad de una escena a la vez. Existen investigaciones [13] que muestran que estas señales son fuentes menores de información al momento de evaluar distancias respecto a otras.

SEÑALES VISUALES MONOCULARES

Las señales monoculares son aquellas que pueden ser percibidas por un solo ojo. Dentro de esta categoría se encuentran la acomodación, tratada en la sección de señales oculomotoras; las señales pictóricas, que proveen de información sobre la profundidad a partir de una imagen bidimensional; y las señales basadas en movimiento, que están basadas en información que se crea del movimiento de una escena.

Las señales pictóricas son fuentes de información de profundidad que se pueden obtener de una imagen, como ser fotografías o imágenes en la retina. A continuación se detallará cada una de las mismas.

Cuando un objeto oculta de manera total o parcial a otro de la vista, estamos frente al fenómeno de oclusión. El objeto oculto es visto como más lejano que aquel que lo cubre. La oclusión no provee de información absoluta sobre la distancia de un objeto, solamente indica la distancia relativa respecto a otros objetos presentes en la escena. Es decir, un objeto oculto se encuentra más lejano que otro objeto, pero simplemente por oclusión no se puede saber cuán mayor puede ser esa distancia.



Figura 6. La imagen de la izquierda contiene señales de oclusión que permiten comprender correctamente la escena. La imagen de la derecha contiene señales contradictorias de oclusión. [17]

La altura relativa es otra señal, objetos que estén por debajo del horizonte y tienen sus bases más altas en el campo de visión usualmente son vistos como más distantes. Esta señal se puede ver en las dos motos de la Figura 7. La parte inferior de la moto más lejana se encuentra más arriba que la parte inferior más cercana. Para objetos por arriba de la línea del horizonte aplica la mismo concepto pero de forma inversa.



Figura 7. [17]

Según las señales de tamaño relativo y tamaño esperado, cuando dos objetos son de igual tamaño, aquel que se encuentre más lejos va a ocupar menos espacio en el campo de visión que el que esté más cerca. Esta señal depende de cierta manera del conocimiento de una persona de los tamaños físicos de los objetos. En la Figura 7, por ejemplo, los dos postes de la línea telefónica son de igual tamaño, al igual que las motos. La Figura 8 muestra como el contexto, sugiriendo diferentes profundidades para los dos monstruos, hace que el monstruo en la parte superior derecha sea percibido de un tamaño mayor aunque los dibujos de ambos monstruos son iguales.

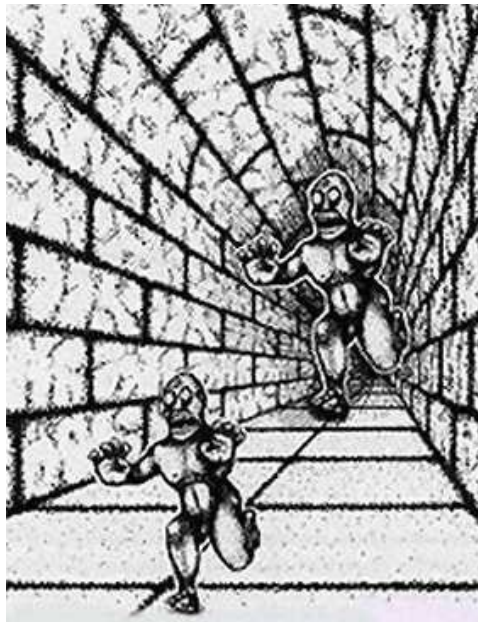


Figura 8. Ambos monstruos se perciben de diferente tamaño sin embargo son dibujos idénticos. [56]

Cuando líneas paralelas se extienden desde el observador, son percibidas como convergentes (se acercan entre sí) a medida que las distancias aumentan. El acercamiento de las líneas paralelas se conoce como convergencia de perspectiva y se puede apreciar en las líneas de las paredes y el piso de la Figura 8.

El tamaño familiar de los objetos se utiliza al juzgar distancias, este se basa en los conocimientos previamente adquiridos por las personas. Un experimento llevado a cabo por William Epstein [11] en 1964 muestra que bajo ciertas condiciones, el conocimiento del tamaño de objetos familiares influencia en la percepción de la distancia. El experimento consistió en fotos de igual tamaño de monedas de distinto tamaño, las cuales fueron posicionadas a una misma distancia del observador. Las fotografías fueron ubicadas en una habitación oscura donde eran iluminadas y donde los sujetos debían observarlas con un solo ojo, creando la ilusión de ser monedas reales. Al momento de evaluar las distancias de cada una de las fotografías, los sujetos estimaron que la foto de la moneda que físicamente era de menor tamaño era la más cercana y respectivamente que la fotografía de la moneda de mayor tamaño era la más lejana. El juicio de los observadores se vio influenciado por el conocimiento del tamaño de las monedas reales. Este resultado no ocurrió cuando los observadores vieron la escena con ambos ojos ya que en este caso otras señales aparecen para aportar información.

La perspectiva atmosférica ocurre cuando objetos más distantes parecen estar menos definidos y frecuentemente tienen un tinte azul (o gris). Cuanto más alejado este un objeto se debe ver a través de más aire y partículas (polvo, polución, etc.), haciendo que éstos se vean menos definidos y más azulados. En la Figura 9 se puede apreciar como las montañas del fondo de la escena tienen una tonalidad más azulada y son menos definidas comparadas con los objetos más cercanos en la escena.

Otra señal de profundidad es el gradiente de textura. El gradiente de textura ocurre al haber elementos espaciados de manera uniforme en una escena que parecen estar más juntos a medida que la distancia aumenta. La Figura 9 es un claro ejemplo de esta señal visual.



Figura 9. Ejemplo de gradiente de textura. También se puede apreciar la perspectiva atmosférica en las montañas en el horizonte. [17]

Las sombras asociadas con objetos pueden proveer información sobre la posición de los mismos. En la Figura 10 se puede apreciar claramente como las sombras resuelven la contradicción asociada con la oclusión, en este caso al dar información de que el florero no se encuentra reposado sobre una superficie sino que está elevado respecto de ésta. Las sombras también permiten percibir la tridimensionalidad de los objetos, en la Figura 10 permiten percibir la curvatura de las copas y el florero.



Figura 10. Izquierda: Oclusión. Medio: Oclusión contradictoria. Derecha: Las sombras resuelven la contradicción. [17]

Las señales que han sido mencionadas funcionan si el observador se encuentra quieto o en movimiento. Además, en el caso de que el observador se encuentre en movimiento nuevas señales surgen que proporcionan aún más información sobre la profundidad. Se describirán dos tipos de señales asociadas con el movimiento: la señal de paralaje de movimiento y la señal de supresión y acrecentamiento.

El paralaje de movimiento ocurre cuando, al moverse el observador, los objetos cercanos aparentan deslizarse rápidamente frente a él, pero los más distantes aparentan moverse más lentamente. Este efecto es claramente notorio al mirar por las ventanas laterales de un auto en movimiento. Los objetos próximos se aceleran al grado de volverse borrosos, mientras que el movimiento de los que están en el horizonte apenas se advierte. El paralaje

de movimiento es una de las fuentes de información más importantes para muchos animales.

Al moverse el observador de forma lateral, algunos objetos en la escena son cubiertos y otros descubiertos. La señal de eliminación y acreción está relacionada tanto al paralaje por movimiento como a la oclusión y ocurre cuando objetos superpuestos en la escena parecen moverse de forma relativa entre sí al desplazarse el observador. Cuando el observador se desplaza en una dirección, el objeto posterior es cubierto, o suprimido, por el anterior. Cuando se mueve en dirección contraria, el objeto posterior es descubierto, o acrecentado.

SEÑALES VISUALES BINOCULARES

Debido a que los ojos en las personas se encuentran separados, éstos tienen puntos de vistas distintos. Es la diferencia entre ambos puntos de vista la que crea la disparidad binocular.

La disparidad binocular se define como la diferencia entre las imágenes en los ojos izquierdo y derecho. Para poder explicar cómo funciona la disparidad hay que introducir el concepto de puntos correspondientes de la retina, éstos son los puntos en el espacio cuyas proyecciones en las retinas tienen la misma ubicación. Si se pudieran juntar ambas retinas, las proyecciones de estos puntos se superpondrían. El lugar geométrico definido por los puntos correspondientes de la retina se denomina horóptera y ésta forma una superficie curva. En teoría, una horóptera debería de estar formada por los puntos de una curva circular en el plano horizontal y los puntos de una línea vertical, sin embargo se ha demostrado que la horóptera empírica está formada por una superficie en el espacio más compleja [53].

Esta diferencia entre la horóptera teórica y la empírica se debe a factores neuronales y ópticos. La cantidad de fotorreceptores por unidad de área en la retina no es homogénea, la mitad de la retina más cercana a la nariz tiene mayor densidad de fotorreceptores, lo que genera una desviación en el mapeo de la horóptera en la corteza visual.

La Figura 11 muestra cómo si un observador mira a un punto F, existe una horóptera en la cual se encuentran los puntos que tienen proyecciones correspondientes en las retinas (en este caso los puntos F, S y H son ejemplos de puntos en la horóptera).

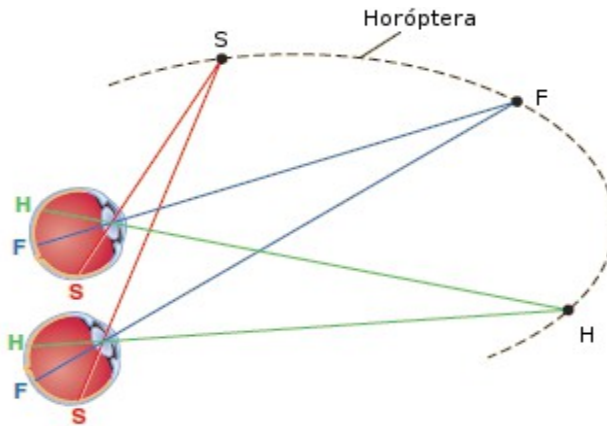


Figura 11. Los puntos en una horóptera se proyectan en puntos correspondientes en las retinas. [17]

En la Figura 12 del lado izquierdo, un observador mira un punto F y en este caso la proyección de F en las retinas se corresponden pero para el punto C, que no pertenece a la horóptera, la proyección en la retina es sobre puntos no correspondientes. La proyección en el ojo izquierdo del punto C es el punto C_L y en el ojo derecho es el punto C_R . Del lado derecho de la Figura 12, el observador fija la vista en el punto C haciendo que la proyección de C sea sobre puntos correspondientes en las retinas, en este caso las proyecciones de F pasan a ser en puntos no correspondientes F_L y F_R .

Analizando la imagen de la izquierda de la Figura 12, la diferencia entre la proyección de C en el ojo derecho (C_R) y el punto correspondiente a C_L en el ojo derecho es el ángulo de disparidad, también llamada disparidad absoluta para C el fijar la mirada en el punto F. La disparidad absoluta es importante porque provee de información sobre la distancia de los objetos. Esta determina que tan distante está un objeto de la horóptera. Mayor disparidad se asocia con una mayor distancia a la horóptera.

Una propiedad de la disparidad absoluta es que cambia cada vez que el observador cambia el punto en donde fija la mirada. Sin embargo, existe información que no cambia sin importar donde fije la mirada el observador. Este dato es llamado disparidad relativa, que se define como la diferencia entre la disparidad absoluta entre dos objetos. Esto se puede apreciar al comparar la disparidad relativa entre los puntos C y F en ambas imágenes de la Figura 12.

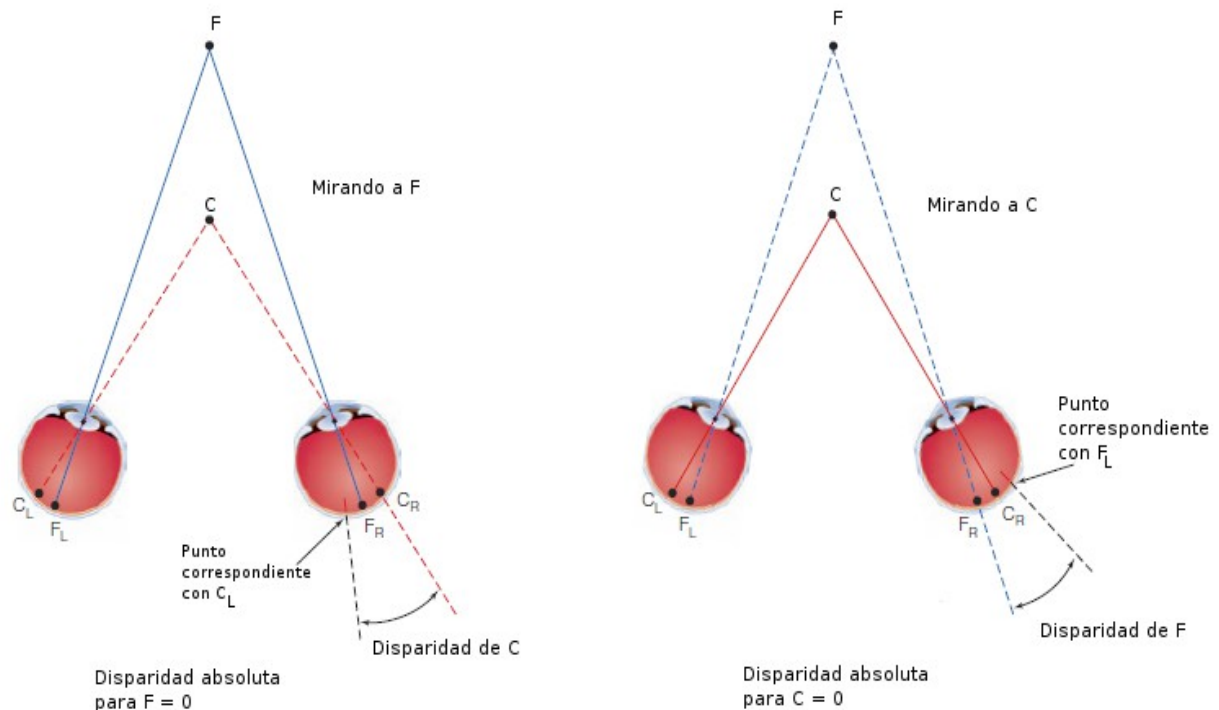


Figura 12. [17]

Tanto la disparidad absoluta como la relativa contienen información sobre la distancia de un objeto al observador. No obstante, la descripción de disparidad brindada es simplemente geométrica y no hace mención de la percepción. Para poder relacionar la disparidad con lo que percibe un observador hay que introducir la estereopsis.

La estereopsis es el fenómeno dentro de la percepción visual, en el que a partir de dos imágenes ligeramente diferentes proyectadas en la retina de cada ojo, el cerebro es capaz de recomponer una imagen tridimensional. Para demostrar que la disparidad por sí sola puede resultar en percepción de profundidad, Bela Julesz [28] creó un estímulo llamado estereograma de puntos aleatorios el cual no contenía ninguna otra señal pictórica. Este estereograma se construyó generando un patrón de puntos aleatorio para una imagen. Luego para la otra imagen, un conjunto de puntos de la primera imagen es desplazado una o varias unidades para generar la disparidad. La Figura 13 es un estereograma de puntos aleatorios, cruzando la vista para lograr que ambas imágenes coincidan se logra ver un cuadrado ubicado al centro por delante del resto de la imagen.

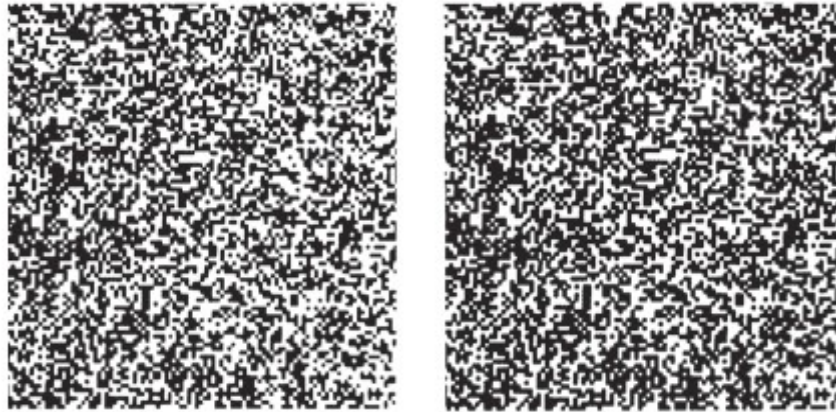


Figura 13. Estereograma de puntos aleatorios. [17]

Resumiendo, la visión humana usa una gran variedad de fuentes (señales visuales) para obtener información de profundidad. Cada señal individualmente puede llegar a ser ambigua pero al combinar todas las señales disponibles y aplicando cierto número de reglas y restricciones (las superficies son localmente suaves, los objetos tienden a ser rígidos, la luz apunta desde arriba), el sistema visual usualmente puede llegar a una solución estable e inequívoca.

3. TÉCNICAS DE REPRESENTACIÓN TRIDIMENSIONAL

Las señales monoculares se basan en propiedades de imágenes bidimensionales, generar imágenes con estas cualidades realistas es un problema medianamente sencillo y pueden ser obtenidas mediante diferentes medios como cámaras o gráficos generados por computadora. Sin embargo, generar disparidad binocular, una de las fuentes más importantes de información de profundidad para la percepción del ser humano, a través de una pantalla o dispositivo similar es una tarea de alta complejidad.

Las técnicas para hacer llegar una imagen distinta a cada ojo, serán agrupadas en: sistemas estereoscópicos basados en accesorios y sistemas autoestereoscópicos.

SISTEMAS ESTEREOSCÓPICOS BASADOS EN ACCESORIOS

Esta sección examina el estado actual de las técnicas estereoscópicas para dispositivos tridimensionales que requieran que el observador utilice algún tipo de accesorio especial (como ser gafas) para poder percibir las tres dimensiones.

MULTIPLEXADO POR COLOR

El ejemplo más conocido de un método de visualización multiplexado por color son las gafas anaglifo. Las gafas anaglifo son gafas que filtran la luz para cada ojo por diferentes longitudes de onda, es decir tienen un filtro de color diferente para cada ojo. Esta es una solución

sencilla al problema de visualización tridimensional y puede ser aplicada junto a cualquier equipamiento de video a color.

El método anaglifo más común utiliza el canal rojo para el ojo izquierdo y el canal cian para el ojo derecho. El observador utiliza gafas coloreadas según el canal del color correspondiente a cada ojo de manera que llegue a cada uno la imagen correcta.

Las mayores desventajas de este tipo de dispositivos son la pérdida de información de color y un alto grado de interferencia entre las imágenes derecha e izquierda. La respuesta del espectro de color del dispositivo (monitor, fotografía, proyector, etc.) y de las gafas, métodos de compresión de imagen y protocolos de transmisión de las imágenes han sido citados como fuentes de la interferencia [63]. Las imágenes anaglifas son creadas usualmente capturando las imágenes con una cámara binocular o usando información de profundidad de los objetos y solamente una imagen [35]. Para disminuir la interferencia y aumentar la calidad de las imágenes, métodos como alineado de imágenes, difuminación de componentes de color y ajustes del mapa de profundidad han demostrado lograr mejoras significativas [21].

Existen actualmente controladores de software de NVIDIA [42] e iZ3D [25] que permiten generar las imágenes en este formato.

Comercialmente existen soluciones que utilizan esta técnica. ColorCode 3-D [7] permite ver imágenes a color utilizando gafas con los colores azul y ámbar. La información tridimensional es representada por pequeñas variaciones de color en la imagen, la cual a simple vista tiene rastros azulados y amarillentos. Las principales ventajas que presenta es que es compatible con cualquier pantalla estándar y puede ofrecer imágenes a color a muy bajo costo. Por otro lado, esta tecnología no ha superado completamente los problemas de representación de los colores e interferencia entre las imágenes correspondientes a cada ojo.

La empresa alemana Infitec GmbH [22] desarrolló una técnica que entra dentro de esta categoría pero logra imágenes de mayor calidad. El dispositivo desarrollado se basa en filtros de interferencia para lograr separar las señales para cada ojo. Dos proyectores a color son adaptados con filtros de interferencia, estos filtros permiten el pasaje de la luz en tres diferentes longitudes de onda dentro del espectro visible de los colores azul, verde y rojo. Además la longitud de onda que le corresponde a cada proyector por cada color es levemente distinta del otro. Las imágenes de ambos proyectores son proyectadas en una misma pantalla a través de los filtros y el observador utilizando gafas con los mismos filtros puede separar la imagen a color en dos imágenes a color, una para cada ojo. La interferencia de las imágenes entre sí al ser demultiplexada por las gafas es menor al 1% en todo el espectro visible [27]. La Figura 14 muestra un esquema del funcionamiento del dispositivo.

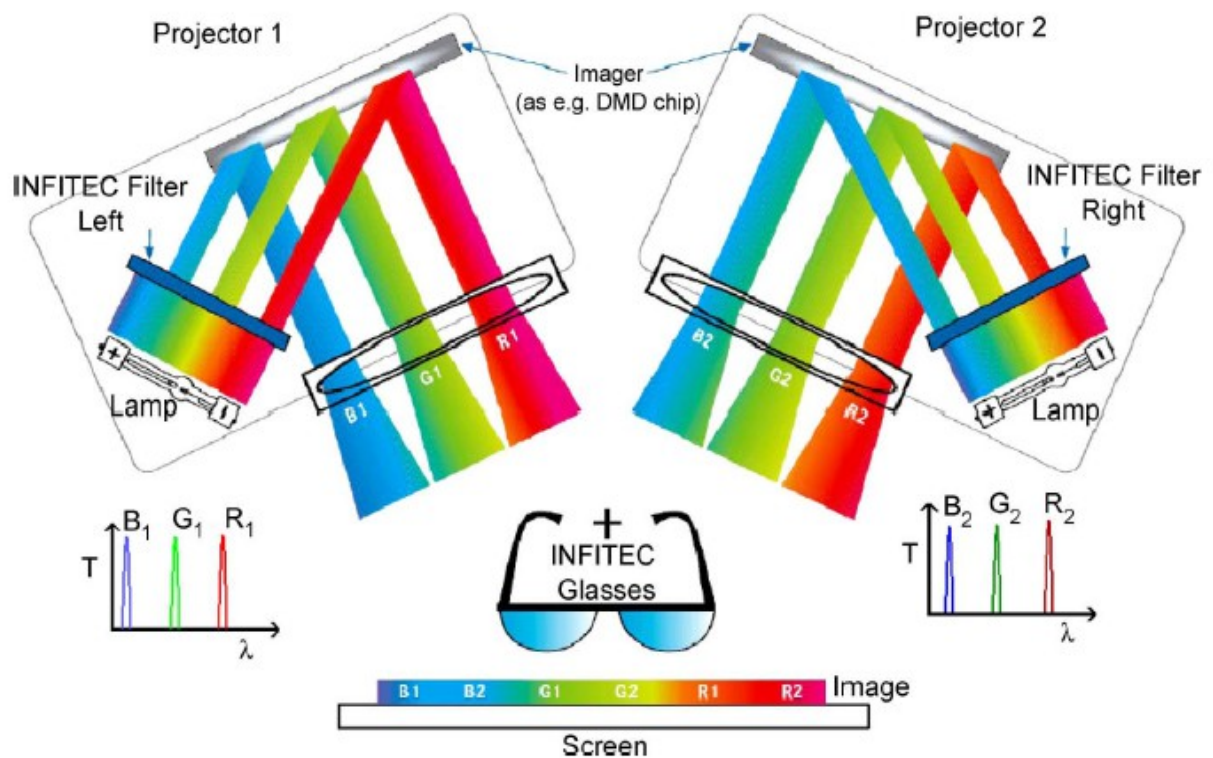


Figura 14. Diagrama del sistema con gafas con filtrado de luz por longitud de onda. El espectro de luz se separa en dos partes (B1, G1 y R1 corresponde a un ojo y B2, G2 y R2 corresponden al otro). [27]

MULTIPLEXADO POR POLARIZADO

En este método, los estados de polarización de la luz correspondiente a las imágenes destinadas a cada ojo son ortogonales entre sí. El observador utiliza gafas con filtros polarizados apropiados para bloquear la imagen que no corresponde a cada ojo. Las imágenes son producidas por dos proyectores donde cada uno proyecta la imagen de cada ojo y con estados de polarización de la luz ortogonales. Este tipo de proyección requiere de una pantalla especial que conserve la polarización de la luz. Comercialmente existen pantallas para proyecciones de frente como 3D Virtual Grey y pantallas para proyecciones desde atrás como 3D Virtual Black, ambas de la empresa Da-Lite [8], que aseguran mantener el 99% de la luz polarizada.

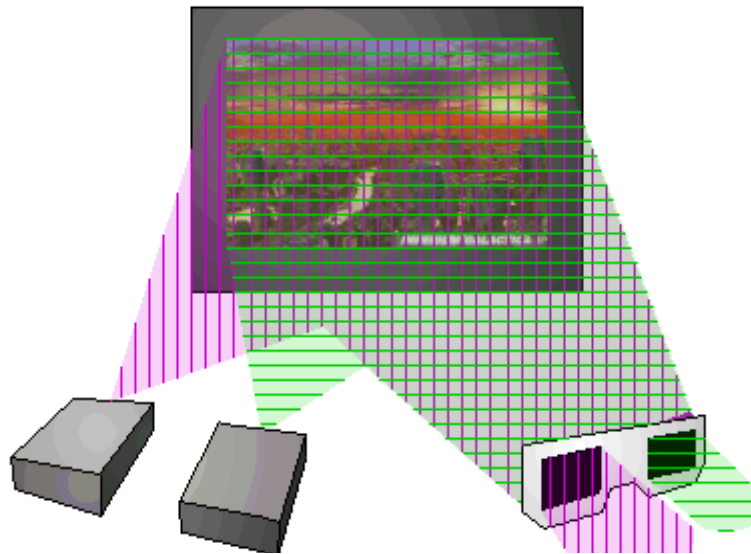


Figura 15. Los proyectores emiten imágenes con luz polarizada (ortogonales entre sí) que es reflejada en la pantalla, luego los lentes filtran la luz polarizada utilizando un ángulo diferente de polarización para cada ojo.

Otro método similar que reduce la complejidad de utilizar dos proyectores es usar un solo proyector equipado con una lente polarizada de cristal líquido capaz de polarizar de diferente modo la luz que pasa a través. La polarización del lente cambia de forma sincronizada con las imágenes que se proyectan alternando la imagen para cada ojo. La empresa RealD [50] ofrece este sistema de forma comercial para salas de cine. Este método tiene como desventaja una disminución del brillo total de la imagen, el filtro de polarizado frente al proyector bloquea la mitad de la luz, y luego se separa nuevamente a la mitad para cada ojo, resultando en solo 25% de la luz proyectada llegando al ojo.

Aunque los filtros de polarización pueden causar errores cromáticos, este tipo de dispositivo de visión tridimensional ofrece altas resoluciones y los problemas de calidad de los colores son generalmente insignificantes comparados con sistemas anaglifos.

MULTIPLEXADO POR TIEMPO

Estas pantallas aprovechan la persistencia de la visión del sistema visual humano para generar la percepción tridimensional. En esta técnica, las imágenes del ojo izquierdo y las del derecho se visualizan en la pantalla en forma alternada a altas velocidades, usualmente de 120 Hz (60 imágenes por segundo para cada ojo). Esto es conocido como secuencia

alternativa de cuadros y el observador debe utilizar gafas activas a batería con obturador, las cuales están sincronizadas con el contenido que se visualiza. Cuando en la pantalla se está mostrando la imagen para el ojo derecho las gafas bloquean la visión del ojo izquierdo y en el instante siguiente al mostrarse la imagen para el ojo izquierdo las gafas bloquean el derecho.

La sincronización entre las gafas y la pantalla se logra utilizando un emisor de luz infrarroja o con el enlace DLP (Digital Light Processor), el cual usa parpadeos codificados de luz blanca detectados por las gafas con obturador entre los cuadros izquierdo y derecho.

El proyector Sharp de una sola lente modelo PG-D45X3D [55], utiliza la tecnología de enlace DLP para su prestación tridimensional. La ventaja de esta técnica para proyecciones es que no requiere de una pantalla especial, solo con un proyector que pueda desplegar imágenes a altas frecuencias es suficiente.

Una desventaja inmediata es el costo de las gafas con obturador y el excesivo ancho de banda de video requerido en comparación con el requerido para los bidimensionales comunes.

Las pantallas tridimensionales multiplexadas por tiempo son las más comunes en el mercado con Samsung, Panasonic y Sony liderando con sus televisores.

Para monitores de escritorio existen una multitud de opciones que emplean diferentes técnicas explicadas anteriormente. También en este caso, la elección es entre las pantallas con gafas con obturador o las polarizadas. En el mercado de monitores NVIDIA ofrece su sistema 3D Vision [42], el cual incluye gafas que funcionan por obturación y un controlador de video especial para generar las imágenes. Este es de los modelos más comúnmente utilizados para los videojuegos.

DISPOSITIVOS DE REALIDAD VIRTUAL

El desarrollo de sistemas que proveen al observador con una sensación de inmersión es un tema que puede desarrollarse como un estudio aparte. En esta sección se discutirán brevemente las técnicas y progreso en el área.

Los dispositivos de realidad virtual son sistemas binoculares estereoscópicos con dos generadores de imagen separados. Generalmente utilizan pantallas de tipo LCD para desplegar la imagen y sistemas ópticos para permitir enfocar la imagen correctamente. Estos son portátiles y son montados en la cabeza del observador, de esta manera los movimientos

no limitan el efecto de tridimensionalidad y permiten dar una sensación de inmersión en las escenas desplegadas. Los primeros avances en este campo fueron dados por Morton Heilig en la década de 1950 al trabajar sobre el Sensorama Simulator [18]. Hoy en día dispositivos de este estilo son mucho más avanzados y tienen una gama de usos variados como entrenamiento militar, medicina y entretenimiento, entre otros.

Sistemas de realidad aumentada utilizan dispositivos de realidad virtual para mejorar el entendimiento del observador del mundo real aportando información mediante gráficos generados por computadora superpuestos sobre la escena del mundo real [1]. Uno de los avances más destacables es en la medicina, donde los cirujanos pueden tener información detallada generada por computadora superpuesta con la imagen de los pacientes en tiempo real. Esta técnica es conocida como cirugía asistida por computadora (CAS por sus siglas en inglés, *Computer-assisted surgery*).

Al igual que otras tecnologías de visión tridimensional los dispositivos de realidad virtual sufren del problema de discrepancia entre la información percibida por la convergencia de los ojos y la acomodación, este problema tiende a desaparecer a medida que el plano de la imagen se mueve hacia el infinito. Además, otro desafío en el desarrollo de este tipo de sistema tridimensional está asociado con el malestar causado por imágenes con diferencias. Las diferencias pueden ser causadas por defectos en los sistemas ópticos correspondientes al canal de cada ojo, por ruido en la señal o problemas de alineación entre las dos pantallas.

Existen varios productos comerciales de realidad virtual disponibles de diferentes compañías como Sony, SiliconMicroDisplay, i-O Display Systems o Vuzix Corporation.

SISTEMAS AUTOESTEREOSCÓPICOS

Los sistemas autoestereoscópicos crean imágenes con la disparidad requerida sin ningún tipo de gafas especiales u otro tipo de aditamento. Éstos pueden ser de dos vistas o multi-vistas. En los dispositivos de dos vistas sólo un par de imágenes estereoscópicas es mostrado, mientras que en los multi-vistas varias parejas de imágenes estereoscópicas son mostradas permitiendo varios observadores simultáneamente a diferentes ángulos. Además las zonas de observación pueden ser fijas o flexibles (utilizando algún método de seguimiento de la posición de los observadores).

Las pantallas autoestereoscópicas utilizan combinaciones de pantallas lenticulares, barreras de paralaje, espejos, micropolarizadores, entre otros para controlar la dirección de la luz emitida. En aquellos sistemas con zonas de observación fijas el observador debe posicionar

sus ojos en un lugar predeterminado para poder percibir las dos imágenes estereoscópicas, en cambio en los sistemas con seguimiento de la cabeza o los ojos el observador puede ver las imágenes con cierta libertad de movimiento. Por lo general los sistemas con seguimiento del usuario solamente permiten un observador.

A continuación se detallaran las diferentes técnicas más utilizadas y los avances realizados en este campo.

BARRERA DE PARALAJE

La barrera de paralaje fue la primera técnica para pantallas autoestereoscópicas. El principio de la barrera de paralaje fue inventado independientemente por Auguste Berthier, quien publicó primero pero no logró ningún resultado práctico [2], y por Frederic Ives, quien hizo y exhibió la primera imagen autoestereoscópica en 1901.

Generalmente las pantallas tienen píxeles que irradian luz de forma difusa, esto es irradian luz de igual manera en todas direcciones. Para poder crear una pantalla autoestereoscópica capaz de generar una vista para cada ojo, la luz emitida por una mitad de los píxeles debe ser dirigida en direcciones que sean vistas por el ojo izquierdo y la otra mitad de los píxeles debe ser dirigida en direcciones que sean vistas por el ojo derecho. Una barrera de paralaje es una solución simple a este problema. Las barreras de paralaje están formadas por un conjunto de líneas verticales negras que funcionan como una máscara sobre la pantalla que contiene los píxeles bloqueando la luz.

La Figura 16 muestra un esquema conceptual del funcionamiento. Las imágenes derecha e izquierda están entrelazadas en columnas en la pantalla y la barrera de paralaje se posiciona de tal manera que píxeles derechos e izquierdos son bloqueados de la vista excepto en las regiones correspondientes a la visión del ojo derecho e izquierdo respectivamente. Pantallas LCD son utilizadas comúnmente ya que ofrecen buenas resoluciones, un grosor controlado del cristal y hoy en día son accesibles económicamente.

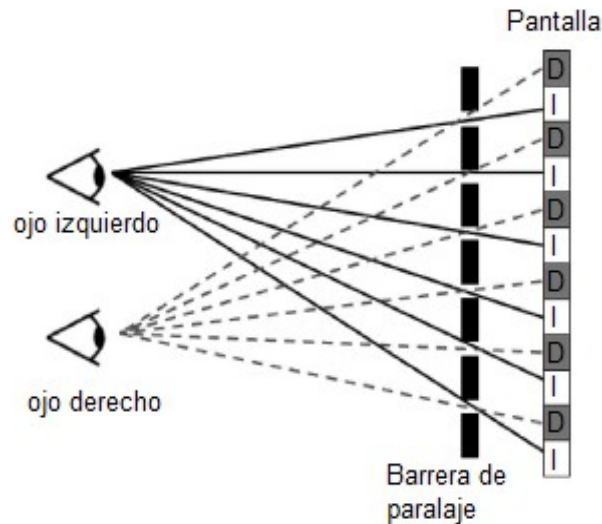


Figura 16. Diagrama de una barrera de paralaje.

Cuanto más cercana esté la barrera de la pantalla mayor será el ángulo de separación de las direcciones de la luz de las imágenes derecha e izquierdas. Esto se traduce en que cuanto más cerca esté la barrera de la pantalla, la distancia óptima de observación de la imagen será menor. A su vez la distancia óptima de observación es inversamente proporcional al tamaño de los píxeles de la pantalla. A mayor resolución de la pantalla, mayor es la distancia de observación óptima [19].

Los principales problemas asociados con los sistemas de barreras de paralajes son la pérdida de brillo y la baja resolución. La barrera al cubrir la luz emitida por la mitad de los píxeles causa un notorio deterioro de la luminosidad de la imagen y disminuye la resolución horizontal en el mejor de los casos a la mitad de la resolución horizontal de la pantalla.

Una solución al problema de la resolución se basa en la persistencia de la visión del ser humano. En un primer cuadro cada ojo ve la mitad de los píxeles y la otra mitad está cubierta por la barrera, en un segundo cuadro un instante después la barrera es modificada junto con los píxeles de la pantalla haciendo que los píxeles cubiertos se vean y los descubiertos sean cubiertos por la barrera. Esta técnica aplicada en una pantalla con alta frecuencia de refresco es capaz de lograr que el observador perciba la resolución horizontal completa [12].

Un problema práctico que se presenta al utilizar sistemas de barreras de paralaje es cómo encontrar la posición óptima de observación. Las barreras de paralaje no producen un solo punto de observación para el ojo izquierdo y para el derecho, estos puntos de observación se repiten alternadamente a una distancia correcta de la barrera. Esto puede causar que un observador vea una imagen estereoscópica invertida, al recibir en su ojo izquierdo la imagen derecha y en el ojo derecho la imagen izquierda. Este efecto causa problemas de falsa

profundidad y puede ser difícil para un observador no acostumbrado a esto darse cuenta de lo que está ocurriendo y si están viendo una imagen tridimensional correcta. Varios dispositivos han solucionado este problema mediante el uso de un indicador de la posición. El indicador se puede ubicar en las filas inferiores de píxeles de la pantalla y la barrera de paralaje es modificada en estas filas de modo que el indicador se ve en negro solamente cuando el observador se encuentra en la posición ideal.

La barrera está formada por líneas verticales que cubren los píxeles de la pantalla. Al estar los ojos alineados horizontalmente el diseño de la barrera es suficiente para hacer llegar imágenes distintas a cada ojo. Sin embargo, este diseño no logra reproducir el paralaje vertical. Es decir, si un observador mueve la cabeza en el eje vertical no cambiará la imagen que ve de manera acorde. Incluso inclinando la cabeza el observador puede dejar de percibir las imágenes correctamente. La solución a estos problemas es que la barrera sea tanto horizontal como vertical, creando una grilla. Esta solución a su vez disminuye la resolución vertical y aún más el brillo de la pantalla, por lo tanto es un punto a evaluar según los requerimientos al momento de diseñar una barrera de paralaje.

Si la barrera de paralaje puede prenderse o apagarse electrónicamente, entonces el sistema puede utilizarse tanto para las imágenes clásicas bidimensionales como para tridimensionales. Un punto interesante visto desde la interacción con el usuario, es poder combinar las imágenes bidimensionales con las tridimensionales al mismo tiempo. Para esto es necesario que la barrera pueda encenderse/apagarse por secciones. En 2001 Sharp produjo una pantalla que podía cambiar de modalidad [26]. En este sistema la barrera era formada por un componente que podía rotar la polarización de la luz (modo tridimensional) o dejarla pasar sin cambios de polarización (modo bidimensional). Avances en la tecnología OLED han permitido a Samsung anunciar un nuevo prototipo capaz que cambiar de modalidades de resolución completa [30].

Las barreras de paralaje no se utilizan como tecnología para televisores o monitores de computadora debido a las limitaciones respecto al punto de observación óptimo. Sin embargo, sí se utiliza para otros dispositivos móviles. Nintendo aprovechó esta técnica para su consola de juegos portátil Nintendo 3DS [41], LG también la ha incluido en sus teléfonos celulares Optimus 3D [33] y Thrill [34], HTC en el teléfono celular EVO 3D [20] (discontinuado) y Sharp con el modelo Aquos [54].

El origen de las imágenes tridimensionales con hojas lenticulares surge como una evolución de las barreras de paralaje. En 1912 Walter Hess patentó las hojas formadas por lentes cilíndricos evitando las imágenes oscuras como las que se obtenían al utilizar barreras de paralaje sobre fotografías impresas.

Los sistemas lenticulares de hoy en día combinan lentes cilíndricos con pantallas LCD para dirigir la luz emitida por los píxeles de tal manera de que solo sea vista en cierto ángulo frente a la pantalla. La pantalla crea zonas de visión para el ojo izquierdo y derecho repitiéndose alternadamente a cierta distancia en dirección paralela a la pantalla. La Figura 17 muestra un esquema del funcionamiento de esta técnica.

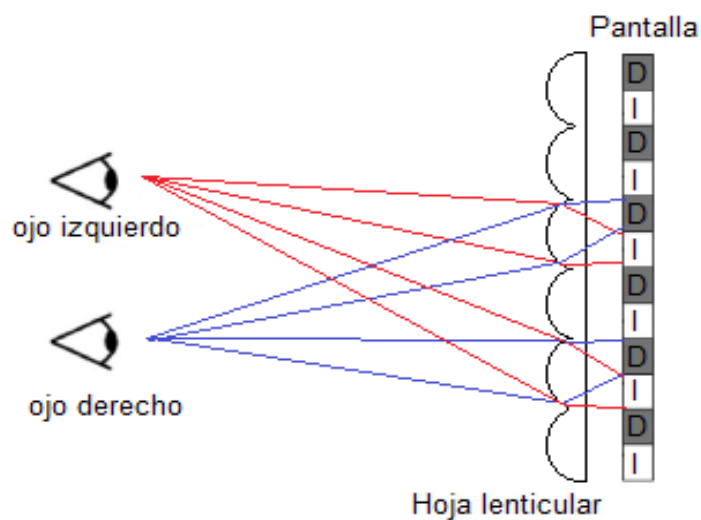


Figura 17. Diagrama de una pantalla lenticular.

La distancia óptima de observación es fija y es proporcional al grosor de la hoja lenticular (longitud focal) e inversamente proporcional al tamaño de los píxeles por cada lente cilíndrico.

La correcta alineación de la hoja lenticular sobre la pantalla LCD es fundamental en los sistemas lenticulares. Esta alineación se hace más difícil a medida que la resolución de la pantalla aumenta y cualquier desfase puede ser causa de distorsiones en las imágenes.

Otro problema que presentan estos sistemas es la variación de la intensidad de la imagen en ciertos ángulos. Los píxeles están compuestos por subpíxeles, uno correspondiente a cada canal de color, por lo general rojo, verde y azul, y en ciertos ángulos el lente puede desviar la luz de un subpíxel en particular a un ojo causando un aumento de la intensidad de la luz y

del color correspondiente al subpíxel. Como los subpíxeles están alineados verticalmente por color el efecto se percibe en líneas verticales en la pantalla. También debido a que entre subpíxeles existe un espacio que no emite luz puede ocurrir el caso contrario. La luz de ningún subpíxel en particular es dirigida al ojo sino que el lente enfoca al espacio entre ellos causando una disminución de la intensidad de la luz. Una solución a este problema fue propuesta por la empresa Philips [62], ésta consiste en inclinar levemente la dirección de los lentes cilíndricos de la hoja lenticular para que no sean verticales. Se logra una transición más homogénea al cambiar el ángulo de visualización y se evitan los cambios de intensidad en la imagen.

Las pantallas lenticulares pueden ser capaces de cambiar de modalidad entre bidimensional y tridimensional si la hoja lenticular está formada por un material especial capaz que alternar entre dos estados refractantes. Se han hecho pruebas con hojas lenticulares formadas por cristal líquido [9] que pueden alternar electrónicamente entre refractante (modo tridimensional) y no refractante (modo bidimensional). La desventaja de esta pantalla es que con ángulos de visualización oblicuos, en modo bidimensional, se produce un efecto de lente residual.

Las pantallas lenticulares con lentes cilíndricos ofrecen solamente paralaje horizontal. Otra técnica similar se basa en lentes esféricos en lugar de cilíndricos, con esto se logra tanto paralaje vertical como horizontal. Son menos comunes que los cilíndricos principalmente porque requieren sacrificar aún más la resolución de la imagen. La Figura 18 muestra un esquema del funcionamiento.

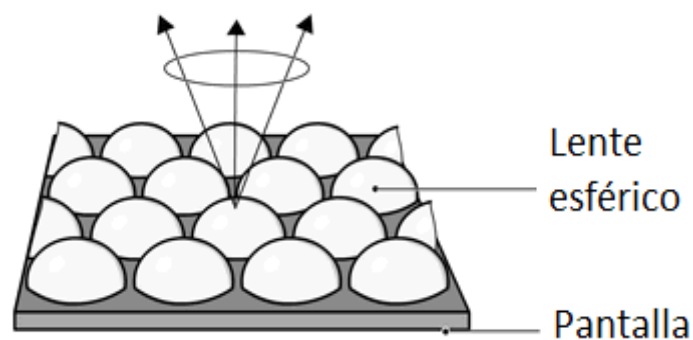


Figura 18. Diagrama de pantalla con lentes esféricos.

Philips produjo una línea de monitores tridimensionales con esta técnica hasta 2009. La empresa SpatialView [57] ofrece un producto basado en hojas lenticulares y un software propietario de la empresa para transformar pantallas de computadoras y teléfonos celulares en pantallas autoestereoscópicas.

Una solución para eliminar la distancia óptima de visión fija y zonas de visualización fijas se puede obtener con una barrera de paralaje dinámica. La Universidad de Illinois de Chicago, a través del Laboratorio de visualización electrónica, ha desarrollado esta solución con el proyecto Dynallax (*Solid State Dynamic Parallax Barrier Autostereoscopic VR Display*) [44] [47]. Dynallax permite hasta dos observadores con perspectivas propias independientes y puede combinar imágenes en dos y tres dimensiones.

La idea de utilizar una barrera de paralaje dinámica no es única a Dynallax, existieron otros proyectos anteriores que han tomado enfoques diferentes en cuanto a cómo construir la barrera. Barreras dinámicas mecánicas han aparecido en la literatura de sistemas autoestereoscópicos. Por ejemplo, Moseley *et al.* [39] propusieron una combinación de barreras electro-mecánicas para producir resultados mecánicos en respuesta al movimiento de la cabeza del observador. Barreras dinámicas de estado sólido también han sido construidas pero la arquitectura de la barrera difiere del propuesto por Dynallax. Un ejemplo es el Cambridge Display [10], el cual multiplexa por tiempo un número de imágenes a ciertas posiciones predeterminadas.

La construcción de la barrera de Dynallax está basada en dos capas de cristal líquido, una delante de la otra y utilizando la misma fuente de luz. En particular, utiliza una pantalla con estas características ofrecida por la empresa PureDepth [49].

Además, Dynallax utiliza el sistema de seguimiento Intersense 900 [23] con dos sensores, para dos observadores. Cada observador utiliza un sensor, sostenido con una banda, sobre la frente.

Una perspectiva independiente y en primera persona es dirigida a cada observador (máximo dos). La interactividad en tiempo real es lograda mediante el uso de un control que contiene un joystick y múltiples botones. En el caso de dos observadores, el control es compartido o utilizado por un observador a la vez.

Rápidos movimientos del observador pueden sobrepasar la velocidad de respuesta de una barrera de paralaje, resultando en una incorrecta visión de la imagen. Para evitar este problema, Dynallax corrige rápidamente la calidad estereoscópica pero no la calidad de la perspectiva. Para lograrlo, separa en dos procesos paralelos la generación de las imágenes a dibujar en cada pantalla. La generación de la imagen de la pantalla trasera tiende ser más lenta debido a que es proporcional a la complejidad de la escena representada. Sin embargo,

la imagen de la barrera de paralaje es de una complejidad más baja y se mantiene sin sufrir grandes cambios. Gracias a la rapidez con que se puede calcular la barrera de paralaje y poder dibujarse independientemente es que Dynallax es capaz que corregir rápidamente la calidad estereoscópica. La calidad de la perspectiva está relacionada al proceso que dibuja la escena por lo que puede sufrir de un desfase temporal.

Dynallax puede alternar fácilmente entre modo tridimensional y modo bidimensional. Para trabajar en modo bidimensional, la pantalla frontal en vez de dibujar la barrera simplemente es dibujada en blanco permitiendo el paso de luz de todos los píxeles de la pantalla trasera. Incluso es capaz de combinar los dos tipos de imagen, para esto se dibuja la barrera de paralaje en toda la pantalla frontal menos en el sector que se mostrará de forma bidimensional (donde se dibuja en blanco). La Figura 19 muestra los distintos modos en los que puede trabajar Dynallax.



Figura 19. Dynallax permite imágenes en clásicas en dos dimensiones (izquierda), imágenes autoestereoscópicas tridimensionales (centro) y la combinación de ambas (derecha). [47]

Dynallax cuenta con algunas limitaciones que no permiten desarrollar el verdadero potencial de esta técnica y requieren ser estudiados en futuras investigaciones. Por ejemplo, las dos pantallas LCD sobrepuestas son notablemente más oscuras que en un sistema con barrera de paralaje estática; de hecho puede llegar tener hasta un 50% menos de contraste y brillo. Otra desventaja es la baja resolución efectiva y la visibilidad de las líneas de las barreras.

HR3D

El *medialab* del MIT a través del grupo *Camera Culture* [5] ha desarrollado la técnica HR3D [32] (*High-Rank 3D Display*) para crear una pantalla de imágenes autoestereoscópicas utilizando dos pantallas LCD sobrepuestas.

HR3D se basa en barreras de paralaje de contenido adaptativo que mejora la técnica de las barreras de paralaje clásicas permitiendo la mayor emisión posible de luz, manteniendo o incluso mejorando en ciertos casos la calidad de las imágenes tridimensionales. La clave está en armar un patrón entre ambas pantallas que no es estático y se modifica dependiendo del contenido de la escena. Eliminando la barrera estática fija, las barreras de paralaje de contenido adaptativo pueden aumentar significativamente el brillo de las pantallas autoestereoscópicas basadas en la tecnología LCD. Además, usando pantallas LCD con altas frecuencias de refresco combinadas con HR3D se ha llegado a lograr altas resoluciones de imágenes tridimensionales. Las pantallas LCD utilizadas son modificadas levemente para mejorar el desempeño, a la pantalla frontal se le remueve el filtro polarizado y la lámina difusora para aumentar el brillo.

En una barrera de paralaje convencional, el panel LCD frontal contiene líneas horizontales o puntos distribuidos uniformemente que permiten el pasaje de luz. El observador ve los píxeles del panel trasero a través de la grilla formada en el panel frontal, seleccionando un subconjunto visible de píxeles dependiendo de la ubicación. HR3D en vez de considerar a los paneles como una simple barrera, los considera como un modulador espacial de la luz que actúa para lograr recrear un campo de luz objetivo (imagen resultante) atenuando la luz emitida por la luz trasera. A diferencia de barreras convencionales, ambas máscaras pueden presentar opacidades no-binarias. Esto permite que la resolución final de la imagen sea la mayor posible, coincidiendo con la resolución del panel LCD frontal.

Una pantalla de barrera de paralaje, que contiene dos máscaras y una luz trasera, puede ser analizada como dispositivo emisor de un campo de luz. Como se puede ver en la Figura 20, un rayo de luz emitido es parametrizable por las coordenadas de intersección con cada máscara. Por lo tanto, el rayo (u, v, s, t) interseca a la máscara trasera en el punto (u, v) y a la máscara frontal en el punto (s, t) .

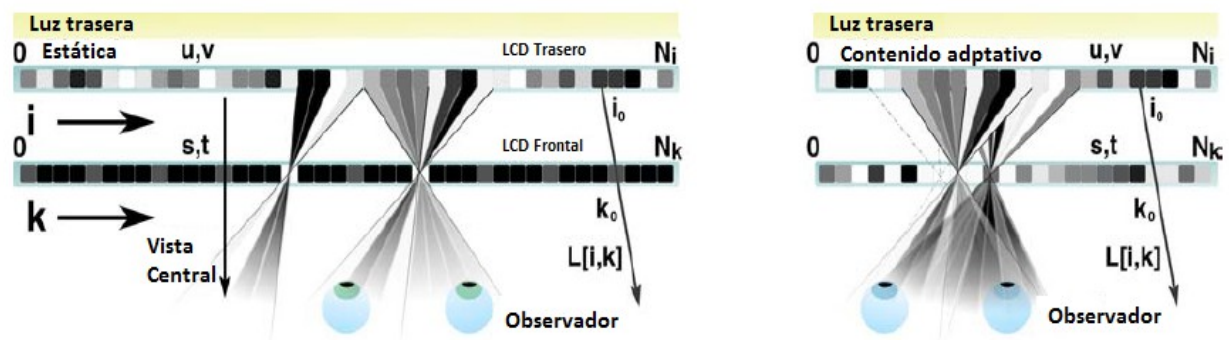


Figura 20. Comparación de barreras de paralaje convencionales (izquierda) contra barreras de contenido adaptativo (derecha). [32]

Dos funciones bidimensionales de atenuación de luz, con parámetros (u, v) y (s, t) respectivamente, son definidas junto con la parametrización. Estas funciones se corresponden con las máscaras trasera y frontal respectivamente. El campo de luz emitido es un función en cuatro dimensiones (u, v, s, t) definida por su producto.

Se han propuesto soluciones para mejorar el brillo de barreras de paralaje estáticas [29] que consisten en alternar en cada instante la posición de la barrera. Si la frecuencia de refresco de las pantallas LCD utilizadas son lo suficientemente altas, no se debería percibir degradación en la imagen. Este concepto es generalizable a las barreras de paralaje de contenido adaptativo.

Las barreras de paralaje estáticas son una solución particular de las funciones de atenuación, donde la función de la barrera frontal toma valores de cero o uno (opaco o transparente). Sin embargo, existen varias soluciones. HR3D lo plantea como un problema de factorización de matrices que minimice la distancia Euclidiana con peso al campo de luz objetivo L , bajo la restricción de no negativos (las máscaras solo pueden quitar luz pero no agregar).

La resolución del problema a la factorización no negativa de matrices es realizado aplicando métodos numéricos y en particular el propuesto por Blondel *et al* [3].

A continuación en la Figura 21 se muestra un ejemplo comparativo visual de los campos de luz obtenidos utilizando una barrera estática y una barrera de contenido adaptativo. También para ambas técnicas se muestra el resultado de aplicar la multiplexación temporal.

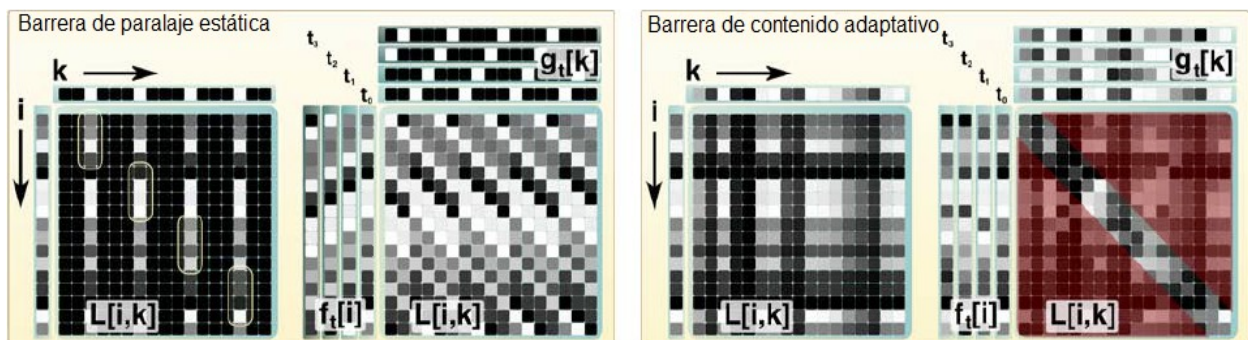


Figura 21. Comparativa de resultados entre una barrera estática y una barrera de contenido adaptativo. [32]

En la imagen de más a la izquierda se puede apreciar como las barreras estáticas logran una reconstrucción precisa en las zonas remarcadas (campos de vista centrales) pero con el costo de tener varias columnas oscuras. Réplicas de las zonas centrales son repetidas fuera de las zonas remarcadas. Los resultados obtenidos con una barrera de contenido adaptativo aumentan el brillo total de la imagen. Las aproximaciones obtenidas con ésta técnica con

multiplexación temporal no fuerzan la aparición de repeticiones periódicas (como sí lo hacen las barreras estáticas) y no limitan los rayos salientes fuera de la zona de visión central (mostrados en rojo).

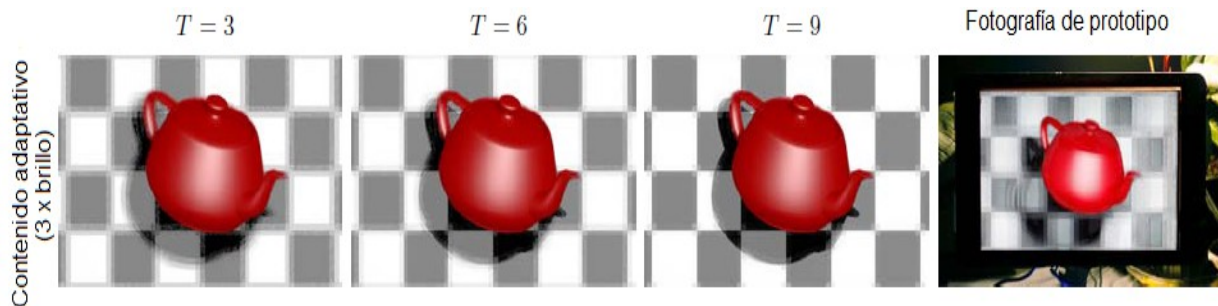


Figura 22. Imágenes obtenidas con barrera de paralaje de contenido adaptativo con multiplexación temporal (T corresponde a la cantidad de imágenes usadas en la multiplexación). [32]

Las imágenes logradas son notablemente más brillantes, sin embargo la técnica sufre de imperfecciones principalmente en los bordes entre objetos que pueden minimizarse aumentando la cantidad de imágenes por segundo y una mejor aproximación en la factorización no negativa. Cualquiera de estas dos mejoras aumenta el tiempo para generar cada imagen siendo una importante desventaja.

El costo de procesamiento de cada cuadro utilizando esta técnica es lento y no permite que sea utilizado para aplicaciones que generen imágenes en tiempo real.

Un trabajo a futuro sobre esta técnica es la investigación sobre el algoritmo de factorización no negativa. Una mejora en este algoritmo implicaría un aumento en la velocidad de convergencia del método numérico utilizado, lo que llevaría a mayor calidad de imagen y menor tiempo de procesamiento.

El enfoque utilizado por esta técnica es prometedor y abre nuevas puertas a la investigación sobre el tema, especialmente a medida que aumente la capacidad de cómputo y mejore la tecnología LCD.

MICROSOFT WEDGE

Microsoft ha realizado avances en este campo, particularmente el Grupo de Ciencias Aplicadas, con el desarrollo del dispositivo *Wedge* [61]. *Wedge* es una guía de luz de acrílico

(Polimetilmetacrilato) especialmente diseñada para emitir luz cuyos rayos son paralelos entre sí (luz colimada) y gracias a esta propiedad sirve para crear un sistema de luz direccional. Combinando el sistema de luz direccional con una pantalla LCD de alta frecuencia de refresco, Microsoft ha logrado construir un prototipo de pantalla autoestereoscópica.

Microsoft *Wedge* no es la primera pantalla autoestereoscópica en utilizar un sistema de luz direccional. Un proyecto anterior basado en el mismo concepto es el Cambridge Display [58], el cual utiliza lentes para dirigir la luz de forma alternada a cada ojo. Sin embargo, es una pantalla aparatosa y de gran tamaño que hace que sea poco práctica como producto comercial. En este punto, *Wedge* se destaca por su delgado tamaño haciéndolo una combinación ideal con las pantallas LCD de hoy en día.

El primer paso para explicar el funcionamiento del *Wedge* es establecer cómo logra emitir luz colimada o paralela. Ubicando una fuente de luz puntual en el extremo más fino de una guía de luz con forma de cuña, los rayos de luz van a estar confinados entre los dos planos de la superficie de la guía abriéndose en abanico como se muestra en la Figura 23. En el extremo grueso, los rayos son reflejados por una superficie espejada con una curvatura esférica con foco al medio del extremo fino (donde se ubica la fuente de luz) y viajan hacia el extremo fino en direcciones paralelas entre sí.

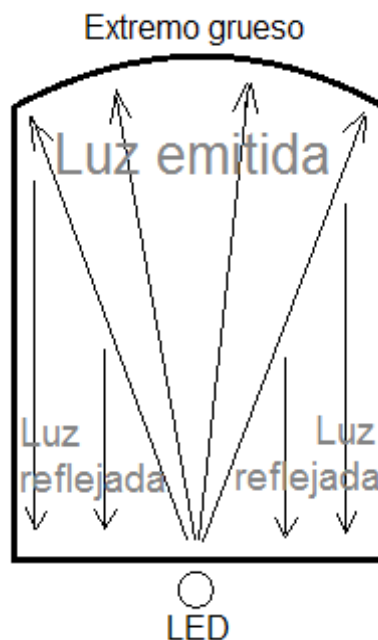


Figura 23. Vista frontal. Los rayos se abren en abanico hacia el extremo grueso donde son reflejados en paralelo. [59]



Figura 24. Sección transversal. El ángulo del rayo es incrementado por la reflexión en las caras anguladas del extremo grueso por lo tanto los rayos emergen al volver hacia el extremo fino. [59]

El corte transversal del *Wedge* es representado por la Figura 24. Luego de ser reflejados los rayos por el extremo grueso y viajar hacia el extremo fino, se busca que alcancen el ángulo crítico y emerjan de la guía de luz. Para poder llegar al ángulo crítico, al extremo grueso se le agregan caras en ángulo que incrementen el ángulo relativo a las superficies del *Wedge*.

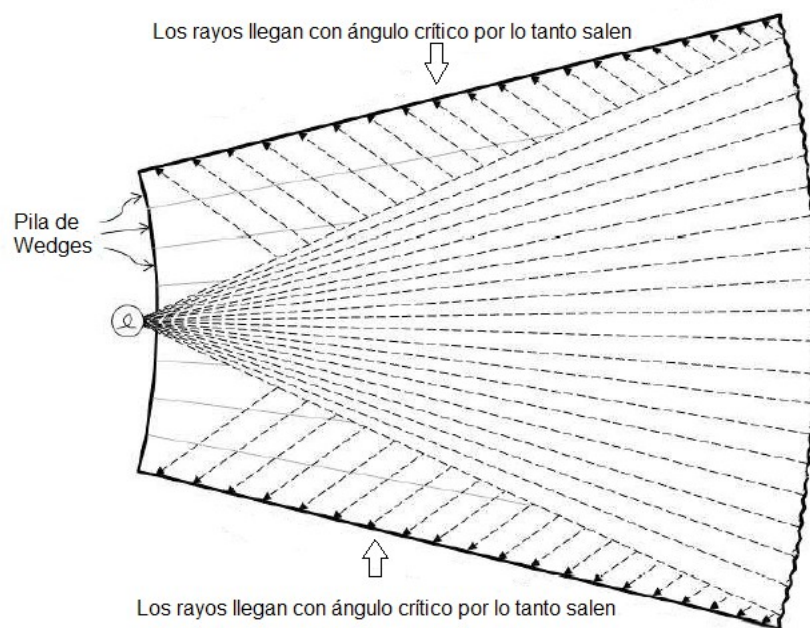


Figura 25. Los rayos viajan de igual manera que si atravesaran una pila de Wedges. La curvatura del extremo grueso colima los rayos y las caras anguladas giran la dirección de manera tal que los rayos iluminen toda la superficie alcanzando ángulo crítico. [59]

La numerosa cantidad de reflexiones de los rayos ocurridas al recorrer el *Wedge* pueden llevar a confusión, es más informativo y ópticamente equivalente seguir el camino de los rayos a través de una pila de réplicas del original como muestra la Figura 25. El ángulo de

curvatura del extremo grueso es elegido, dejando de lado las caras anguladas, manteniendo el mismo radio de curvatura constante que el de la Figura 23.

Todos los rayos deben alcanzar el ángulo crítico antes de salir por la superficie del *Wedge* y se quiere que la emisión sea uniforme en todos los puntos de la superficie. Para lograr que la emisión sea uniforme hay que definir adecuadamente el ángulo de las caras anguladas del extremo grueso. Este ángulo se calcula haciendo el camino inverso de los rayos salientes. Los rayos salientes son distribuidos uniformemente con ángulo de incidencia crítico y son rastreados hasta el extremo grueso. Luego, el ángulo de las caras del extremo grueso es definido como el ángulo necesario para que los rayos hagan foco en la fuente de luz del extremo fino.

Debe haber la misma cantidad de caras angulares frontales como de traseras en el extremo grueso ya que es igualmente probable que los rayos incidan en cualquiera de las dos direcciones. La Figura 25 muestra como hay rayos que no se reflejan como para emerger por la superficie frontal del *Wedge* y emergen por la superficie trasera. Esto se soluciona simplemente espejando la cara trasera del *Wedge* de forma de que los rayos emerjan por la superficie frontal en combinación con el resto.

Los rayos emitidos a través de la superficie del *Wedge* salen en un ángulo no perpendicular (salen en el ángulo crítico refractante) a la superficie. Para corregir el ángulo y hacer que sean efectivamente perpendiculares, se agrega un filtro sobre la superficie formado por un arreglo de prismas que modifica la dirección de los rayos.

Mirando la Figura 23 se puede apreciar como los rayos de luz saldrán perpendiculares a la superficie si la fuente emisora se encuentra al medio. En caso de que la fuente emisora se encuentre a la izquierda, los rayos emitidos se dirigirán a la derecha. De igual modo, si la fuente emisora se encuentra a la derecha, los rayos emitidos se dirigirán a la izquierda.

Una desventaja del sistema de direccionamiento de la luz en el *Wedge* es la aparición de sombras y zonas con mayor luminosidad al dirigir la luz hacia algún costado. La Figura 26 muestra el efecto ocurrido. Como se puede ver, si los costados del *Wedge* son reflexivos, se forma un triángulo de doble brillo en el lado opuesto. Por lo tanto, para eliminar las sombras por iluminar desde un costado se debe iluminar de la misma manera del lado opuesto. Esta solución no es aplicable en todos los casos, particularmente para una pantalla autoestereoscópica no es válido ya que puede generar interferencia entre las imágenes.

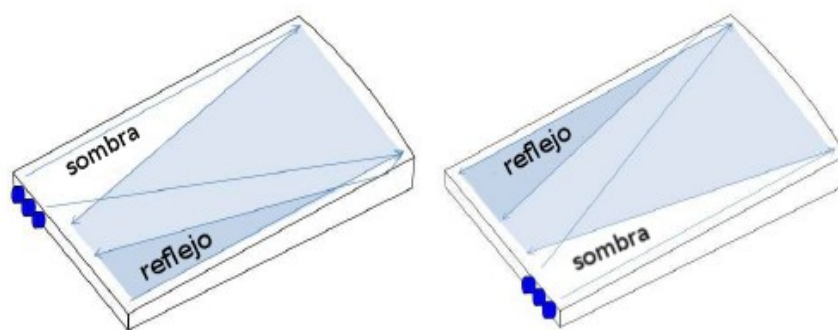


Figura 26. Rayos emitidos por LEDs hacia algún costado deja una sombra. [59]

En resumen *Wedge* es un dispositivo capaz de emitir luz colimada, uniformemente distribuida a través de su superficie y direccionable horizontalmente. La Figura 27 corresponde a una prueba realizada con LEDs rojos ubicados a la derecha y LEDs azules a la izquierda. La proyección generada en la pantalla tiene a la derecha el color azul y a la izquierda el color rojo, debido al diseño del direccionamiento de la luz detallado en párrafos anteriores.



Figura 27. Proyección sobre una pantalla a dos metros de distancia del Wedge. Los LEDs rojos y azules se encuentran separados 30 mm entre sí. [60]

Utilizando un arreglo de LEDs blancos a lo largo del extremo fino del *Wedge*, y encendiendo los LEDs adecuados en cada momento, se puede cambiar rápidamente la dirección horizontal de la luz. Al encenderse los LEDs a la izquierda, la luz es dirigida horizontalmente a la derecha; al encenderse los LEDs centrales, la luz es emitida perpendicular al plano del *Wedge* y al encenderse los LEDs de la derecha, la luz emitida es dirigida horizontalmente hacia la izquierda. Esto permite emitir luz direccional a través del *Wedge* de forma tal que solamente le llegue a un ojo a la vez. La dirección con que se emite la luz es sincronizada con la imagen dibujada en un LCD con alta frecuencia de refresco. Esta sincronización debe ser de tal manera que al emitir luz que llegue al ojo izquierdo, el LCD dibuja la imagen

correspondiente a la perspectiva del ojo izquierdo y en el instante siguiente lo mismo ocurre para el ojo derecho. Alternando constantemente la emisión de imágenes entre los ojos a altas frecuencias es que logra emitir una imagen distinta a cada ojo, logrando así la autoestereoscopia.

Para saber la posición de los ojos del observador en el espacio se utiliza una cámara Kinect de Microsoft [36]. Esta cámara facilita obtener información tridimensional sobre las escenas, en este caso la ubicación del observador. Con la tecnología actual de LCD existen pantallas con frecuencias de refresco de hasta 240Hz que permiten soportar hasta dos observadores con *Wedge* (cuatro ojos con una imagen por ojo a 60Hz cada una).

Entre las principales ventajas de esta técnica están el tamaño compacto, la capacidad de lograr imágenes en altas resoluciones y que no hay grandes pérdidas de brillo en las imágenes. Como desventaja principal se puede destacar las zonas sombreadas en la pantalla al observar desde ciertos ángulos. Otras desventajas son la alta complejidad en la construcción de la guía de luz de acrílico y rango de funcionamiento limitado a los ángulos máximos de direccionamiento de la luz. El *Wedge* de Microsoft además de lograr buenos resultados abre una puerta a nuevos avances e investigaciones en el campo gracias a su enfoque innovador al problema.

EVALUACIÓN DE TÉCNICAS

En base al objetivo planteado de construir un prototipo que no requiera el uso de aditamentos por parte del observador, se evaluaron algunas de las técnicas relevadas en este capítulo como punto de partida para diseñar una solución.

En primer lugar se estudió la posibilidad de crear un dispositivo similar al *Wedge* de Microsoft. Esta técnica cumple con ser autoestereoscópica y dar libertad de movimiento al usuario. Sin embargo, no se pudo conseguir en el mercado local a ningún proveedor que pudiese fabricar la pieza de acrílico que sirve como guía de la luz. Proveedores en el extranjero fueron descartados por motivos de falta de presupuesto y además incrementarían los tiempos del proyecto.

También se hicieron pruebas con hojas lenticulares en conjunto con una pantalla LCD para el prototipo. Estas hojas lenticulares tienen la ventaja de tener un costo accesible y permite una rápida construcción de una prueba de concepto del funcionamiento.

Sin embargo, esta técnica basada en el uso de hojas lenticulares tiene varios inconvenientes. El principal inconveniente es que no le permite libertad de movimiento al usuario, el uso de

láminas lenticulares es efectivo a una distancia definida por su longitud focal. Incluso manteniendo la distancia apropiada a la pantalla tampoco permite al usuario moverse hacia los costados, ya que se intercalan imágenes para ojo izquierdo y derecho, y se cruzan las imágenes del ojo izquierdo con el derecho perdiendo la correcta estereoscopía de la imagen. Las hojas lenticulares al estar formadas por lentes hacen aumento de los píxeles de la pantalla LCD, bajando considerablemente la resolución e incluso llegando a percibir los subpíxeles individualmente (subdivisiones del píxel en los colores rojo, verde y azul).

Por último, se estudió la técnica de barreras de paralaje. Las barreras de paralaje por lo general son estáticas y esto no permite la libertad de movimiento del usuario. Para poder lograr que funcione correctamente y que el usuario pueda moverse la barrera tiene que ser dinámica, es decir, la barrera que filtra las imágenes para cada ojo y la imagen a ser filtrada deben ser generadas dinámicamente en base a la ubicación de los ojos del usuario. Utilizando una barrera de paralaje dinámica se puede cumplir con todos los requerimientos establecidos. Por lo tanto esta técnica es la seleccionada como base para la construcción del prototipo.

4. SOLUCIÓN PLANTEADA

DESCRIPCIÓN GENERAL

En este proyecto se desarrolla un prototipo de dispositivo de visualización tridimensional autoestereoscópico y una aplicación para demostrar los posibles usos del dispositivo.

El dispositivo construido es un monitor tridimensional en el cual el usuario no necesita de gafas para percibir la profundidad. El monitor funciona en base a una barrera de paralaje dinámica y la detección de posición de los ojos del usuario.

La barrera de paralaje dinámica está formada por dos pantallas superpuestas. En la pantalla trasera es donde se dibujan las imágenes correspondientes a cada ojo combinadas en un patrón adecuado a la posición del usuario. La pantalla frontal es transparente y se ubica por delante de la otra pantalla. En la pantalla frontal, en base a la ubicación del usuario, se dibujan barreras verticales negras que filtran las imágenes de la pantalla trasera de forma tal que cada ojo pueda ver solamente los píxeles de la imagen que le corresponde.



Figura 28. Imagen del prototipo construido, formado por dos pantallas LCD superpuestas y una cámara Kinect.

Para la construcción del dispositivo se utilizó una pantalla LCD regular (trasera) con una pantalla LCD transparente delante (frontal). Ambas pantallas son conectadas a una computadora mediante los conectores de salida estándar de video (VGA, DVI o HDMI) y son utilizadas por el sistema operativo como dos monitores externos comunes e independientes entre sí. Una aplicación basada en OpenFrameworks [43] y OpenGL [44] se encarga de generar los contenidos y dibujarlos en cada una de las pantallas LCD, y realiza el seguimiento de la posición de los ojos del usuario.

Se utiliza la cámara Kinect de Microsoft para hacer el seguimiento de la posición de los ojos del usuario. Esta cámara tiene la capacidad de capturar información de profundidad, además de la imagen, que facilita el reconocimiento de personas en la escena. La API (siglas en inglés de Interfaz de Programación de Aplicaciones, *Application Programming Interface*) de Kinect para Windows [31] de Microsoft es utilizada, en conjunto con la cámara, para el seguimiento de la posición de los ojos y manos del usuario. La posición de la cámara no es fija respecto de la ubicación del dispositivo y puede ubicarse libremente para filmar al usuario.

Para el desarrollo de la aplicación se ha creado una biblioteca que facilita la programación de aplicaciones para el prototipo de barrera de paralaje dinámica. La biblioteca se encarga de procesar las imágenes correspondientes a cada ojo del usuario y la posición del usuario, proporcionadas por la aplicación, para crear y desplegar las imágenes con los patrones apropiados en el monitor. Las imágenes creadas son dos, una es la imagen a ser desplegada en la pantalla LCD trasera, que combina verticalmente las imágenes correspondientes a cada ojo, y la otra es la imagen a ser desplegada en la pantalla LCD frontal, correspondiente a la barrera de líneas verticales que filtra la imagen del LCD anterior. La creación de ambas imágenes se hace en la tarjeta de video, aprovechando la capacidad de procesar en forma paralela por cada píxel, para obtener un alto rendimiento y poder generar contenido en tiempo real.

La biblioteca creada es un programa multiplataforma de código abierto [51], también basada en OpenFrameworks y OpenGL.

La aplicación desarrollada para el dispositivo muestra pruebas de concepto de visualización de fotografías estereoscópicas, videos tridimensionales, aplicaciones interactivas y visualización de escenas tridimensionales de gráficos creados por computadora en tiempo real.

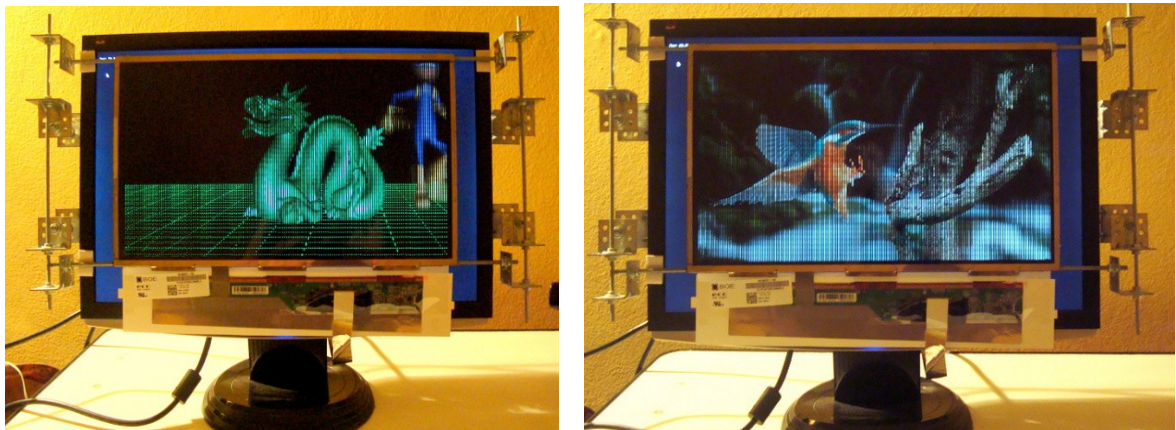


Figura 29. Imágenes del funcionamiento del prototipo. La imagen de la izquierda muestra gráficos generados en tiempo real mientras que la imagen de la derecha muestra una fotografía estereoscópica.

REQUERIMIENTOS

Para el diseño de la técnica a utilizar en el prototipo se tuvieron en cuenta los siguientes requerimientos funcionales:

1. El dispositivo debe ser autoestereoscópico, es decir, el usuario debe poder percibir la profundidad de la imagen sin la necesidad de aditamento alguno.
2. El usuario debe poder percibir la profundidad sin necesidad de estar ubicado en un punto fijo. El dispositivo debe dar la posibilidad de libre movimiento al usuario dentro de un espacio acotado.

Además, también se consideraron los siguientes requerimientos no funcionales:

1. El tiempo de procesamiento necesario para la desplegar una imagen en el dispositivo debe ser lo suficientemente rápido para permitir que una aplicación genere en contenido en tiempo real. Esto implica una velocidad de por lo menos treinta cuadros por segundo.
2. Para facilitar el desarrollo de aplicaciones para el dispositivo, se debe contar con una herramienta (biblioteca o framework) de software que faciliten el uso del mismo.
3. Los recursos necesarios para el desarrollo del prototipo deben poder conseguirse dentro del mercado local.

BARRERA DE PARALAJE DINÁMICA

En esta sección se describe el diseño utilizado para el desarrollo del prototipo de la barrera de paralaje dinámica. En las siguientes subsecciones se cubre la base matemática de la solución, el diseño de hardware y software.

SOLUCIÓN MATEMÁTICA

Si bien las imágenes en el prototipo se dibujan sobre una pantalla discreta dividida en píxeles, para el análisis matemático del problema se trabajó sobre una superficie continua para encontrar la solución. Una vez encontrada la solución en la superficie continua, el resultado es aproximado a los puntos discretos (píxeles) de manera de introducir el menor error posible.

La barrera de paralaje está formada por barras verticales tanto en la imagen a filtrar como en la imagen que sirve de filtro. Por lo tanto se trabajará la solución en un espacio bidimensional, en el cual se ha eliminado el eje vertical para simplificar los cálculos.

La lógica detrás de una barrera de paralaje consiste en que existe una imagen que combina verticalmente las vistas para cada ojo, y existe un filtro de barreras verticales, por delante de la imagen, que permiten que cada ojo pueda ver la vista que le corresponde y bloquea la vista correspondiente al otro ojo.

La composición de la imagen con las vistas combinadas y la imagen de filtro se define por los pasos siguientes:

1. Definir como punto de inicio de la iteración al extremo izquierdo de pantalla con la imagen de las vistas combinadas. Se denomina este punto como P.
2. Trazar una recta definida por la posición del ojo derecho y P. Al punto de la intersección de la recta con el filtro se la denomina F.

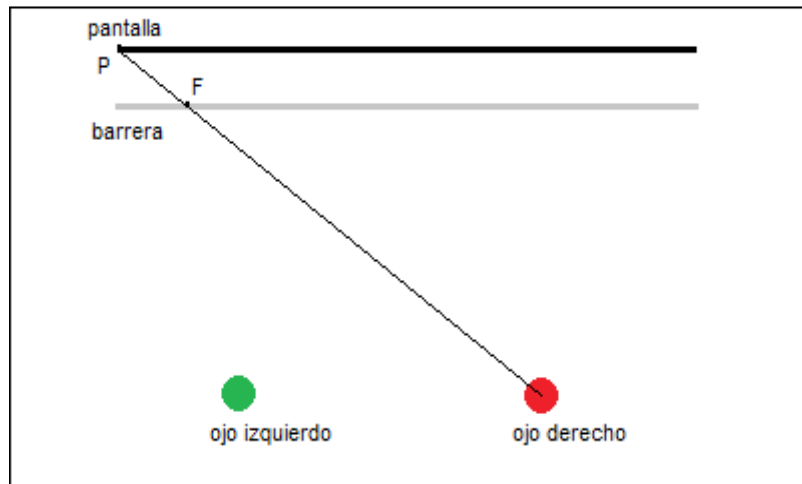


Figura 30. Representación gráfica del paso 2.

3. Trazar una recta definida por la posición del ojo izquierdo y F. El punto de la intersección de la recta con la pantalla define un nuevo punto P.

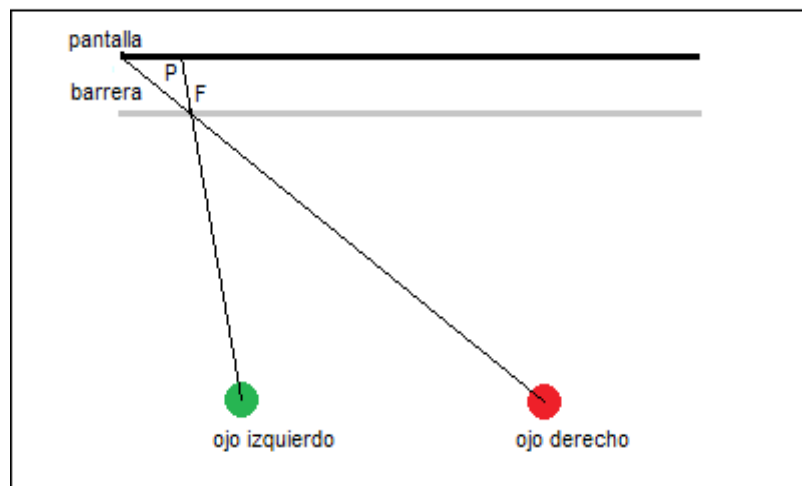


Figura 31. Representación gráfica del paso 3.

4. Iterar a partir del paso 2 hasta alcanzar el extremo derecho de la pantalla.

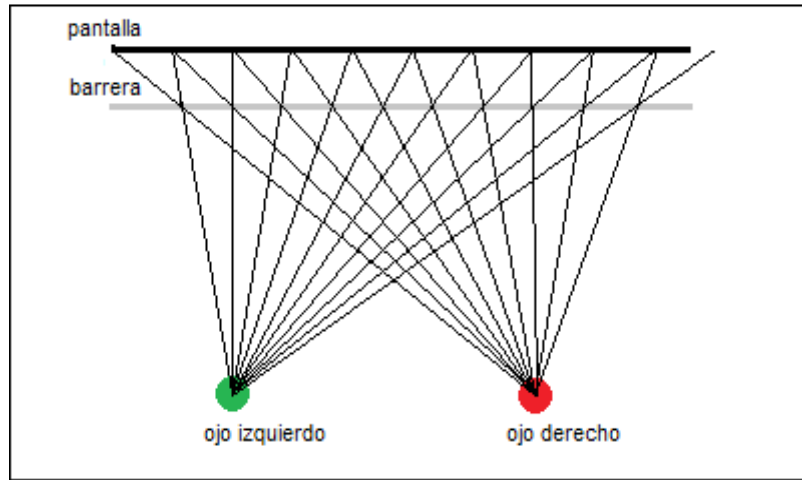


Figura 32. Representación gráfica del resultado de la iteración del paso 4.

El conjunto de puntos P y el conjunto de puntos F definidos en las iteraciones de los pasos numerados, definen cómo estarán compuestas la imagen de la combinación de las vistas y el filtro. Los tramos definidos por los puntos P corresponden a la vista de cada ojo de forma intercalada. Los tramos definidos por los puntos F corresponden a zonas donde se permite el paso de la luz y zonas donde no, de forma intercalada.

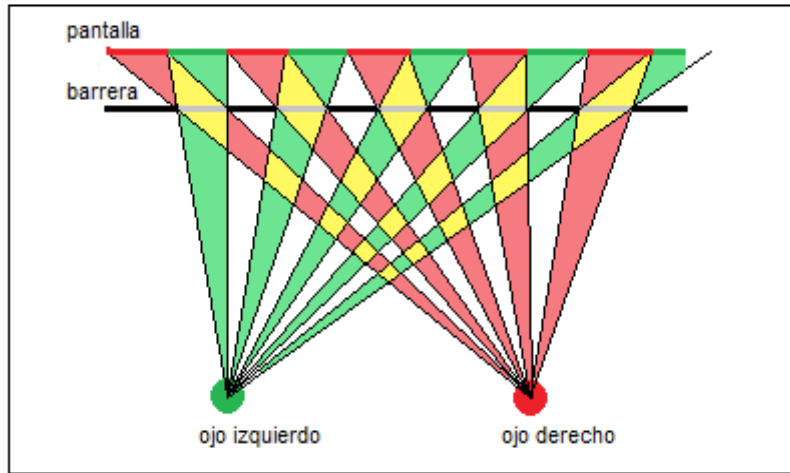


Figura 33. Campo de visión de los ojos en la barrera de paralaje dinámica. En rojo se muestra la vista del ojo derecho y en verde la del ojo izquierdo.

A continuación se presentan los cálculos necesarios para realizar estos mismos pasos. Se va a trabajar asumiendo que la pantalla se encuentra sobre la recta $y=0$, la barrera se encuentra sobre la recta $y=1$ y las posiciones de los ojos son $i = (i_x, i_y)$ y $d = (d_x, d_y)$.

Para el paso 1, se define el punto P de origen de la iteración como $P=(\alpha,0)$. Al comenzar la iteración en el extremo izquierdo de la pantalla se asigna $\alpha = 0$.

Para los pasos 2 y 3, es necesario plantear las ecuaciones de las rectas y resolver las intersecciones correspondientes que veremos ahora.

Dada una posición $(x_p,0)$ en la pantalla, la ecuación de la recta que pasa por $(x_p,0)$ y d está dada por:

$$\frac{x-x_p}{d_x-x_p} = \frac{y-0}{d_y-0} \quad , \text{ que puede ser planteada como } x = \frac{y \cdot d_x}{d_y} + x_p \cdot \left(1 - \frac{1}{d_y}\right)$$

En base a la recta anterior, se define una función para obtener la coordenada x de la intersección de la barrera ($y = 1$) con una recta definida por d y cualquier punto de la pantalla:

$$barrera_d(x) = \frac{d_x}{d_y} + x \cdot \left(1 - \frac{1}{d_y}\right)$$

Dado un punto cualquiera $(x,1)$ de la barrera, se puede encontrar el valor correspondiente en la pantalla invirtiendo la ecuación anterior:

$$pantalla_d(x) = \frac{\left(x - \frac{d_x}{d_y}\right)}{\left(1 - \frac{1}{d_y}\right)}$$

Por lo tanto, reproduciendo los pasos 2 y 3 descritos anteriormente, dada una posición X_n en la pantalla, el siguiente punto X_{n+1} está definido por:

$$X_{n+1} = pantalla_i(barrera_d(X_n))$$

Esta sucesión corresponde al paso 4. El último paso de la sucesión está definido por el primero en igualar o superar el extremo derecho de la pantalla.

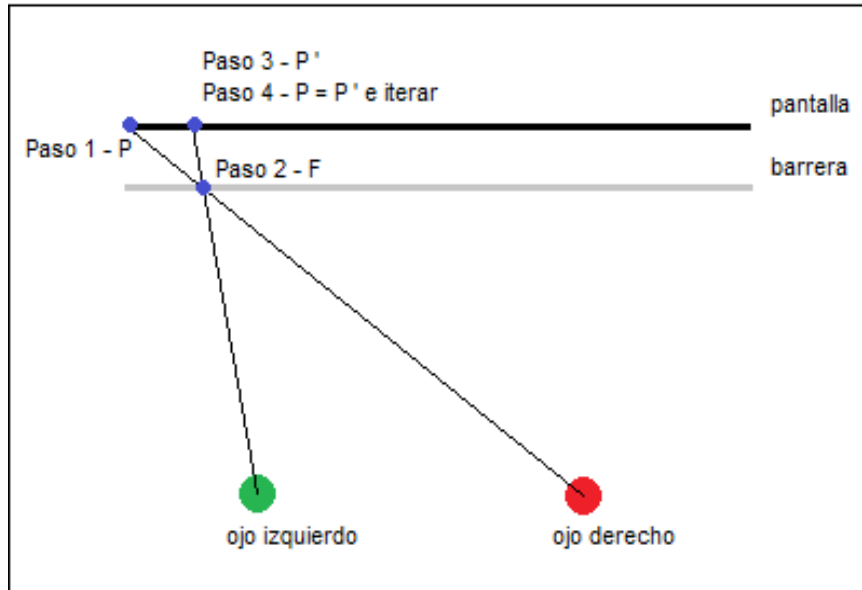


Figura 34. Representación gráfica de los pasos para armar dinámicamente la barrera de paralaje.

Ahora trabajaremos sobre la sucesión de modo de llegar a una expresión que dependa de n y no del paso anterior.

Expandiendo la fórmula se obtiene:

$$X_{n+1} = \frac{\left(\frac{d_x}{d_y} + X_n \cdot \left(1 - \frac{1}{d_y}\right) - \frac{i_x}{i_y}\right)}{\left(1 - \frac{1}{i_y}\right)}$$

Lo cual puede expresarse como una ecuación lineal $X_{n+1} = A X_n + B$, donde:

$$A = \frac{\left(1 - \frac{1}{d_y}\right)}{\left(1 - \frac{1}{i_y}\right)}, \quad B = \frac{\left(\frac{d_x}{d_y} - \frac{i_x}{i_y}\right)}{\left(1 - \frac{1}{i_y}\right)}$$

Dado α como valor inicial para iterar (por lo general el extremo izquierdo de la pantalla), el n -ésimo valor de la sucesión puede calcularse iterando en $X_{n+1} = A X_n + B$:

$$\begin{aligned}
X_0 &= \alpha \\
X_1 &= A\alpha + B \\
X_2 &= A^2\alpha + AB + B \\
X_3 &= A^3\alpha + A^2B + AB + B \\
&\vdots \\
X_n &= A^n\alpha + B(A^{n-1} + \dots + A + 1)
\end{aligned}$$

Para optimizar el código que implementa esta solución, es útil calcular en paralelo a que vista corresponde cada píxel en la pantalla. Para saber a qué vista corresponde un píxel, hay que saber cuál es el elemento anterior de la sucesión y cuál es el siguiente.

Viendo a la sucesión como una función que varía en función de n , es necesario definir una función inversa que en base a un valor de x devuelva el correspondiente valor de n .

Se consideran dos casos para definir esta nueva función dependiendo del valor de A .

Si $A = 1$, esto ocurre cuando ambos ojos están a la misma distancia de la pantalla, entonces:

$$X_n = \alpha + Bn, \text{ y por lo tanto: } inversoSucesion(x) = \frac{x - \alpha}{B}$$

Si $A \neq 1$, los ojos están a distancias distintas de la pantalla, entonces:

$$X_n = A^n\alpha + B \overbrace{(A^{n-1} + \dots + A + 1)}^{\text{serie geométrica}} \Rightarrow X_n = A^n\alpha + \frac{B(1 - A^n)}{1 - A}$$

$$X_n(1 - A) = A^n\alpha(1 - A) + B(1 - A^n) = A^n\alpha - A^{n+1}\alpha + B - BA^n$$

$$X_n(1 - A) = -\alpha A^{n+1} + (\alpha - B)A^n + B$$

$$X_n(1 - A) - B = (\alpha - B - A\alpha)A^n$$

$$\frac{X_n(1 - A) - B}{(\alpha - B - A\alpha)} = A^n \Rightarrow \log_A\left(\frac{X_n(1 - A) - B}{(\alpha - B - A\alpha)}\right) = n$$

$$\text{y por lo tanto: } inversoSucesion(x) = \log_A\left(\frac{x(1 - A) - B}{(\alpha - B - A\alpha)}\right)$$

La función *inversoSucesion* permite saber a qué vista corresponde un píxel de la pantalla. Evaluando con esta función ambos extremos de un píxel de la pantalla se puede saber entre cuáles puntos de la sucesión se encuentra, lo que define a que vista corresponde el píxel. Si el valor entero del resultado para ambos extremos del píxel es el mismo y el valor entero es impar, entonces el píxel pertenece a la vista del ojo izquierdo del usuario. Si el valor entero del resultado para ambos extremos del píxel es el mismo y el valor entero es par, entonces el píxel pertenece a la vista del ojo derecho del usuario.

En el caso que el valor entero del resultado para ambos extremos del píxel de la pantalla no coincide y si alguna de las vistas llega a cubrir un porcentaje superior al 95% del píxel, entonces se asigna el color de dicha vista. De lo contrario el píxel se deja negro de modo de no introducir error en la imagen a mostrar al usuario.

Para el cálculo de los píxeles de la barrera se proyectan los puntos sobre la pantalla, a través de la recta que une un punto en la barrera y la posición del ojo derecho, y se evalúan de manera similar a los de la pantalla. La diferencia en este caso es que los píxeles de la barrera toman valores blanco (permite el paso de la luz) o negro (no permite el paso de la luz), dependiendo entre qué puntos de la sucesión se encuentran.

HARDWARE

Para la construcción del prototipo de la barrera de paralaje dinámica se decidió trabajar utilizando monitores LCD. Los monitores LCD ofrecen varias ventajas útiles. En primer lugar y fundamental para armar una barrera de paralaje, es que el panel LCD (removiéndolo de la estructura del monitor) es transparente. Esto permite filtrar la luz que atraviesa un panel LCD, en particular utilizando barreras de líneas verticales. La segunda ventaja que ofrece es que soportan altas resoluciones con tamaños de píxel pequeños, esto permite el dibujado de las barreras de forma precisa. Por último, el bajo costo y popularidad en la actualidad de estos monitores hace que sean sencillos de conseguir en el mercado local.

Cada píxel en un LCD está formado por subpíxeles por cada color base (rojo, azul y verde, por lo general). Cada subpíxel consiste de una capa de cristal líquido entre dos electrodos transparentes y dos filtros polarizados (perpendiculares entre sí). Mientras no se le aplica corriente eléctrica, la orientación de las moléculas del cristal líquido es determinada por la orientación de la superficie de los electrodos. Las orientaciones de las superficies tienen una diferencia de 90°, por lo tanto, las moléculas forman una estructura helicoidal. Esta estructura induce a la rotación de la polarización de la luz incidente y permite que la luz no

sea bloqueada por el segundo filtro polarizado. Al no ser bloqueada la luz, el subpíxel se ve del color definido por el filtro de color. [14]

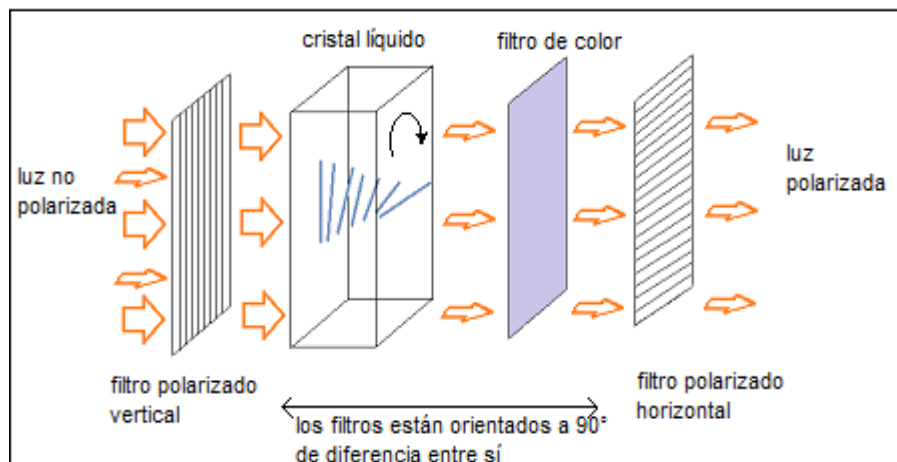


Figura 35. Diagrama de un subpíxel de un monitor LCD.

Cuando una corriente eléctrica es aplicada con el voltaje adecuado al cristal líquido, las moléculas se orientan paralelas al campo eléctrico y pierden la estructura helicoidal. En esta nueva estructura se reduce casi completamente la rotación de la polarización de luz y la polarización de luz incidente es perpendicular al segundo filtro al llegar. El segundo filtro bloquea la luz incidente y en consecuencia el subpíxel se ve negro. Diferentes voltajes permiten reducir en menor medida la pérdida de la estructura helicoidal del cristal líquido, rotando en menor medida la polarización y, por lo tanto, permitiendo variar la intensidad de luz del subpíxel.

El prototipo está formado por dos pantallas LCD superpuestas una delante de la otra. Para la pantalla trasera se utiliza un monitor LCD sin ninguna modificación en particular. Esta pantalla es la encargada de desplegar la imagen de la vista para ambos ojos combinados según el patrón vertical calculado.



Figura 36. Imagen de aplicación interactiva de dibujado desarrollada para el dispositivo.

Para la pantalla frontal se utiliza como base un monitor LCD el cual es modificado. Se le remueve el armazón y la base, y se utiliza solamente la pantalla LCD. A la pantalla se le remueve la fuente de luz y las láminas reflectoras de luz, de modo que solamente quede una pantalla transparente con los píxeles. Esta pantalla transparente es ubicada por delante de la pantalla trasera y será la que dibuje la barrera que filtra las vistas correspondientes a cada ojo.

Como se vio en el funcionamiento de los subpíxeles LCD, por sí misma la pantalla no emite luz sino que solamente deja o no deja pasar la luz incidente. En este caso la fuente de luz de la pantalla frontal pasa a ser la imagen desplegada por la pantalla trasera del prototipo. Para que pueda funcionar correctamente este sistema, la polarización de la luz que emite la pantalla trasera debe coincidir con el polarizado del primer filtro de la pantalla delantera. Utilizar simplemente el mismo modelo de monitor para las pantallas trasera y frontal no sirve ya que la polarización de la luz saliente de la pantalla trasera sería perpendicular al primer filtro polarizado de la pantalla frontal. Las alternativas para la pantalla frontal son utilizar el mismo modelo de monitor rotado 90° o utilizar un modelo de monitor diferente cuyo primer filtro polarizado coincida con el de la luz entrante.

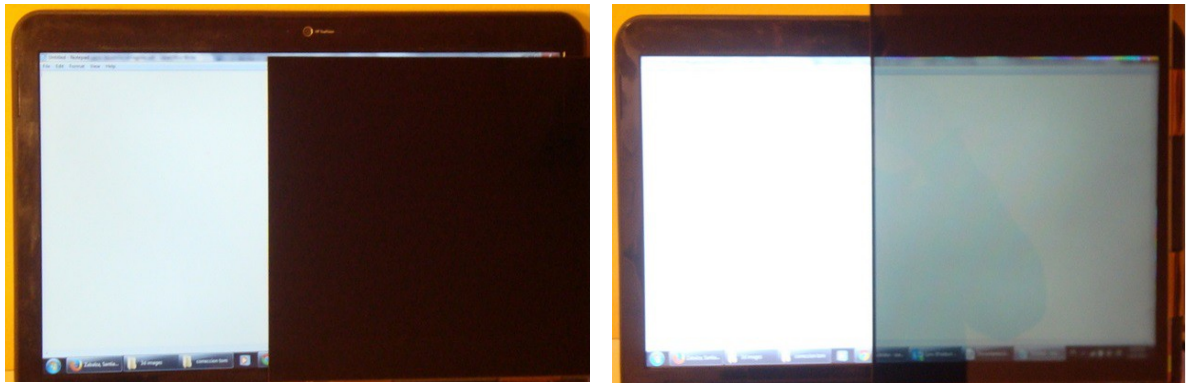


Figura 37. En la imagen de la izquierda la pantalla frontal tiene un filtro de luz polarizada perpendicular a la polarización de la luz que sale de la pantalla trasera. En la imagen de la derecha la pantalla frontal está rotada 90° permitiendo el paso de la luz de la pantalla trasera.

Otro punto a evaluar en la construcción del prototipo, es el tipo de terminado de la superficie de las pantallas.

La pantalla frontal debe tener terminado brillante. El terminado mate de los monitores difumina la luz, lo cual funciona correctamente cuando los píxeles a mostrar se encuentran en contacto con la capa difusora. En el caso de la pantalla frontal, ésta se encuentra separada de la capa de píxeles a mostrar (los píxeles a mostrar son los de la pantalla trasera) y un terminado mate hace perder nitidez de la imagen resultante.

Para la pantalla trasera es conveniente usar un monitor con terminado mate, el uso de un terminado brillante genera un patrón de moiré (patrón de interferencia visual causado por la superposición de dos capas con patrones similares [4]) causado por la interferencia entre los subpíxeles. El terminado mate, al difuminar levemente los píxeles, evita la creación del patrón.

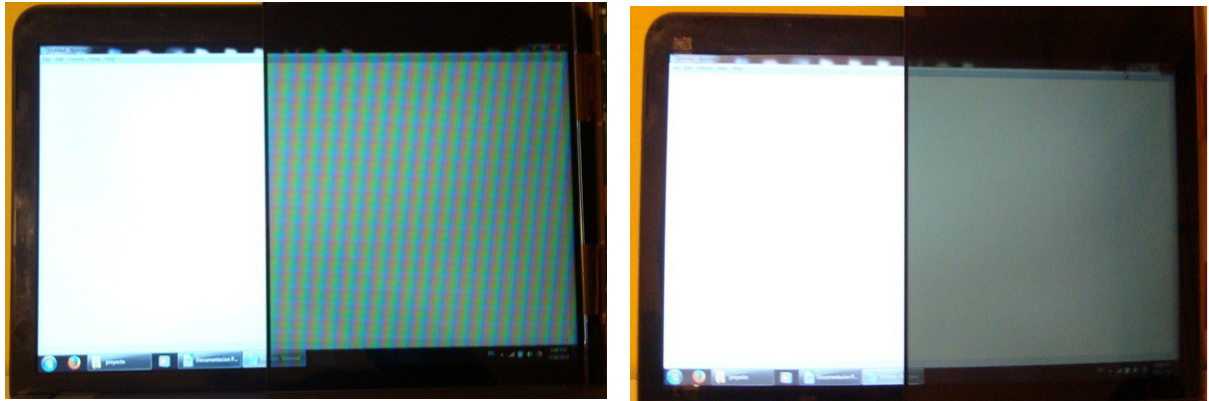


Figura 38. La imagen de la izquierda muestra el patrón de moiré que se genera al usar una pantalla trasera con terminación brillante. La imagen de la derecha usa una pantalla trasera con terminación mate que no genera patrones visuales que distorsionen la imagen.

Para definir la distancia de separación entre las pantallas trasera y frontal hay que considerar varios factores. El objetivo principal al elegir la distancia de separación es que las barreras de filtro de la barrera de paralaje sean lo más angostas posibles. Cuanto más angostas sean las barreras, son menos perceptibles para el observador y permiten una mejor comprensión de las vistas.

Si nos basamos únicamente en la solución matemática, suponiendo un modelo continuo sin separación por píxeles, al ser menor la distancia entre las pantallas también es menor el ancho de las barreras. Esto llevaría a pensar que cuanto más juntas estén las pantallas es mejor.

Sin embargo, al tener en cuenta que las pantallas están formadas por píxeles discretos, si las pantallas están muy juntas el algoritmo puede definir varios puntos de las iteraciones dentro de un mismo píxel. Estos píxeles en la barrera de paralaje quedarían indefinidos, es decir deberían ser en parte negros y en parte blancos. Los casos de los píxeles indefinidos se resuelven dejando el píxel en negro para que no le llegue información cruzada al observador (pueda ver con un ojo píxeles de la vista correspondiente al otro ojo). Esto lleva a que la distancia entre las pantallas no puede ser la mínima posible, sino que debe ser la menor posible de forma que los puntos en la pantalla frontal de las iteraciones del algoritmo estén separados a una distancia mayor al tamaño de un píxel de la pantalla frontal.

Si bien el terminado mate de la pantalla trasera evita en gran parte la formación del patrón de moiré, para distancias menores a un centímetro de separación entre las pantallas se puede percibir de forma leve el patrón. La distancia de separación debe estar dentro del rango de 1cm a 1.5cm, cuanto menor sea la distancia menor será el ancho de las barreras

pero se percibirá más el patrón de moiré (aunque a un nivel aceptable donde no afecta notoriamente la visualización de las imágenes).

En base a todos estos puntos se construyó el prototipo utilizando un monitor Viewsonic VX1940w [38] para la pantalla trasera y un monitor AOC E1660SW [37] para la pantalla frontal. Estos dos monitores utilizan filtros polarizados perpendiculares por lo que no hay necesidad de rotar la pantalla frontal. Al monitor AOC se le removió el armazón, la fuente de luz y láminas reflectoras. Se conservó la pantalla LCD y las placas controladoras.



Figura 39. La imagen izquierda muestra el monitor AOC sin el armazón y separada la pantalla de la placa controladora. En la imagen de la derecha se le quita al monitor la fuente de luz y las láminas reflectoras.

Al monitor Viewsonic que funciona como pantalla trasera, se le agregó una estructura metálica para poder soportar la pantalla LCD frontal y su placa de control de manera estable. La distancia de separación entre las pantallas es de 1.4cm, es una distancia aceptable donde prácticamente no se ve ningún patrón que distorsione la imagen.

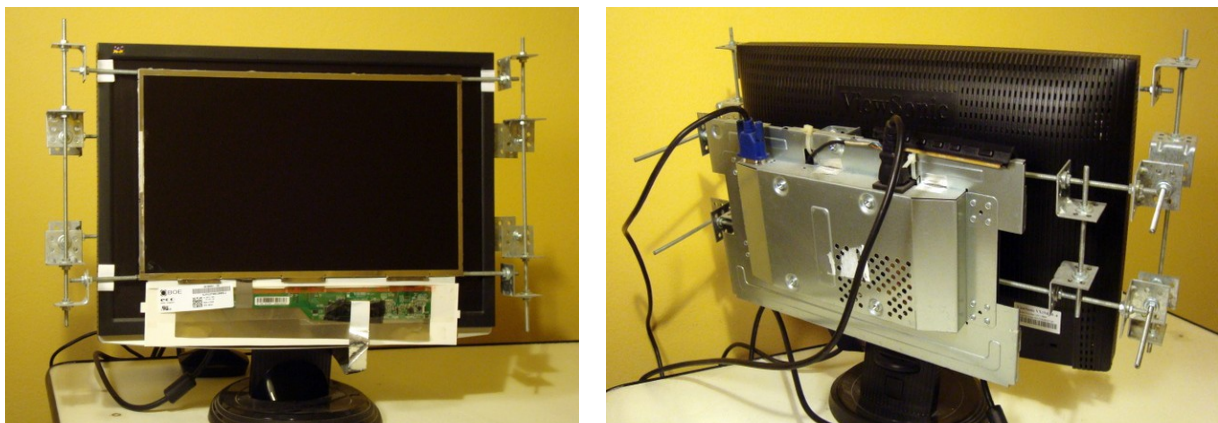


Figura 40. Imagen del prototipo construido. El monitor utilizado como pantalla trasera soporta mediante una estructura a la pantalla frontal y sus placas controladoras que van por detrás.

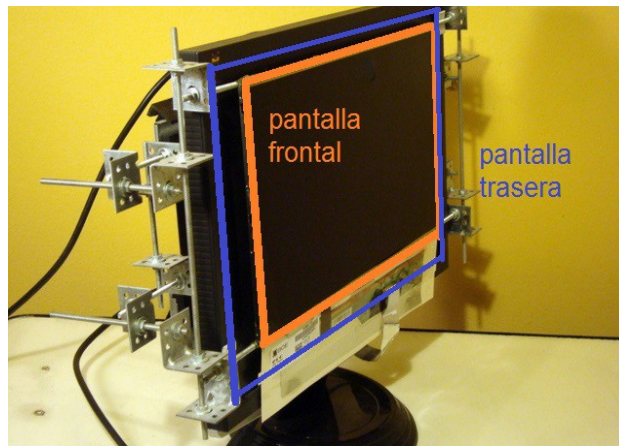


Figura 41. Identificación de las pantallas del prototipo.

Si bien el monitor Viewsonic de la pantalla trasera es de 19" y el AOC de la pantalla frontal es de 17" y además trabajan a distintas resoluciones nativas, no es un problema para el correcto funcionamiento y el software se encarga de tener estos valores en cuenta.

Ambas pantallas son conectadas a una tarjeta de video con doble conexión de salida y son utilizadas por el sistema operativo como dos monitores externos comunes. En particular el monitor Viewsonic es conectado mediante la salida DVI y el monitor AOC mediante VGA. Es importante que ambos monitores sean utilizados en sus resoluciones nativas para que funcione correctamente el prototipo, al usar resoluciones no nativas la interpolación de los píxeles en las pantallas genera valores incorrectos.

Para seguir la posición del usuario y ubicar sus ojos se utiliza una cámara Kinect de Microsoft. La cámara no lleva una posición fija en el prototipo y puede ser ubicada donde sea mejor para cada caso. La cámara Kinect tiene un rango de funcionamiento de entre 0.8m y 4m, fuera de este rango no logra identificar correctamente al usuario. La libre ubicación de la cámara permite mayor libertad al usar el prototipo, como por ejemplo si la cámara es ubicada por detrás de la pantalla se puede trabajar a distancias menores a 0.8m de la pantalla. La libre ubicación de la cámara es acompañada necesariamente por una calibración por software al comenzar a trabajar con el prototipo para poder relacionar la posición de la cámara con la de la pantalla.

El software para el prototipo está desarrollado utilizando el lenguaje C++. Además, utiliza OpenGL como biblioteca para los gráficos y el framework OpenFrameworks funcionando sobre OpenGL para facilitar el desarrollo de la solución.

La aplicación dibuja en la pantalla frontal del prototipo una imagen formada por líneas verticales blancas y negras calculada en base a la posición del usuario. Las líneas blancas en la pantalla frontal son vistas como zonas transparentes por el usuario y es lo que permite ver a pantalla trasera. En la pantalla trasera se dibuja una imagen formada por una combinación, con un patrón de líneas verticales, de las vistas correspondientes a cada ojo. Esta imagen es luego filtrada por la pantalla frontal permitiendo que llegue la vista correcta a cada ojo.

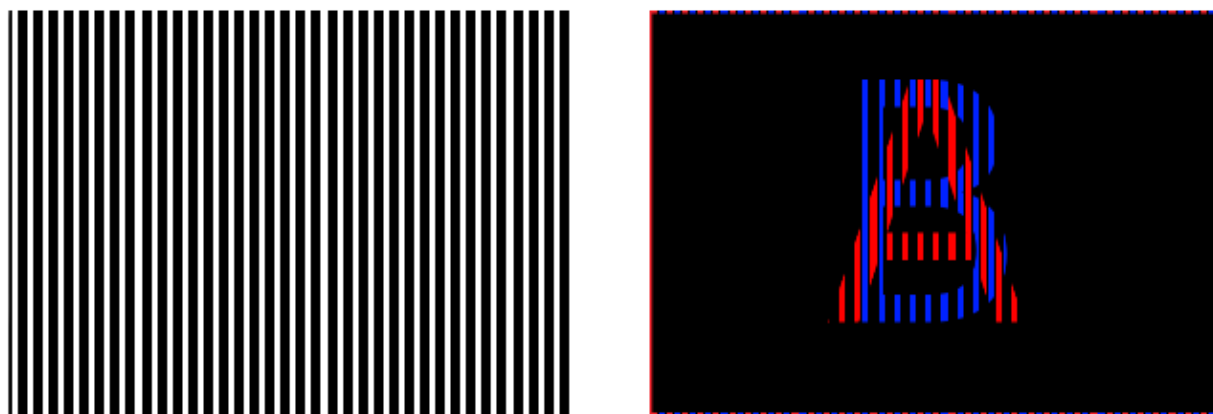


Figura 42. En la imagen de la izquierda se ve un ejemplo de la imagen dibujada en la pantalla frontal del prototipo. La imagen de la derecha es un ejemplo de la imagen dibujada en la pantalla trasera, en este caso corresponde a una letra A roja para la vista del ojo izquierdo y una letra B azul para la vista del ojo derecho.

El diseño del software se hizo pensando en facilitar la creación de aplicaciones para el prototipo. En base a esto, la lógica relacionada con los cálculos de la barrera de paralaje y la generación de las imágenes para las pantallas del dispositivo está centralizada en una biblioteca llamada *Parallax Barrier Library* (ParallaxBarrierLib).

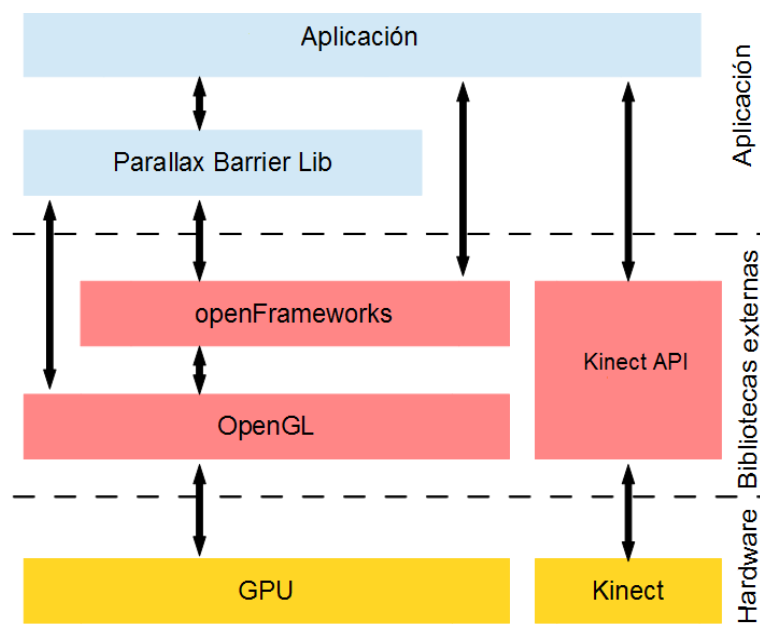


Figura 43. Diagrama de arquitectura del software.

PARALLAX BARRIER LIBRARY

Esta biblioteca fue desarrollada con el objetivo de encapsular la lógica correspondiente a la barrera de paralaje dinámica. Está basada en el flujo de trabajo de OpenFrameworks y permite crear fácilmente aplicaciones que utilicen el prototipo.

Las aplicaciones deben implementar las funciones del dibujado de las vistas y obtención de la posición de los ojos del usuario, luego, la biblioteca se encarga de procesar las imágenes generadas de las vistas según el algoritmo visto anteriormente para dibujar en cada una de las pantallas del prototipo las imágenes correspondientes.

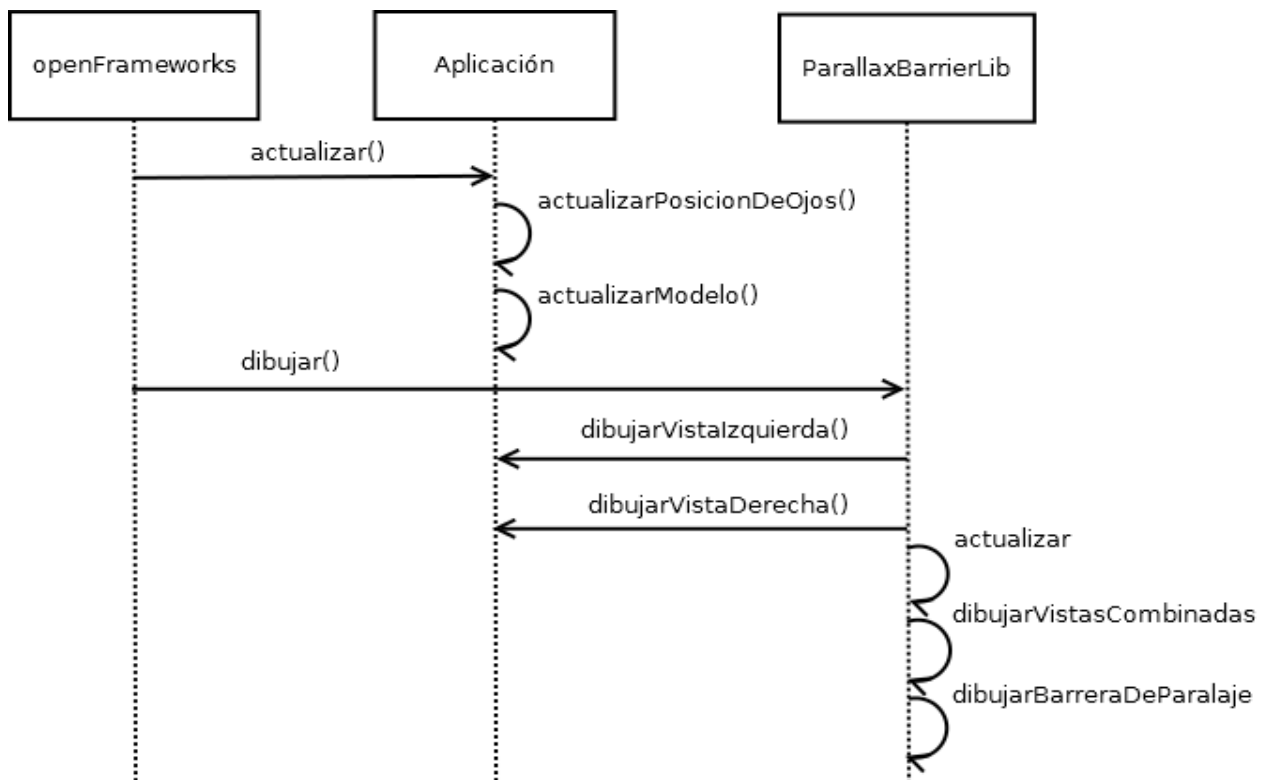


Figura 44. Diagrama de flujo de procesamiento de dibujo de un cuadro de la aplicación.

El dibujo de las vistas, hecho por la aplicación, es almacenado cada uno directamente en una textura de OpenGL. Para esto, ParallaxBarrierLib redirige lo dibujado a texturas mediante FrameBuffersObjects y no es dibujado directamente en pantalla.

El algoritmo de la solución matemática propuesta anteriormente simplifica las cuentas al suponer que la pantalla frontal se encuentra sobre la recta $y=1$, y la pantalla trasera se encuentra sobre el eje y . Es necesario realizar una transformación geométrica para realizar un cambio de coordenadas que ubique a las pantallas en estas posiciones.

La calibración que se realiza al comenzar a utilizar el prototipo permite ubicar al usuario y al monitor en coordenadas que tienen como origen a la posición de la cámara Kinect. La transformación por lo tanto debe trasladar, rotar y escalar de modo que en el nuevo sistema de coordenadas las pantalla se encuentren en las posiciones correctas. Esta transformación también debe aplicarse a la posición de los ojos del usuario para trabajar con toda la información en el mismo sistema de coordenadas.

Después de hecho el cambio de coordenadas se comienza con la actualización de las imágenes para dibujar en las pantallas del prototipo.

La imagen para la pantalla trasera se crea en base a las dos texturas correspondientes a las vistas de cada ojo. Como el procesamiento y combinación de las texturas de las vistas se debe hacer píxel por píxel, es muy costoso para ser ejecutado en el procesador de forma lineal. Además, es importante aprovechar que la información de las texturas ya se encuentra en la memoria de la tarjeta de video y evitar el pasaje de la información a la memoria del procesador. Por estos motivos, las texturas de las vistas son combinadas en la tarjeta de video a través de un *Fragment Shader* de OpenGL. El *Fragment Shader* es un programa del cual se ejecuta una instancia por cada píxel de forma paralela en la tarjeta de video. En este caso por cada píxel a ser dibujado en la pantalla se ejecuta el algoritmo de combinación de las vistas definiendo el color final del píxel a ser dibujado.

La imagen de la barrera para la pantalla frontal es creada también de forma paralela utilizando otro *Fragment Shader*. Basado en el algoritmo de la solución, calcula si cada píxel debe ser blanco o negro para formar correctamente la barrera del prototipo.

APLICACIÓN

La aplicación se encarga de hacer el seguimiento de la posición del usuario y de dibujar las vistas para cada ojo en base a la posición obtenida.

El seguimiento de la posición del usuario se hace en la aplicación para no condicionar a ParallaxBarrierLib a utilizar la cámara Kinect ni la API de Kinect para Windows. ParallaxBarrierLib recibe como parámetro la posición de los ojos del usuario desde la aplicación para los cálculos.

El seguimiento de la posición del usuario, y en particular de sus ojos, se realiza en un hilo en paralelo al del hilo principal de la aplicación. Esto permite que el rendimiento general de la aplicación no sea limitado por los tiempos de la cámara Kinect o los algoritmos de reconocimiento facial.

El acceso a los datos de la cámara Kinect se hace usando la API de Kinect para Windows de Microsoft. Ésta permite el acceso a la información de color y profundidad de la cámara, pero además puede hacer seguimiento del cuerpo del usuario y se integra con una API de seguimiento del rostro en tiempo real de alta precisión.

La API de seguimiento del rostro, una vez detectado un rostro, devuelve la posición de un conjunto de puntos preestablecidos. Particularmente para la aplicación se toman dos puntos por cada ojo que luego se promedian para obtener el centro del mismo.

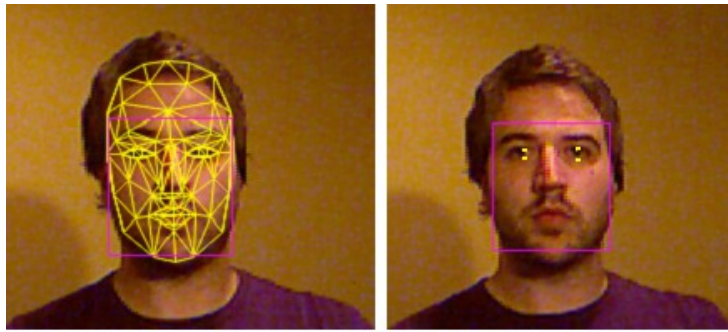


Figura 45. A la izquierda se ve la malla formada por el conjunto de puntos detectado por el seguimiento del rostro. A la derecha se muestran únicamente los puntos utilizados para ubicar los ojos.

La posición de la cámara es libre y se puede ubicar donde se precise según cada caso. Esto permite ubicarla por detrás o delante del dispositivo, pudiendo utilizar el prototipo a distancias menores o mayores al del rango de trabajo de la cámara.

Para poder ubicar libremente la cámara el software viene configurado, al iniciar, con una imagen de calibración que solamente es vista correctamente a 90cm por delante del monitor. El observador debe ubicarse en esta posición e indicarle al software que calibre. Al calibrar se relaciona la posición de la cámara con la posición del observador y la posición del observador con la posición del monitor. De esta manera se establece la posición de la cámara respecto del monitor y se pueden transformar las posiciones del observador relativas a la ubicación de la cámara a posiciones relativas a la ubicación del monitor.

Luego de la calibración el usuario es libre de moverse dentro del rango de la cámara.



Figura 46. Ubicado el observador en la posición de calibración correcta, la imagen de muestra una letra A roja para la vista del ojo izquierdo y una letra B azul para la vista del ojo derecho. Fuera de esa posición las imágenes no se muestran correctamente y se ven entrelazadas.

La aplicación es también responsable del dibujado de las vistas para cada ojo. Las vistas se crean por dibujado mediante instrucciones de OpenGL o funciones de OpenFrameworks

(que utilizan OpenGL), sin embargo estas vistas no son dibujadas directamente en pantalla sino que son procesadas por la biblioteca ParallaxBarrierLib para crear las imágenes correctas en el prototipo.

Se implementaron varios módulos en la aplicación sobre diferentes pruebas de concepto de usos del prototipo. Entre los módulos implementados se encuentran:

- Visualización de fotografías estereoscópicas

Se utilizan fotografías obtenidas con cámaras estereoscópicas para desplegar la vistas correspondientes a cada ojo, logrando así visualizar la tridimensionalidad de las escenas fotografiadas.

Las fotografías son cargadas desde archivos de imagen en formato PNG [48].

- Visualización de videos tridimensionales.

Los archivos de video en formato estereoscópico generalmente son almacenados como archivos de video regulares donde los videos para cada ojo son combinados lado a lado en un mismo video. Este módulo procesa los archivos en este formato, separando las vistas correspondientes a cada ojo, para luego desplegar el video autoestereoscópico correspondiente.

Los videos son cargados desde archivos en formato MP4 [40].

- Escenas tridimensionales generadas por computadora en tiempo real.

Este módulo utiliza modelos tridimensionales de objetos para armar una escena en tiempo real.

La escena es mostrada mediante el uso de dos cámaras para generar las vistas de cada ojo. Estas cámaras son ubicadas en la posición de los ojos del espectador utilizando la información proporcionada por la cámara Kinect. Combinando la posición de los ojos con la posición del dispositivo, se configuran las perspectivas de las cámaras de modo que al mirar en la pantalla simule estar mirando la escena a través de una ventana.

Los modelos tridimensionales son cargados a partir de archivos en formato COLLADA [6] y en la escena se presentan tanto objetos fijos como animados.

- Aplicación interactiva de dibujo en base a los movimientos de las manos del usuario. En este caso la cámara Kinect se usa también para detectar y seguir las manos del usuario.

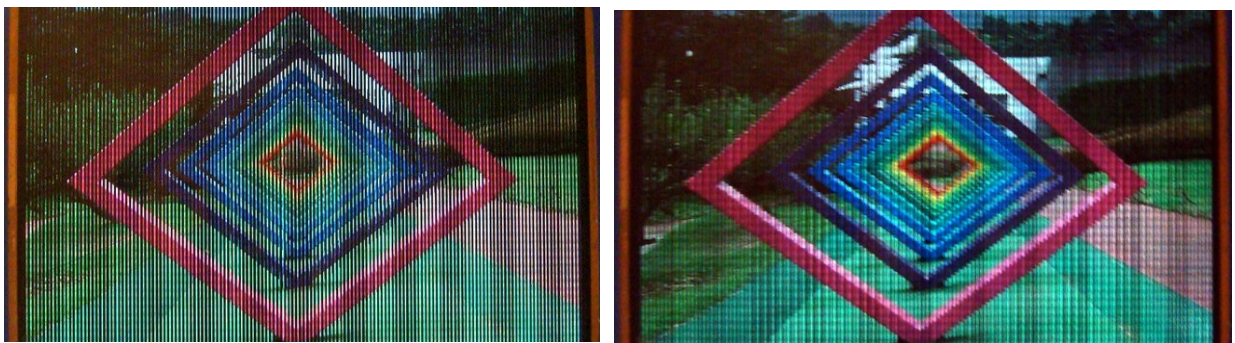
Este módulo permite al espectador dibujar en la pantalla con el movimiento de las manos. Al igual que en el módulo de la escena tridimensional de gráficos creados por computadora, se utilizan dos cámaras (una para la vista de cada ojo) que dan la sensación de estar mirando a través de una ventana.

Al mover una mano por delante del dispositivo se comienza a dibujar una cinta de color que sigue los movimientos de la mano en la pantalla. Esta cinta se desplaza alejándose del espectador y va cambiando su color con el paso del tiempo.

Ambas manos son detectadas y seguidas independientemente por la cámara Kinect, lo que permite dibujar libremente con cada mano combinando los diferentes colores en pantalla.



*Figura 47. Ejemplo de fotografías estereoscópicas mostradas por el prototipo.
En este caso una sola de las vistas es mostrada.*



*Figura 48. Vista de fotografía estereoscópica. La imagen de la izquierda
corresponde a la vista del ojo izquierdo y la imagen de la derecha
corresponde a la vista del ojo derecho.*

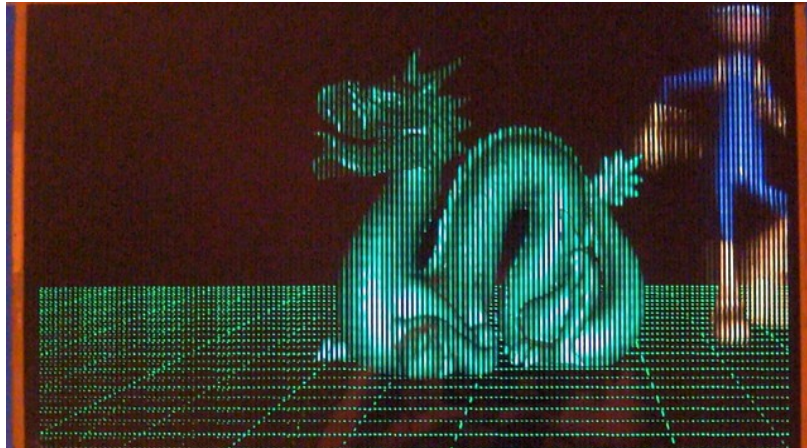


Figura 49. Imagen de contenido generado en tiempo real utilizando OpenGL por aplicación desarrollada para el prototipo.



Figura 50. El prototipo también puede reproducir archivos de video en formato estereoscópico.

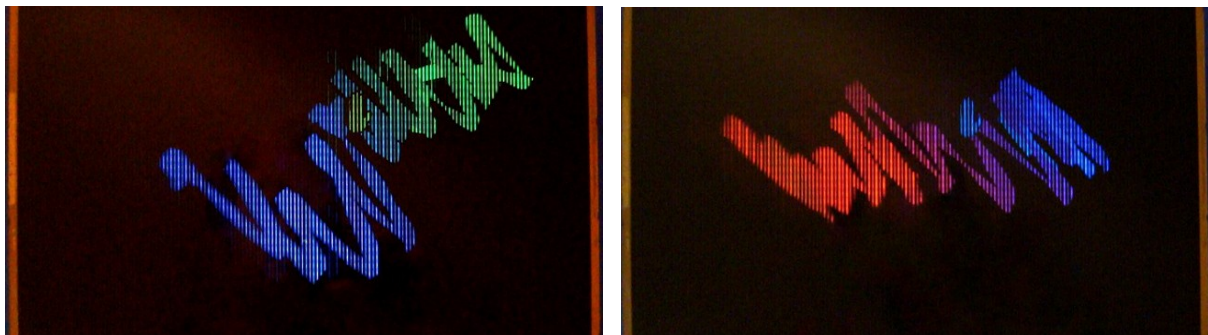


Figura 51. Imágenes de aplicación interactiva de dibujado. Las manos del usuario son seguidas mediante la información obtenida de la cámara Kinect para, basado en sus posiciones, ir dibujando cintas tridimensionales de colores en pantalla.

PROBLEMAS IDENTIFICADOS

Durante la construcción del prototipo se detectaron problemas que dificultaron el desarrollo y la evaluación de la efectividad de la técnica.

El primero, está relacionado a la precisión de la construcción del hardware del prototipo. Durante la construcción del hardware se introdujeron errores causados por fallas en el armado o falta de precisión en las mediciones.

Otro problema importante es el nivel de precisión del seguimiento de la posición de los ojos del usuario en tiempo real. La información de posición obtenida de la cámara Kinect y la biblioteca de reconocimiento facial son fuentes de introducción de error.

Los resultados de la técnica desarrollada son afectados considerablemente ante pequeños cambios en la información de entrada en la que se basa para hacer los cálculos. Estos problemas en las primeras etapas fueron una traba para el desarrollo y la verificación de la teoría de la técnica en la práctica. Para evitar estos problemas y poder trabajar en un ambiente controlado se creó una aplicación de simulación del prototipo para poder verificar el correcto funcionamiento de la técnica.

La aplicación de simulación utiliza el algoritmo desarrollado en ParallaxBarrierLib para crear la barrera de paralaje dinámica y dibuja en pantalla la vista desde arriba del dispositivo, basada en los resultados del algoritmo.

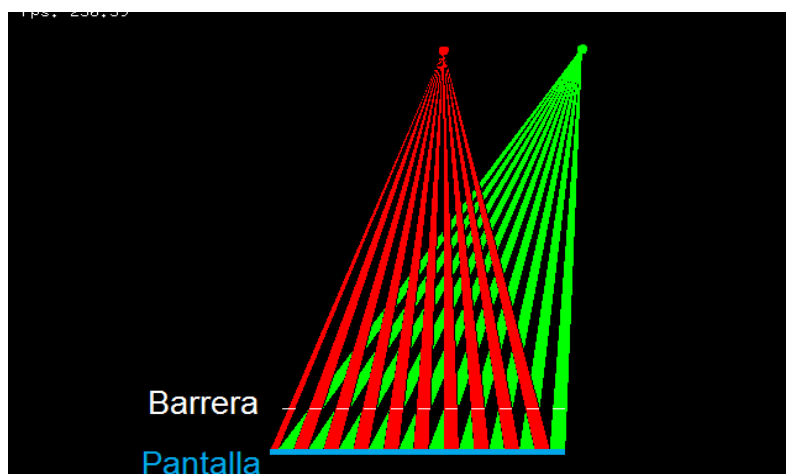


Figura 52. Captura de pantalla de la aplicación de simulación. Ésta muestra una vista de arriba del funcionamiento del dispositivo. El color rojo corresponde a la proyección de la imagen de la vista del ojo derecho desde la pantalla hasta el ojo derecho y el color verde a la proyección de la imagen de la vista del ojo izquierdo desde la pantalla hasta el ojo izquierdo. Cuando las proyecciones pasan por los espacios abiertos de la barrera (representados por las líneas blancas) el funcionamiento del dispositivo es correcto.

La simulación permite modificar diferentes parámetros que inciden en el algoritmo como son la posición de los ojos o la distancia entre la barrera y la pantalla. Luego, los resultados son dibujados y pueden ser evaluados visualmente de forma rápida.

Al asegurar el correcto funcionamiento de la técnica con la aplicación de simulación, se pudo detectar y corregir algunas de las fallas en las mediciones y construcción del hardware del prototipo.

5. CONCLUSIONES Y TRABAJO FUTURO

CONCLUSIONES

Se cumplió ampliamente con los objetivos fijados al principio del proyecto, los cuales eran hacer un relevamiento de las diferentes técnicas existentes hoy en día de visualización tridimensional y la construcción de un prototipo de dispositivo de visualización tridimensional que no requiera uso de aditamentos.

Respecto al estudio del estado del arte, se hizo énfasis en los últimos avances en las técnicas menos intrusivas para el espectador. Al investigar entre las técnicas relevadas para el desarrollo del prototipo se buscó principalmente que ésta diera libertad de movimiento y no requiera del uso de gafas o ningún otro elemento para su funcionamiento.

Como resultado de la investigación se llegó a la tecnología *Wedge* de Microsoft como primera opción para construir el dispositivo ya que cumple con los requerimientos establecidos. Sin embargo, debido a limitaciones del mercado local para construir este tipo de instrumento y restricciones de presupuesto fue descartado. Como segunda opción se trabajó en la creación de un barrera de paralaje generada dinámicamente basada en la posición del espectador. Para el seguimiento de la ubicación de los ojos del usuario se decidió el uso de una cámara Kinect que permite hacerlo de forma no intrusiva. La construcción del prototipo con esta técnica fue hecha basada en elementos fácilmente accesibles y de bajo costo en el mercado local como los son los monitores LCD.

El prototipo cumple con creces los objetivos establecidos y presenta varias ventajas a destacar.

En primer lugar, vale remarcar que es no intrusivo para el espectador y no requiere del uso de ningún elemento para el funcionamiento. Lo que facilita el uso y mejora la interacción con el usuario.

También permite visualizar contenido en tres dimensiones sin importar la ubicación del usuario, permitiendo que se traslade manteniendo la imagen actualizada. Esto se logra en base al seguimiento del usuario aprovechando los últimos avances en las herramientas disponibles para Kinect.

Además, la implementación del algoritmo permite generar el contenido para el monitor en tiempo real gracias al bajo costo de procesamiento requerido. Esto posibilita que sea utilizado para videojuegos o cualquier otro tipo de aplicación interactiva.

El bajo costo de su construcción es otra ventaja adicional, todos los elementos necesarios son de uso común y pueden ser adquiridos en el mercado local a precios accesibles.

Se crearon satisfactoriamente pruebas de concepto para validar las ventajas mencionadas y los posibles usos. Estas pruebas incluyeron reproducción de videos en formato tridimensional, generación en tiempo real de escenas tridimensionales, visualización de imágenes estereoscópicas y aplicaciones de interactivas utilizando Kinect para reconocimiento de gesticulación del usuario.

Como último punto vale destacar que todo el código desarrollado relacionado a la lógica de la técnica y está disponible como código libre [51] dando la posibilidad a que sea utilizado por el resto de la comunidad.

TRABAJO FUTURO

A lo largo del transcurso del proyecto, durante el estudio del estado del arte y el desarrollo del prototipo, se identificaron varios puntos a mejorar que dan lugar a nuevos trabajos a futuro.

Los monitores con los que fue construido el prototipo pueden ser reemplazados por monitores de alta frecuencia. Los monitores tridimensionales comerciales con gafas tienen frecuencias de refresco de 120Hz, el doble de lo normal. El uso de monitores de alta frecuencia permitiría probar una técnica para disminuir la percepción de baja resolución en el prototipo. En la implementación actual la pantalla trasera del prototipo dedica aproximadamente la mitad de sus píxeles, agrupados en líneas verticales, para la vista de un solo ojo lo que causa una reducción de la resolución horizontal a la mitad. Utilizando monitores de alta frecuencia se pueden alternar los patrones de la barrera de paralaje cuadro a cuadro manteniendo una frecuencia de refresco que no genera una imagen intermitente. La alternación de los patrones consiste en, para cada cuadro nuevo, invertir la asignación de los píxeles entre las vistas en la pantalla trasera e invertir en la pantalla frontal

la ubicación de las barreras. Esto permite que los píxeles negros vistos por un ojo en un instante al siguiente cuadro muestren la imagen de la vista, con una alta frecuencia de refresco este efecto puede lograr una percepción de resolución horizontal completa.

El prototipo tiene una distancia óptima de trabajo dentro del rango entre 0.6m y 1m. Fuera de este rango se pueden ver en algunos casos información cruzada entre las vistas. Esto es debido a falta de precisión en los datos obtenidos de la cámara Kinect, errores de medición y falta de precisión en la construcción del prototipo. La nueva versión de Kinect de Microsoft mejora la precisión de seguimiento y el tiempo de respuesta. El uso de esta nueva cámara mejoraría con poco esfuerzo la calidad del seguimiento del usuario. Otra alternativa para mejorar la precisión del seguimiento del usuario es el uso de varias cámaras Kinect. El uso de varias cámaras sirve también para ampliar el rango de movimiento del usuario y no estar limitado al espacio de funcionamiento de una sola cámara.

La API de Kinect para Windows solamente soporta el sistema operativo de Microsoft. Existen alternativas multiplataforma de código abierto, como OpenNI [45], que pueden reemplazar la biblioteca de Microsoft. Este cambio permitiría que el prototipo pueda correr sobre varios sistemas operativos.

El seguimiento de las manos en la aplicación de dibujo utilizando la cámara Kinect no es de alta precisión y en determinadas condiciones puede no lograr seguir el movimiento de las manos. Para mejorar el seguimiento de las manos se puede utilizar el dispositivo Leap Motion [65]. Esta herramienta permite seguir el movimiento de las manos con mayor precisión, alta estabilidad y menor tiempo de respuesta que la cámara Kinect.

Actualmente el prototipo de la barrera de paralaje dinámica soporta un sólo usuario. Un punto a mejorar es su modificación para que soporte varios espectadores al mismo tiempo. El cambio en el algoritmo no sería de gran magnitud, actualmente está hecho para un espectador (o dos vistas) y adicionar nuevos espectadores implicaría aumentar el número de vistas en el algoritmo. Vale destacar que cada vista nueva adicionada disminuye la resolución horizontal del prototipo proporcionalmente.

Otro punto a profundizar en futuros trabajos es el estudio técnicas de conversión de imágenes bidimensionales a tridimensionales y su integración con el software del dispositivo creado. Hoy en día existen varias técnicas de conversión automática de contenido en formato bidimensional a tridimensional [66] que, basándose en la generación de mapas de profundidad de imágenes o en la reconstrucción de la estructura tridimensional de las escenas, permitirían la generación de imágenes estereoscópicas para el dispositivo a partir de contenido bidimensional convencional.

Por último, sería valioso la creación de aplicaciones específicas que hagan provecho de la visualización de imágenes tridimensionales en el dispositivo creado. Las aplicaciones

interactivas, como ser de tele-presencia o videojuegos, pueden hacer uso de la tridimensionalidad de la imagen para mejorar la experiencia del usuario. También puede utilizarse para la visualización de datos complejos tales como en la fotogeología (estudio del terreno a partir de imágenes estereoscópicas aéreas) o en química para la representación de moléculas complejas. Otro campo de uso potencial es la medicina, donde se pueden crear aplicaciones médicas para visualizar radiografías estereoscópicas, visualización del interior del cuerpo humano a partir de tomografías asistidas por computador, entre otros.

6. BIBLIOGRAFÍA

- [1] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, y B. MacIntyre, Recent advances in augmented reality, IEEE Comput. Graph. Appl., 2001, Vol. 21, No. 6, pp. 34–47.
- [2] A. Berthier, Images stéréoscopiques de grand format, Cosmos 34, 1896, pp.227-233.
- [3] V. D. Blondel, N.-D. Ho y P. van Dooren, Weighted Nonnegative Matrix Factorization and Face Feature Extraction, Image and Vision Computing, 2008.
- [4] O. Bryngdahl, Moiré: Formation and interpretation, J. Opt. Soc. Am. 64, 1974, pp. 1287-1294
- [5] Camera Culture, MIT Media Lab. cameraculture.media.mit.edu (accedido Septiembre 2014).
- [6] COLLADA ISO Standard http://www.iso.org/iso/catalogue_detail.htm?csnumber=59902 (accedido Septiembre 2014).
- [7] ColorCode 3-D. www.colorcode3d.com (accedido Septiembre 2014).
- [8] DaLite. www.da-lite.com (accedido Septiembre 2014).
- [9] D. K. G. de Boer, M. G. H. Hiddink, M. Sluijter, O. H. Willemsen y S. T. de Zwart, Switchable lenticular based 2D/3D displays, Proc. SPIE 6490, 2007, Stereoscopic Displays and Virtual Reality Systems XIV, 64900R.
- [10] N. Dodgson, J. Moore, S. Lang, G. Martin y P. Canepa, A 50" Time-Multiplexed Autostereoscopic Display, Proc. SPIE 3957, Stereoscopic Displays and Virtual Reality Systems VII, 2000, pp. 177-183.
- [11] W. Epstein y S. S. Baratz, Relative size in isolation as a stimulus for relative perceived distance, Journal of Experimental Psychology 1964, Vol. 67, No. 6, pp. 507-513.
- [12] K. Erbey, Apparatus for the generation of a stereoscopic display, U.S. Patent US6476850, 2002.
- [13] J. M. Foley, Binocular distance perception, Psychological Review, Vol. 87(5), Sep 1980, pp. 411-434.

- [14] Fundamentals of Liquid Crystal Displays, Fujitsu. www.fujitsu.com/downloads/MICRO/fma/pdf/LCD_Backgrounder.pdf (accedido Septiembre 2014).
- [15] J. Gibson, The Ecological Approach to Visual Perception (ISBN 0-313-23961-4).
- [16] B. Gillam, The Perception of Spatial Layout from Static Optical Information en W. Epstein y S. Rogers, Perception of Space and Motion, New York: Academic Press , 1995, pp. 23-67.
- [17] E. B. Goldstein, Sensation and Perception (8° edición) (ISBN 0-495-60149-7) , pp. 230-248.
- [18] M. Heilig, Sensorama simulator, U.S. Patent US3050870 , 1962.
- [19] N. Holliman. 3D Display Systems. Technical report, Department of Computer Science, University of Durham, 2003.
- [20] HTC EVO 3D www.electronista.com/articles/11/03/22/htc.evo.3d.brings.dual.core.and.glasses.free.3d/ (accedido Septiembre 2014).
- [21] I. Ideses y L. Yaroslavsky, New methods to produce high quality color anaglyphs for 3-D visualization, Image Analysis and Recognition, Vol. 3212/2004, pp. 273–280.
- [22] INFITEC GmbH. www.infitec.net (accedido Septiembre 2014).
- [23] InterSense. InterSense IS-900 Precision Motion Tracker. www.intersense.com/pages/20/14 (accedido Septiembre 2014).
- [24] F. E. Ives, A novel stereogram, Journal of the Franklin Institute 153, 1902, pp. 51–52.
- [25] iZ3D. www.iz3d.com (accedido Septiembre 2014).
- [26] A. Jacobs, J. Mather, R. Winlow, D. Montgomery, G. Jones, M. Willis, M. Tillin, L. Hill, M. Khazova, H. Stevenson y G. Bourhill, 2D/3D Switchable Displays, Sharp Tech. J., 2003, Vol. 85, pp. 15–18.
- [27] H. Jorke, A. Simon, y M. Fritz, Advanced stereo projection using interference filters, J. Soc. Inf. Display, 2009, Vol. 17, No. 5, pp. 407-410.
- [28] B. Julesz, Foundations of cyclopean perception (1971) (ISBN 0-226-41527-9).
- [29] Y. Kim, J. Kim, J.-M. Kang, J.-H. Jung, H. Choi y B. Lee, Point light source integral imaging with improved resolution and viewing angle by the use of electrically movable pinhole array, Optics Express 15, 26, 2007, pp. 18253–18267.

- [30] D.-S. Kim, S. Shestak, K.-H. Cha, S.-M. Park y S.-D. Hwang, Time-sequential autostereoscopic OLED display with segmented scanning parallax barrier, Proc. SPIE 7329, Three-Dimensional Imaging, Visualization, and Display 2009, 73290U.
- [31] Kinect for Windows SDK, Microsoft. www.microsoft.com/en-us/kinectforwindowsdev/start.aspx (accedido Septiembre 2014).
- [32] D. Lanman, M. Hirsch, Y. Kim y R. Raskar, Content-adaptive parallax barriers: optimizing dual-layer 3D displays using low-rank light field factorization. ACM TRANS. GRAPH. 29, 6, Artículo 163, Diciembre 2010, 10 páginas.
- [33] LG Optimus 3D www.lg.com/es/telefonos-moviles/lg-P920-Optimus-3D (accedido Septiembre 2014).
- [34] LG Thrill www.lg.com/us/cell-phones/lg-P925-thrill-4g (accedido Septiembre 2014).
- [35] F. Matsuura y N. Fujisawa, Anaglyph stereo visualization by the use of a single image and depth information, J. Vis. 2008 Vol. 11, pp. 79–86.
- [36] Microsoft Kinect. www.microsoft.com/en-us/kinectforwindows/ (accedido Septiembre 2014).
- [37] Monitor AOC E1660SW. uy.aoc.com/monitor_displays/e1660sw (accedido Septiembre 2014).
- [38] Monitor Viewsonic VX1940w. www1.viewsonic.com/products/archive/vx1940w.htm (accedido Septiembre 2014).
- [39] R. Moseley, G. Woodgate y D. Ezra. Observer Tracking Directional Display, U.S. Patent US6377295, 2002.
- [40] MP4 (MPEG-4) ISO Standard http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=38538 (accedido Septiembre 2014).
- [41] Nintendo 3DS www.nintendo.com/3ds/hardware (accedido Septiembre 2014).
- [42] NVIDIA. www.nvidia.com (accedido Septiembre 2014).
- [43] OpenFrameworks. www.openframeworks.cc/ (accedido Septiembre 2014).
- [44] OpenGL. www.opengl.org/ (accedido Septiembre 2014).
- [45] OpenNI. structure.io/openni (accedido Septiembre 2014).

- [46] T. Peterka , R. L. Kooima , J. I. Girado , J. Ge , Daniel J. S , A. Johnson , J. Leigh , J. Schulze y T. A. Defanti, Dynallax: Solid State Dynamic Parallax Barrier Autostereoscopic VR Display, In Proceedings of IEEE Virtual Reality 2007, 2007, pp. 155-162.
- [47] T. Peterka, R. L. Kooima, D. J. Sandin, A. Johnson, J. Leigh y T. A. DeFanti, Advances in the Dynallax Solid-State Dynamic Parallax Barrier Autostereoscopic Visualization Display System, IEEE Transactions on Visualization and Computer Graphics, 2008, pp. 487-499.
- [48] PNG (Portable Networks Graphics) ISO Standard http://www.iso.org/iso/catalogue_detail.htm?csnumber=29581 (accedido Septiembre 2014).
- [49] Pureddepth. www.pureddepth.com (accedido Septiembre 2014).
- [50] RealD. www.reald.com (accedido Septiembre 2014).
- [51] Repositorio de código de ParallaxBarrierLib. github.com/szabalza/ParallaxBarrierLib (accedido Septiembre 2014).
- [52] O. Sacks, The Man Who Mistook His Wife For A Hat: And Other Clinical Tales (ISBN 0-684-85394-9), pp. 8-22.
- [53] K. M. Schreiber, D. B. Tweed y C. M. Schor, The extended horopter: quantifying retinal correspondence across changes of 3D eye position, Journal of Vision 6, 2006, art. 6.
- [54] Sharp Aquos www.sharp-phone.com/de-en/products/sh80f/ (accedido Septiembre 2014).
- [55] SHARP Electronics. www.sharppusa.com (accedido Septiembre 2014).
- [56] R. N. Shepard, Mind Sights: Original Visual Illusions, Ambiguities, and other Anomalies, New York: WH Freeman and Company, 1990.
- [57] Spatial View. www.spatialview.com/ (accedido Septiembre 2014).
- [58] A.R.L. Travis et al, Time-multiplexed three-dimensional video display, SID 95 DIGEST, 1995, p. 851.
- [59] A. Travis, N. Emerton , T. Large, y S. Bathiche, Collimated light from a waveguide for a display backlight, Optics Express 17, 2009, pp.19714-19719.
- [60] A. Travis, N. Emerton, T. Large y S. Bathiche, The design of backlights for view-sequential 3D. download.microsoft.com/download/D/2/E/D2E425F8-CF3C-4C71-A4A2-70F9D4081007/Backlightforviewsequentialautostereo.docx (accedido Septiembre 2014).

- [61] A. Travis, T. Large, N. Emerton y S. Bathiche, Wedge Optics in Flat Panel Displays, Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 1-16.
- [62] C. van Berkel y J. A. Clarke, Characterization and optimization of 3D-LCD module design, Proc. SPIE 3012, 1997, Stereoscopic Displays and Virtual Reality Systems IV, p. 179.
- [63] A. J. Woods y T. Rourke, Ghosting in anaglyphic stereoscopic images, Proc. SPIE 5291 Stereoscopic Displays and Virtual Reality Systems XI, Ed. 2004, pp. 354–365.
- [64] Laboratorio de Medios del Instituto de Computación de la Facultad de Ingeniería, Universidad de la República. www.fing.edu.uy/grupos/medialab/index.html (accedido Septiembre 2014).
- [65] Leap Motion <https://www.leapmotion.com/> (accedido Septiembre 2014).
- [66] L. Zhang, C. Vazquez y S. Knorr, 3D-TV Content Creation: Automatic 2D-to-3D Video Conversion, IEEE Transactions on Broadcasting, 2011, vol. 57, no. 2, pp. 372-383.