



UNIVERSITAT POLITÈCNICA  
DE CATALUNYA

## **Técnicas y algoritmos para la adquisición, transmisión y visualización de escenas 3D**

**Diana Beltrán Guerrero, Luis Basañez Villaluenga.**

*IOC- Divisió de Robòtica*

*IOC-DT-P-2008-12  
Setembre 2008*

**Institut d'Organització i Control  
de Sistemes Industrials**



# TÉCNICAS Y ALGORITMOS PARA LA ADQUISICIÓN, TRANSMISIÓN Y VISUALIZACIÓN DE ESCENAS 3D

Diana Beltrán Guerrero, Luis Basañez Villaluenga

Universidad Politécnica de Catalunya, Av Diagonal, 647, Planta 11, 08028, Barcelona, España

Instituto de Organización y Control de Sistemas Industriales

Julio, 2008

## ÍNDICE:

1. Objetivo	2
2. Introducción	2
3. Técnicas de Adquisición de Imágenes 3D	3
3.1 Visión Estereo	4
3.1.1 El Problema de Correspondencia	5
3.1.1.1. Técnicas de Correspondencia	6
3.1.1.2. Selección y Representación de Primitivas	7
3.1.1.3. Estrategias de Correspondencia	8
3.2 Visión Activa	8
3.2.1 Estrategias para Desarrollar Sistemas de Visión Activa	9
3.2.2. Sistemas de visión foveal	12
3.2.3 Visión Reconstructiva Frente A Visión selectiva	12
3.3 Técnicas de Luz Estructurada	13
3.4 Telemetría Laser	16
3.5 Holografía	16
3.5.1 La Holografía De Exhibición	19
3.5.2 La Holografía Como Instrumento De Medida	20
3.5.3 La Holografía Como Almacén De Información	20
3.5.4 La Holografía Como Dispositivo de Seguridad	20
4. Técnicas de Transmisión	21
4.1 H.261.	23
4.2 MPEG-1.	21
4.3 MPEG-2.	24
4.4 H.263.	24
4.5 MPEG-4.	25
4.5.1 Códecs MPEG-4	26
4.6 H.264/AVC.	27
4.7 MPEG-7	28

5. Técnicas De Visualización 3D	29
5.1 Visualización volumétrica	30
5.1.1 Dispositivos con volumen dinámico	31
5.1.2 Dispositivos de espejo oscilante	31
5.1.3 Dispositivos de espejo varifocal	31
5.1.4 Félix 3D	32
5.1.5 Perspecta Spatial 3D	33
5.2. Dispositivos con volumen estático	33
5.2.1. Helio Diplay	34
5.2.2. FogScreen	35
5.2.3. Monitores con varios LCDs	35
5.3. Autoestereoscopía	36
5.3.1. Estereogramas	37
5.3.2. Estereogramas holográficos	37
5.3.3. Dispositivos de barrera de paralaje	38
5.3.4. Lenticulares	39
5.3.5. El dispositivo de Holográfica	40
5.3.6. The Light Field Display	40
5.4. Estereoscopía	41
5.4.1 Anáglifos	41
5.4.2 ChromaDepth	42
5.4.3. Gafas de lentes polarizadas	43
5.4.4. Gafas de conmutación	43
6 Conclusiones	44
7 Anexo 1 (Estado del Arte)	46
8 Bibliografía	63

## 1. Objetivo

El objetivo principal de este trabajo es realizar un estudio sobre los diferentes métodos de adquisición, visualización y transmisión para obtener escenas 3d existentes hasta el momento.

## 2. Introducción

La visión artificial constituye uno de los temas de investigación que posee en la actualidad un espectro más amplio de posibles aplicaciones industriales, y que en un futuro adquirirá todavía una mayor relevancia. Muestra de ello son tanto los esfuerzos que dedican al tema los principales centros de investigación del mundo entero como el interés que demanda la industria en estas aplicaciones. La mayor parte de las realizaciones prácticas existentes, trabajan sobre imágenes bidimensionales, bien por manejar objetos planos, o bien por considerar que la información del objeto a analizar está suficientemente condensada en una o varias proyecciones. Esto supone una fuerte restricción en la gama de productos a analizar y en sus resultados. En la actualidad, el desarrollo de nuevas técnicas de procesamiento de imágenes, así como la espectacular evolución de los equipos informáticos, permite incluir la tercera dimensión como un objetivo adicional, permitiendo una adecuada adquisición y un correcto tratamiento de la información tridimensional de los objetos.

Las soluciones de la visión computacional para cada aplicación, pueden ser más o menos complicadas dependiendo de múltiples factores. Actualmente no existe, y probablemente no

existirá, una solución única general para todo problema que se plantee, debido a que cada situación a resolver tiene sus propias dificultades que hay que solventar. La forma de llegar a conseguir un sistema práctico que nos solucione el problema planteado consiste en aplicar aquellas técnicas, de entre todas las conocidas, que mejor se adecuen tanto en efectividad del reconocimiento como en velocidad.

El aspecto de la velocidad es en este caso muy importante ya que, prácticamente en cualquier aplicación industrial, se requiere un sistema que funcione en tiempo real. Para ello es necesario estudiar todos los métodos posibles y la forma de realizarlos con la menor carga computacional, es decir, el menor tiempo de procesado que nos permita acercarnos a una eficiencia óptima.

Actualmente la visión computacional, está en fase de crecimiento. Dada la enorme complejidad de ésta se van solucionando por etapas problemas cada vez más complicados. Actualmente las líneas de estudio e investigación se dividen en múltiples campos, desde la visión tridimensional, visión con moción, a la segmentación y agrupamiento de múltiples objetos diferentes en entornos no controlados, etc.

En este trabajo se presenta un estudio de los principales métodos de adquisición, transmisión y visualización de escenas 3D y se muestra los diferentes trabajos realizados los últimos años sobre estos temas.

### **3. TÉCNICAS DE ADQUISICION DE IMAGENES 3D**

El desarrollo de un sistema de Visión 3D requiere la resolución de una serie de aspectos ó etapas: recuperación de la estructura tridimensional de la escena, modelado y representación de objetos, reconocimiento y localización, y la interpretación de la escena.

Las diferentes técnicas existentes para la recuperación de la estructura tridimensional de la escena presentan características específicas en todos los niveles del proceso de interpretación visual Jiménez et al (1999), desde la etapa de formación de la imagen, análisis e interpretación de la misma, cada método requiere tanto de equipos como de algoritmos específicos. La bibliografía desarrollada en los últimos años ha permitido disponer de un conjunto de técnicas que se podrían clasificar en seis grupos:

- a. Visión estéreo: permite extraer la información tridimensional mediante la puesta en correspondencia de las informaciones bidimensionales procedentes de dos o más captadores de imagen.
- b. Cámara móvil: denominada en alguna bibliografía como técnicas de visión activa, permite extraer la información 3D a partir del flujo de imagen obtenido por un sensor, conocido el flujo de velocidades de la cámara.
- c. Técnicas de luz estructurada: son técnicas de visión activa en cuanto que modifican las condiciones del entorno. La distorsión producida por la proyección de patrones simples (rayos o planos), generados mediante luz coherente o luz láser, permite la extracción de la información tridimensional.
- d. Telemetría láser: permiten determinar el mapa de profundidad de la escena con base al tiempo transcurrido entre la emisión y detección de un pulso láser.
- e. Control de parámetros ópticos (Análisis enfoque/desenfoque): permiten determinar el mapa de profundidad de la escena a partir del nivel de enfoque en cada pixel de la escena.
- f. Holografía: es una técnica avanzada de fotografía, que consiste en crear imágenes tridimensionales. Para esto se utiliza un rayo láser, que graba microscópicamente una película fotosensible. Ésta, al recibir la luz desde la perspectiva adecuada, proyecta una imagen en tres dimensiones.

Cada una de estas técnicas se ha desarrollado tratando de resolver problemas concretos de la visión 3D y presentan características específicas en todos los niveles del proceso de interpretación visual, por ello gran parte de las investigaciones actuales se dirigen hacia la integración de la información obtenida mediante diferentes sensores o técnicas de extracción, de forma que el sistema pueda trabajar con entornos complejos y no estructurados, en los cuales cada técnica por separado presenta graves dificultades tanto algorítmicas como en tiempo de cálculo y precisión cuando se trabaja con entornos no controlados.

### 3.1 Visión Estéreo

El desarrollo de las técnicas de Visión Estéreo ha sido objeto de un gran esfuerzo de investigación en los últimos años. El objetivo de la Visión Estéreo es resolver dos problemas: el problema de correspondencia consistente en decidir para un punto del plano de imagen izquierdo, que punto en el plano de imagen derecho es correspondiente (son imágenes del mismo punto físico). El segundo problema es el problema de reconstrucción que trata de obtener, dados dos puntos correspondientes en ambos planos de imagen, las coordenadas 3D del punto en el espacio respecto a un sistema de coordenadas del mundo.

Dentro de este segundo problema subyace el problema de calibración del sistema visual que determina con qué aproximación conocemos las posiciones de los centros focales de cada sistema óptico, y la posición de los planos de imagen en el sistema de coordenadas del mundo. La solución del problema de la visión estereo debe asumir desde el primer momento que las posiciones de los puntos en la imagen y los parámetros de calibración del sistema visual son conocidos de forma imperfecta. Por lo tanto no se puede confiar en modelos geométricos rígidos del proceso de formación y reconstrucción de la imagen.

El uso de sistemas de visión estereoscópicos puede mejorar en gran medida el desempeño de la teleoperación y la satisfacción del usuario, y puede, de hecho, hacer posible la realización de tareas que de otro modo no podrían ser llevadas a cabo. Además ha quedado totalmente demostrado que a medida que la calidad de la imagen es menor o la complejidad de la escena aumenta, las ventajas de utilizar el sistema de visión de este tipo son mayores.

Algunos de los beneficios de la utilización de los sistemas de visión estereoscópicos son los siguientes:

- \* Mejora de la interpretación de la imagen.
- \* Filtro del ruido visual.
- \* Mejora de la calidad efectiva de la imagen.
- \* Campo efectivo de vista más ancho.
- \* Mejora de la detección de pendientes y depresiones.

Sumándose a estos beneficios se pueden añadir otros no tan explícitos:

- \* Una percepción más exacta de la profundidad.
- \* Una más rápida percepción de las relaciones espaciales en el mundo remoto.
- \* Una más completa información acerca de las relaciones espaciales en el mundo remoto.
- \* Mayor facilidad de uso para los novatos y reducción en el tiempo de entrenamiento y práctica.

Todas estas consideraciones demuestran, con total rotundidad, que el uso de un SISTEMA DE VISIÓN ESTEREOSCÓPICO reporta IMPORTANTES VENTAJAS que pueden compensar su mayor costo.

### 3.1.1 El Problema de Correspondencia.

El problema de correspondencia tal como se ha definido es en principio un problema indeterminado: dado un punto  $m_1$  en el plano de imagen  $R_1$  puede ser el correspondiente de cualquier punto  $m_2$  en el plano de imagen  $R_2$ . Para resolver esta dificultad se deben identificar un conjunto de restricciones que permitan reducir el número de correspondencias potenciales. Estas restricciones se pueden clasificar en cuatro tipos básicos:

#### a. Restricciones geométricas impuestas por el sistema de formación de imagen.

La más importante es la condición de epipolaridad por la cual los puntos correspondientes deben estar contenidos en la línea epipolar correspondiente Marr (1982). Esta restricción permite reducir el problema de una búsqueda bidimensional a uno unidimensional. Otras restricciones son las impuestas por la convergencia de los ejes focales de las cámaras que impone la condición de disparidad nula en el centro de la imagen.

#### b. Restricciones geométricas precedentes de los objetos visualizados.

Se han identificado un conjunto amplio de condiciones de este tipo que permiten de forma general resolver la ambigüedad del problema estéreo. La aplicación de estas condiciones depende en gran medida del tipo de escenas analizadas:

- Condición de unicidad: si no existen superficies transparentes la correspondencia para un punto  $m_1$  en el plano de imagen  $R_1$  debe ser única en el plano de imagen  $R_2$ .
- Condición de continuidad de superficie: La distancia al sistema de imagen varía suavemente dentro de las superficies, por lo tanto las disparidades deben variar también suavemente.
- Condición del gradiente de disparidad: restringe el máximo gradiente de disparidad permitido entre correspondencias de primitivas.
- Condición de continuidad figural: la disparidad varía continuamente a lo largo de los contornos de las superficies, pero puede cambiar abruptamente entre contornos.
- Condición de orden: la correspondencia de primitivas debe ocurrir en el mismo orden izquierda derecha a lo largo de las líneas epipolares. La inversión de orden puede ocurrir en ciertos entornos (p.e. objetos transparentes, objetos estrechos, etc.) pero son escasas en escenas naturales.
- Condición de posición general: ciertos eventos ocurren bastante infrecuentemente, en un sentido estadístico, de forma que permite desestimar correspondencias improbables.

#### c. Restricciones físicas.

Procedentes de modelos de como los objetos interactúan con la iluminación. La luz reflejada por una superficie es función del tipo de superficie, su orientación, la posición de los focos de luz y el punto de vista. Parten de un modelo de reflectancia de las superficies (generalmente un modelo Lambertiano) y un modelo de iluminación (difusa). El problema surge por el hecho de que las superficies reales no son Lambertianas y la iluminación nunca es perfectamente difusa.

#### *d. Restricciones locales de la primitiva.*

Estas condiciones determinan el grado de probabilidad de correspondencias de primitivas seleccionadas de la escena, como pueden ser signo y orientación del gradiente de intensidad, o condiciones asociadas a primitivas de mayor nivel como líneas o regiones que atesoran un conjunto más rico de características.

### **3.1.1.1. Técnicas de Correspondencia.**

Los algoritmos de reconstrucción estéreo dependen en gran medida del tipo de primitivas candidatas de la imagen. Medioni y Nevatia (1985) distinguen entre dos tipos de técnicas estéreo según la primitiva utilizada: las basadas en regiones, y las basadas en características dispersas de la escena.

#### *A. Técnicas basadas en regiones.*

Idealmente, el método más simple sería tratar de buscar correspondencias para cada pixel de la imagen, pero por un lado la información semántica almacenada en un pixel es muy escasa (nivel de intensidad) y por otro es el resultado de una discretización de un fenómeno continuo como es la proyección de la luminosidad reflejada por las superficies de la escena. En esta discretización subyace un error de precisión y posible enmascaramiento de proyecciones próximas. Por lo tanto para poder atacar el problema de correspondencia las técnicas basadas en regiones utilizan la información en un área o entono de vecindad de cada pixel. La obtención de una correspondencia se realizaría mediante la búsqueda en la otra imagen mediante una medida de correlación cruzada suma de las diferencias absolutas de intensidad u otras medidas similares. Este tipo de técnicas hacen uso de simplificaciones del espacio de búsqueda mediante la condición de epipolaridad y la utilización de estrategias multinivel «tosco-a-fino». Deriche y Cocquerez (1987) utilizan regiones, contornos cerrados sobre la imagen de intensidad luminosa. Estas regiones se segmentan mediante alguna característica de textura, tamaño, forma, etc. Estas técnicas plantean problemas de correspondencia de los puntos dentro de las regiones, pero facilitan la correspondencia entre regiones al ser entidades de mayor nivel, con más características diferenciadoras.

Los sistemas basados en correlación de regiones sufren de las siguientes limitaciones:

- a. Requieren la presencia de texturas dentro de cada ventana de correlación, por lo tanto tienden a comportarse mal ante entornos con falta de características o por la presencia de texturas repetitivas.
- b. Tienden a confundirse por la presencia de discontinuidades en superficies dentro de la ventana de correlación.
- c. Son sensibles a intensidades absolutas, contraste e iluminación.
- d. Tienen problemas con campos de profundidad que cambian rápidamente.

#### *B. Técnicas basadas en características.*

Estas técnicas se basan en la extracción de un conjunto de primitivas características de cada imagen para realizar la correspondencia entre estos datos a partir de las propiedades diferenciadoras. Las primitivas pueden ser píxeles que representen características especiales de la escena como son los píxeles de bordes, correspondientes a cambios bruscos de intensidad o los pasos por cero del operador Laplaciano. Este tipo de datos presentan unas características estables frente a variaciones fotométricas. Como características se utilizan el contraste del gradiente, orientación, radiancia. Un

paso adicional consiste en la agrupación de píxeles en primitivas de mayor nivel de abstracción. El caso más simple es la línea recta que es invariante a la proyección de perspectiva, esta representación es compacta, intrínseca, discriminable, precisa y densa. Esta representación permite modelar fácilmente modelos poliédricos, pero tiene problemas con contornos curvos, para los cuales se debe utilizar aproximaciones polinomiales de orden superior para preservar la compacidad de la descripción y discriminar mejor la característica. Aún así pueden presentar problemas adicionales debido a oclusiones parciales entre primitivas. Los segmentos de línea y curvas disponen de características geométricas y de intensidad incorporadas como longitud, orientación, curvatura, contraste medio, etc. Entre las ventajas de los sistemas basados en primitivas se encuentran:

\* Son más rápidos que los métodos basados en área, ya que se reduce el número de primitivas a considerar.

\* La correspondencia es más precisa ya que la primitiva puede ser localizada con precisión subpíxel.

\* Son menos sensibles a variaciones fotométricas, ya que representan propiedades de la escena. Por otra parte presentan el inconveniente de utilizar datos dispersos de la escena, por lo cual para obtener un mapa denso de profundidad de la escena será necesario interpolar los datos obtenidos en primitivas de superficie.

### **3.1.1.2. Selección y Representación de Primitivas.**

Respecto a la selección de las primitivas de la escena se pueden establecer un conjunto de propiedades que permitan evaluar su bondad e idoneidad para resolver el problema de correspondencia. Para ello se evalúan dos características básicas: la potencia descriptiva para resolver ambigüedades, y la invarianza ante pequeñas variaciones en el punto de vista. Para evaluar estas propiedades se puede identificar un conjunto de estimadores:

a. Dimensionalidad: Los puntos se ven menos afectados por distorsiones de perspectiva, pero las características tipo borde permiten encontrar correspondencias más fácilmente.

b. Tamaño: (frecuencia espacial) Para primitivas prominentes es más fácil encontrar la correspondencia.

c. Contraste: primitivas de bajo contraste son difíciles de distinguir.

d. Contenido semántico: que permita resolver falsas correspondencias.

e. Densidad de ocurrencia: (disperso frente a denso). Primitivas dispersas son más fáciles de corresponder pero generan sólo disparidades dispersas.

f. Localización y estabilidad del detector de primitivas. El detector debe localizar la primitiva en la posición real correspondiente a la proyección de la primitiva tridimensional, y esta debe ser estable entre las diferentes imágenes.

g. Distinguibilidad/Unicidad. La primitiva debe ser distinguible en ambas imágenes y cada característica de la escena debe generar una única primitiva en cada imagen. Ejemplos de utilización de primitivas incluyen la utilización de intensidades de pixel, signo del filtro Laplaciana-Gausiana, zero-crossings de la Laplaciana-Gaussiana, elementos de borde, segmentos enlazados de borde. Además de obtener las características (contornos o regiones), es necesario utilizar una representación estructurada de esta información.



### 3.1.1.3. Estrategias de Correspondencia.

Independientemente de la técnica de correspondencia utilizada se han desarrollado un conjunto de estrategias cuyo objetivo es reducir la ambigüedad y mejorar la correspondencia estéreo. Entre estos métodos se pueden citar:

- \* Estrategias tosco-a-fino. Son estrategias multinivel que utiliza la información obtenida en una escala tosca que se utiliza para guiar y limitar la búsqueda de correspondencia de las primitivas en la escala fina. En esta aproximación la correspondencia se inicia en la escala grosera donde el número de primitivas es muy inferior, lo cual disminuye el espacio de búsqueda. Pero su precisión es muy pequeña.
- \* Estrategias estructurales. Utilizan relaciones estructurales entre las primitivas, dependen en general del tipo de escena a analizar.
- \* Estrategias jerárquicas. Estos métodos utilizan diferentes niveles de descripción de la imagen formados por primitivas de diferente nivel de abstracción. En este caso la reducción de la complejidad de la correspondencia se logra mediante la utilización de estructuras de más alto nivel que sirven para decidir correspondencias en un nivel inferior.
- \* Integración, de las diferentes fases de extracción de primitivas, correspondencia e interpolación de superficies.
- \* Detección y corrección de errores en las disparidades.
- \* Estrategias híbridas regiones-características.
- \* Estrategias iterativas de predicción-propagación. Estos métodos realizan una primera predicción de correspondencias sobre algunos buenos candidatos y tratan de propagarlas a sus vecinos mediante grafos de disparidad.

Una parte importante de los algoritmos de correspondencia suele ser la técnica de "relajación", que permite reducir el número de correspondencias hipotéticas, disminuyendo el cálculo computacional necesario.

Este algoritmo se inicializa con todas las correspondencias que satisfagan la condición epipolar. Se asocia un peso a cada correspondencia (modelo neuronal) y se utiliza un mecanismo iterativo: en cada iteración y para cada correspondencia, un proceso "INHIBITORIO" reduce los pesos de correspondencias conflictivas (aquellas que violan la condición de unicidad con la correspondencia actual) y un proceso "EXCITATORIO" incrementa los pesos de las correspondencias con disparidades similares (condición de continuidad). Las correspondencias con pesos insuficientes son eliminadas.

## 3.2 Visión Activa

La mayoría de los sistemas de visión por computador consideran la Visión como un proceso reconstructivo que aborda el problema creando representaciones a diferentes niveles con un grado de abstracción progresivamente más alto. Según Marr (1982) el proceso de la visión consiste en descubrir mediante imágenes qué es lo que hay presente en una escena y dónde está ubicado. Hay que resaltar que esta definición no capta el hecho de que la visión, habitualmente, es usada para guiar nuestras actividades, es decir, los procesos de adquisición y extracción de la información

visual dependen en gran medida de la tarea a realizar y sirven para interactuar con el entorno de una manera intencionada y activa: dirigimos la mirada y atendemos a eventos selectivamente en función de nuestras necesidades. El mayor problema que presentan los sistemas de visión artificial surge de la enorme cantidad de datos contenidos en las imágenes convencionales muestreadas uniformemente, donde la resolución es constante en todo el campo visual; este volumen de datos hace prácticamente imposible que tales sistemas puedan procesar en tiempo real toda la información de dichas imágenes. Con el fin de acelerar y simplificar el proceso de comprensión de una escena, en los últimos años ha surgido una tendencia dentro del campo de la visión por computador denominada Visión Activa, que tiene como objetivo el estudio y desarrollo de sistemas de percepción en tiempo real, que integrados en un robot permitan que este interactúe en un mundo complejo y dinámico. El paradigma de la visión activa toma como punto de partida la idea de que la visión no se debe llevar a cabo de manera aislada, sino que debe servir para algún propósito y, en particular, debería permitir a la gente conocer el entorno. Los precursores de este paradigma fueron Bajcsy (1988), Aloimonos (1990) y Ballard (1992), que propusieron, inspirándose en los sistemas biológicos, actuar de forma activa en el proceso de percepción, o sea, controlar los parámetros de la cámara y los módulos de procesamiento, para simplificar la comprensión de las escenas. La visión activa es definida por Bajcsy como un proceso inteligente de adquisición de información, consistente en elaborar estrategias de control para ajustar los parámetros del sensor de cara a mejorar el conocimiento que se tiene del entorno. Por otro lado, mientras Aloimonos se fijó en la intencionalidad del agente, Ballard hizo hincapié en las conductas de éste y sugirió que un sistema de visión de propósito general puede ser implementado a partir de la combinación de múltiples y simples conductas.

Otra forma de simplificar el proceso de la visión, y a la vez aumentar la eficiencia de procesamiento, consiste en extraer y procesar únicamente los datos que son útiles para una tarea determinada; en este sentido, surge el concepto de percepción selectiva, basado en procesar la escena sólo donde sea necesario, con el nivel de detalle necesario y sólo con las operaciones necesarias; es decir, el sistema no requiere comprender completamente la escena, sino que debe ser capaz de detectar y reconocer sólo ciertas situaciones y patrones que son precisas para efectuar una actividad visual. Dentro del campo de la percepción selectiva cabe distinguir aquellos sistemas que emplean sensores que realizan un muestreo no uniforme del campo de visión, imitando así la distribución de los campos receptivos de la retina humana y dando lugar al concepto de sensores foveales.

### **3.2.1 Estrategias para Desarrollar Sistemas de Visión Activa**

A partir de lo indicado anteriormente se deduce que para desarrollar un sistema de visión activa pueden aplicarse diferentes estrategias Arrebola F. (2000): a) controlar los parámetros de la cámara, b) efectuar una reducción selectiva de los datos a procesar mediante mecanismos atencionales, c) emplear sensores foveales y d) explotar en conjunto más de una de las técnicas anteriores. A continuación se describen en detalle cada una de estas estrategias.

#### **Control de los parámetros del sistema de visión**

De igual forma que nuestro sistema visual se adapta a la actividad que estamos realizando y posee la capacidad de enfocar objetos o adecuarse a las condiciones de iluminación, se pueden controlar ciertos parámetros de la cámara para adaptar las imágenes a la tarea a realizar. Por ejemplo, cuando paseamos, nuestra mirada, cíclica y sucesivamente cambia: se dirige hacia un punto lejano para controlar la dirección que llevamos y, a continuación, se dirige y enfoca hacia el suelo para detectar obstáculos.

Entre los parámetros a controlar tenemos: el movimiento y posición de la cámara, el foco y apertura de la lente, el zoom, el ángulo de vergencia, la resolución y las condiciones de iluminación. Los

sistemas de visión activa, a medida que controlan tales parámetros, van adquiriendo múltiples vistas de la escena y extraen de cada una pocas características con algoritmos simples. Para determinadas aplicaciones de robots guiados visualmente, este modo de operación es mucho más eficiente y robusto que realizar una única observación y, a continuación, aplicar numerosos y complejos algoritmos. Sin embargo, como consecuencia de efectuar el control de los parámetros del sistema de visión surge la necesidad de introducir un subsistema capaz de planificar el conjunto de observaciones: el subsistema de Control de la Mirada será el encargado de implementar las estrategias de percepción, y puede, a su vez, estar compuesto por dos módulos: uno para efectuar las fijaciones -estabilización de la mirada-, y otro para realizar los cambios de la mirada o sacadas. Se entiende por sacadas los reajustes rápidos de la mirada, para atender a eventos de interés, que realizan los sistemas de visión biológicos.

#### Reducción selectiva mediante mecanismos atencionales

Partiendo de que la señal que entrega una cámara proporciona un volumen enorme de información -formas, color, movimiento, etc.-, y teniendo en cuenta que tales datos son extremadamente redundantes, esta estrategia trata de explotar el hecho de que para realizar una tarea visual determinada, únicamente una pequeña fracción de dicha información es relevante; la dificultad principal radica en determinar qué parte de la información es la de utilidad y cuál se puede obviar para cada tarea. Los sistemas de visión activa, emulando a los sistemas de visión biológicos, disponen de la capacidad de detectar rápidamente las zonas más importantes de las entradas visuales, para ello, suelen ejecutar algoritmos de reducción selectiva en profundidad -mediante mecanismos estéreo o técnicas de enfoque que efectúan una segmentación o descomposición de la escena en función de la distancia- o en movimiento -de cara a realizar seguimiento de móviles y evitar obstáculos.

Una vez determinado el volumen de información relevante, el sistema debe dirigir sus recursos computacionales a tales zonas de interés, para lo cual deben existir bucles de realimentación, así como mecanismos de alerta y de detección de eventos relevantes.

Por otra parte, el sistema atencional debe también decidir dónde y qué mirar a continuación, o sea, determinar, para el instante siguiente, dónde y cómo dirigir nuestro foco de atención. Dicha tarea puede venir controlada no sólo en función de las características extraídas del procesamiento -movimiento, distancia, color, sino también en función de la tarea a realizar y del contexto: se pueden establecer restricciones tanto espaciales como de comportamiento que simplifican los procesos a aplicar, disminuyen considerablemente el tiempo de cómputo y, por tanto, mejoran en gran medida el comportamiento y eficiencia del sistema. Volviendo al ejemplo del paseo, un mecanismo atencional para seleccionar el punto de fijación, puede realizarse simplemente empleando consideraciones geométricas, midiendo en ángulo entre el eje óptico y el plano del suelo, con lo que se puede obviar el resto de la información entregada por la cámara. En resumen, la capacidad de sensación selectiva hace que el proceso de percepción sea tratable desde el punto de vista computacional.

#### Reducción selectiva mediante sensores con resolución variante en espacio

Con este tipo de sensores se pretende simular la retina de los sistemas de visión biológicos, que presentan, por un lado, una zona central donde la densidad de fotorreceptores es elevada, es decir, en dicha zona la resolución es máxima -región foveal-, y por otro, una zona donde la densidad de fotorreceptores decrece monótonamente hacia la periferia del campo de visión, o sea, la resolución disminuye a medida que nos alejamos del centro del campo visual. Combinando esta disposición no uniforme de los elementos sensores, junto con la capacidad para mover la cámara de manera controlada, es posible adquirir la información necesaria para efectuar una actividad sin tener que considerar todos los datos contenidos en las imágenes de resolución uniforme. Los sistemas de

visión foveal efectúan una reducción selectiva en espacio, resolución y tiempo, y después de efectuar cada sacada deben realizar dos tareas, asociadas a la dicotomía qué -analizar la información correspondiente a la fovea- y dónde -seleccionar en la región periférica el siguiente punto de fijación-; esta última tarea es la principal característica distintiva entre la visión convencional y la visión foveal.

La principal ventaja que ofrecen las imágenes foveales es que se puede disponer, con un reducido número de elementos sensores, de un elevado campo de visión con una resolución muy alta en el centro del mismo. Por contra, las dificultades que conlleva el utilizar este tipo de imágenes son:

\*Hay que dotar al sistema de mecanismos que permitan reubicar la zona de máxima resolución o fovea en las regiones de interés. Los movimientos oculomotores del ojo humano permiten posicionar la fovea en las regiones relevantes de la escena, lo cual compensa la falta de detalle en la periferia del campo visual.

\*Se deben desarrollar, por un lado, nuevas estructuras de datos para manipular tales imágenes y, por otro, nuevos algoritmos para extraer información de las zonas de baja resolución, de cara a guiar los movimientos sacádicos de la cámara.

Dentro de los sensores variantes en espacio, se pueden distinguir varios tipos según la estrategia de muestreo se corresponda a una topología logaritmo-polar o a una topología cartesiana exponencial.

En la logaritmo-polar, los campos receptivos están ubicados en círculos concéntricos, su tamaño crece linealmente desde el círculo más interno hasta la periferia y su posición sensor queda especificada por su radio y su orientación.

En la topología cartesiana exponencial, la resolución también es variable, y disminuye según nos alejamos del centro del campo de visión de forma exponencial. Sin embargo, a diferencia del caso anterior, el tamaño o dimensiones de los elementos sensores -denominados rexels para diferenciarlos de los pixels- siempre es un número entero con respecto a los de la fovea y, dentro de ésta o de cada anillo, la resolución o agudeza visual permanece constante. Las imágenes que responden a estas geometrías quedan definidas únicamente por dos parámetros,  $m$  o número de anillos o niveles de resolución que rodean a la fovea, y de número de subanillos en cada nivel de resolución.

La estrategia de muestreo cartesiana conlleva una serie de ventajas con respecto a la logaritmo-polar, pues en cada nivel de resolución pueden aplicarse algoritmos convencionales de análisis y proceso de imagen, incluso la imagen completa puede ser procesada fácilmente mediante técnicas multiresolución previamente desarrolladas para estructuras piramidales, además, por la propia naturaleza de la estrategia de muestreo, la implementación VLSI sería menos compleja. Desde el punto de vista de la implementación, Generalmente los sistemas de visión artificial, desde los empleados para inspección industrial hasta los usados por satélites, captan el entorno con una resolución espacial constante, o sea, ésta es idéntica en todos los puntos del campo visual y, por tanto, no consideran la posible relevancia de alguna región de éste. Además, la cadena de procesamiento (data path) del sistema, típicamente está diseñada sin que exista realimentación entre las diferentes etapas, lo que provoca que las etapas de preproceso trabajen sin tener en cuenta el contexto de alto nivel y, como consecuencia, se procesan datos irrelevantes que pueden reducir la capacidad de respuesta del sistema. Por el contrario, los sistemas biológicos perciben el entorno con una agudeza visual variable: nuestra retina presenta una pequeña región central, denominada fovea, donde la resolución espacial es muy elevada, pero a medida que nos alejamos de ella la resolución disminuye de manera progresiva (si la agudeza fuera constante e igual a la de la fovea, el número de fibras del nervio óptico crecería varios órdenes de magnitud y nuestro cerebro, al tener que procesar tanta información, respondería mucho más lento). En estos sistemas, la falta de detalle en la

percepción de la periferia se resuelve mediante la capacidad para saber dónde mirar y cómo mover la cabeza y los ojos para hacerlo. Ciertas aplicaciones guiadas visualmente están destinadas a operar en entornos dinámicos y exigen realizar múltiples tareas heterogéneas de forma paralela -detección y reconocimiento de objetos, evitar obstáculos, buscar y seguir cierto móvil e identificarlo, navegar inteligentemente etc.-. Tales aplicaciones requieren una elevada resolución tanto espacial -que incrementa la precisión de los procesos de identificación- como temporal -que facilita las tareas de seguimiento- y un amplio campo visual -que aumenta la probabilidad de detección de eventos-. El empleo de un sensor foveal constituye una solución de compromiso para lograr estos requisitos, sin que por ello se dispare el número de datos de las imágenes a procesar, cosa que ocurriría si el sensor es de resolución constante.

Por otra parte, estos sistemas están dotados de un control de la mirada inteligente, que proporciona caminos de realimentación entre las diferentes etapas de proceso, lo que permite aumentar la eficiencia del procesamiento (las primeras etapas pueden trabajar en función del contexto detectado en etapas posteriores) y, consecuentemente, se mejora aún más el tiempo de respuesta del sistema.

En cuanto a productos relativos a la visión foveal, cabe indicar que, desde sensores foveales hasta los dispositivos para realizar la generación de sacadas, pasando por algoritmos para implementar la atención visual, están en fase de construcción de prototipos. En este sentido cabe destacar: a) el sensor CCD logaritmo-polar desarrollado por el grupo LIRA de la Universidad de Génova en colaboración el Dpto. de Ingeniería Eléctrica de la Universidad de Pennsylvania; b) el sensor CCD cartesiano exponencial desarrollado por Amherst Systems y el Dpto. de Ingeniería Eléctrica de la Universidad de Buffalo en el estado de Nueva York; c) el filtro hardware generador de imágenes multiresolución con tamaño y posición de la fovea reconfigurables que está siendo desarrollado por el Dpto. de Tecnología Electrónica de la Universidad de Málaga.

### **3.2.2. Sistemas de visión foveal**

Los sensores foveales pueden obtenerse mediante diferentes procedimientos:

- a) Métodos ópticos: construyendo una lente específica, se puede lograr que al proyectar la señal luminosa sobre un CCD uniforme se obtenga la topología deseada.
- b) Técnicas VLSI: en este caso la lente es convencional pero el CCD presenta una distribución no uniforme.
- c) Filtros hardware: se emplea un hardware especial para convertir una señal de vídeo normal en una imagen foveal.
- d) Filtros software: a partir de las imágenes convencionales se aplica un algoritmo que transforma el perfil de resolución constante de éstas en el perfil de resolución deseado.

### **3.2.3. Visión Reconstructiva Frente a Visión Selectiva**

Cada una de las estrategias para implementar sistemas de percepción en tiempo real, expuestas anteriormente, no son excluyentes, sino todo lo contrario; los sistemas de visión activa actuales tratan de explotar e integrar en conjunto estas estrategias y, como consecuencia, implican unas características específicas en este tipo de sistemas:

\*Operación continua: el sistema está siempre en ejecución, el procesamiento se realiza de forma cíclica y depende de los resultados previamente obtenidos en fijaciones anteriores y de los eventos que surjan en el entorno donde actúa el sistema.

\* Filtrado de información: el sistema ha de comportarse como un filtro que determina, en cada momento, cuál es la información relevante a procesar para realizar una actividad.

\*Tiempo real: para que el sistema sea útil, ha de entregar resultados dentro de un intervalo fijo de tiempo, o sea, los recursos computacionales han de estar limitados.

\*Control del procesamiento: la exigencia de que la respuesta del sistema esté disponible en un intervalo de tiempo fijo, supone restringir el procesamiento masivo a pequeñas regiones de interés; para esto hay que aplicar previamente pasos atencionales.

### **3.3 Técnicas de Luz Estructurada**

Dentro del campo de la Visión tridimensional, existen un gran número de técnicas que hoy son empleadas con éxito en numerosas aplicaciones industriales. Entre todas ellas, se encuentra lo que se conoce como la Luz Estructurada.

Este tipo de sistema se caracteriza por ser un método directo y activo. Un método directo se caracteriza por que se pueden obtener conclusiones estudiando los datos obtenidos directamente de las imágenes. En cuanto a lo de activo, hay que tener en cuenta que este sistema emplea algún sistema generador de luz, por lo que introduce un tipo de energía al entorno donde se realiza el estudio.

Los sistemas de luz estructurada se basan en estudiar la deformación que sufre un patrón de luz al ser interseccionado por cualquier objeto. Este es el problema principal de este tipo de herramientas, ya que se necesita un tipo de luz concentrada en un punto. No valdría como sistema de iluminación, cualquiera de los sistemas normales que se emplean actualmente, como bombillas, fluorescentes, etc, ya que, están compuestos por ondas de diferentes frecuencias provocando que el haz se difumine por todo el entorno.

Una de las mejores soluciones es emplear un haz láser. Debido a sus características de coherencia, divergencia, y direccionalidad, ésta se comporta en una luz ideal para este tipo de sistemas. Esto es así debido a que la coherencia provoca que todas las ondas del haz tengan la misma frecuencia, y junto a su poca divergencia y alta direccionalidad permiten que se pueda dirigir un haz láser a cualquier punto que se quiera.

Una vez que ya se conoce el tipo de luz que se va a emplear, será necesario elegir un patrón adecuado. Las diferentes soluciones van desde la utilización de puntos, hasta rejillas con láseres de diferentes colores.

El emplear puntos implica tener que recorrer el objeto por toda su superficie tomando una gran cantidad de puntos y pudiendo perder algunas zonas de éste.

El empleo de un plano parece una mejor solución que la anterior. El plano iluminará un conjunto de puntos con las mismas características y que cumplen con la condición de un plano en el espacio. El uso de una rejilla con diferentes puntos de colores implica el tener toda la superficie del objeto iluminada de una vez y el problema consistiría en encontrar los diferentes puntos y su situación.

Por supuesto, además del patrón de luz, es necesario tener una cámara que recoja todas las imágenes de la deformación del plano láser. La posición de la cámara en el conjunto deberá ser aquella, que permita obtener, tanto la mejor resolución como evitar que existan zonas oscuras, es decir, que no existan zonas del objeto que no sean iluminadas por el láser.

Una vez que se ha decidido por uno u otro patrón y la correspondiente cámara, será necesario calibrar los diferentes parámetros. Este paso es el más importante, ya que para obtener las coordenadas de diferentes puntos en el objeto se va a emplear el método de triangulación.

La triangulación consiste en obtener la situación de un punto en relación a la situación de la cámara y el plano. La figura de abajo explica el método de la triangulación:

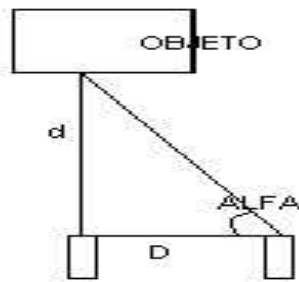


Figura 1: Método de triangulación

Si se conoce la distancia de la cámara a un punto del objeto, que será la base del triángulo, la distancia entre la cámara y el láser, uno de los lados, y el ángulo del plano del láser, se puede conocer las coordenadas tridimensionales de dicho punto.

Pero para ello, es necesario conocer las posiciones en el espacio tanto de la cámara como del plano láser, por lo que será necesario, un proceso de calibración de ambos sistemas.

Su principio de funcionamiento puede resumirse como sigue Granero L. (2002):

1. Se proyecta sobre el objeto en cuestión un patrón (o conjunto de patrones) cuya codificación es conocida.
2. Se realiza un análisis de las variaciones de este patrón respecto a la codificación original.
3. Se genera mediante algoritmos de cálculo y depuración la nube de puntos que posteriormente compondrán la estructura tridimensional (malla) del objeto.

Los patrones proyectados están diseñados de manera que a cada píxel de la imagen se le pueda asignar un código bien definido. Las técnicas de proyección de estos patrones pueden clasificarse en función de la estrategia de codificación utilizada, dividiéndose en tres grandes grupos:

- Técnica de Multiplexado en tiempo.

- Técnicas de Vecindad espacial.

- Codificación directa.

En nuestro caso consideraremos únicamente las llamadas técnicas de multiplexado en tiempo una de las más extendidas en los sistemas comerciales de digitalización tridimensional por luz estructurada.

Dentro de las técnicas de multiplexado temporal podemos encontrar a su vez tres tipos de estrategias y una combinada de proyección de patrones:

- Técnicas binarias: En este caso, se utilizan únicamente dos niveles de iluminación, codificados por 0 ó 1 (Fig. 1.1), cada píxel del patrón tiene su propia codificación compuesta por una secuencia de 0s u 1s. Este tipo de patrones son sencillos de codificar y de procesar.



Fig. 1.1 Proyección de patrones binarios

- Técnicas de N-ary códigos: Estas técnicas reducen el número de patrones a proyectar sobre la pieza, sin embargo, utilizan un mayor número de niveles de iluminación (niveles en escala de gris/color Fig. 1.2) incrementando así el nivel de información codificada en estos patrones.



Fig. 1.2. Proyección de patrones en escala de gris

- Técnicas de codificación en grises y desplazamiento de fase: La técnica consiste en la proyección de un conjunto de patrones binarios que producen una división del objeto en diferentes regiones. Además de este conjunto de patrones, se proyecta un patrón complementario que se desplaza con el fin de aumentar la resolución (Fig. 1.3).



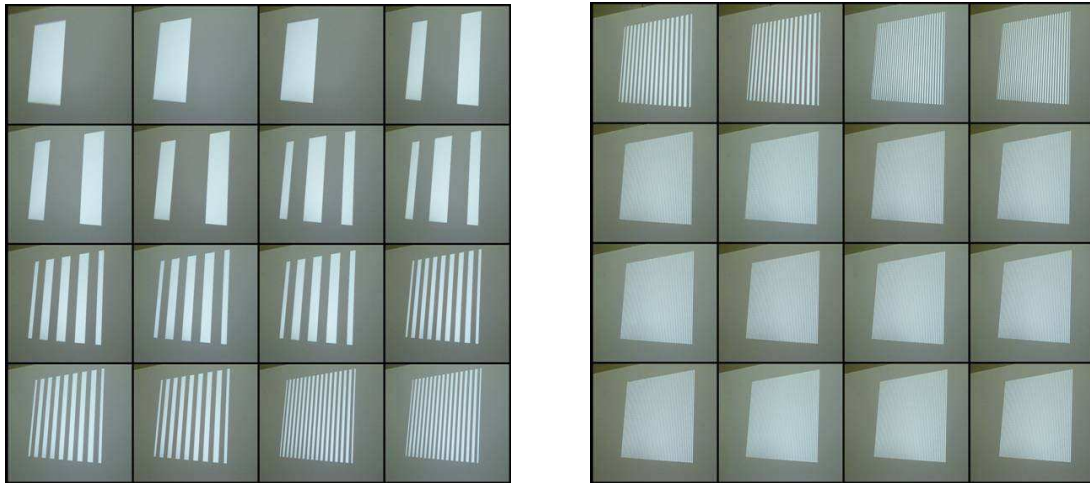


Fig. 1.3. Proyección simultánea de patrones en combinado

### 3.4 Telemetría Láser

La telemetría láser consiste en medir el tiempo de recorrido de un rayo luminoso (láser) hasta la superficie de medida. Se puede medir de dos formas: con la medida del tiempo de vuelo y el cálculo por diferencia de fase. En el primer caso los datos se obtienen midiendo el tiempo entre la emisión del impulso luminoso y la observación del retorno. En el segundo se regula el impulso luminoso siguiendo una frecuencia determinada y se mide el desfase entre el rayo emitido y la luz retornada.

### 3.5 Holografía

La holografía es una técnica especial de producción de fotografías tridimensionales de un objeto. El termino holograma fue acuñado por el inventor de la holografía, el científico húngaro Dennis Gabor, a partir de las palabras “grama” (mensaje), y “halos” (toda, completa). En realidad un holograma contiene más información sobre la forma de un objeto que una fotografía simple, ya que permite verla en relieve, y variando la posición del observador, obtener diferentes perspectivas del objeto holografiado. Dennis Gabor inventó la holografía en 1947, y recibió el premio Nóbel de Física en 1971 por su descubrimiento.

La holografía ha progresado de una manera impresionante y rápida debido a la gran cantidad de aplicaciones que se le están encontrando día a día. Los hologramas se pueden ahora hacer de muy diferentes maneras, pero todos con el mismo principio básico. Los principales tipos de hologramas son los siguientes:

a) Hologramas de Fresnel. Éstos son los hologramas más simples, también son los hologramas más reales e impresionantes, pero tienen el problema de que sólo pueden ser observados con la luz de un láser.

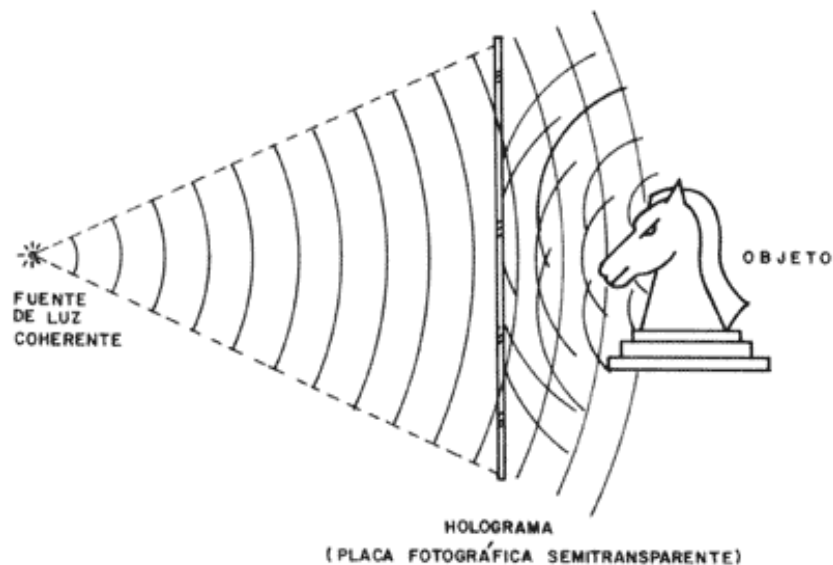


Figura 2. Formación de un holograma de reflexión.

b) Hologramas de reflexión. Los hologramas de reflexión, inventados por Y N. Denisyuk en la Unión Soviética, se diferencian de los de Fresnel en que el haz de referencia, a la hora de tomar el holograma, llega por detrás y no por el frente, como se muestra en la figura 2. La imagen de este tipo de hologramas tiene la enorme ventaja de que puede ser observada con una lámpara de tungsteno común y corriente. En cambio, durante la toma del holograma se requiere una gran estabilidad y ausencia de vibraciones, mucho mayor que con los hologramas de Fresnel. Este tipo de holograma tiene mucho en común con el método de fotografía a color por medio de capas de interferencia, inventado en Francia en 1891 por Gabriel Lippmann, y por el cual obtuvo el premio Nobel en 1908.

c) Hologramas de plano imagen. Un holograma de plano imagen es aquel en el que el objeto se coloca sobre el plano del holograma. Naturalmente, el objeto no está físicamente colocado en ese plano, pues esto no sería posible. La imagen real del objeto, formada a su vez por una lente, espejo u otro holograma, es la que se coloca en el plano de la placa fotográfica. Al igual que los hologramas de reflexión, éstos también se pueden observar con una fuente luminosa ordinaria, aunque sí es necesario láser para su exposición.

d) Hologramas de arco iris. Estos hologramas fueron inventados por Stephen Benton, de la Polaroid Corporation, en 1969. Con estos hologramas no solamente se reproduce la imagen del objeto deseado, sino que además se reproduce la imagen real de una rendija horizontal sobre los ojos del observador. A través de esta imagen de la rendija que aparece flotando en el aire se observa el objeto holografiado, como se muestra en la figura 3. Naturalmente, esta rendija hace que se pierda la tridimensionalidad de la imagen si los ojos se colocan sobre una línea vertical, es decir, si el observador está acostado. Ésta no es una desventaja, pues generalmente el observador no está en esta posición durante la observación. Una segunda condición durante la toma de este tipo de hologramas es que el haz de referencia no esté colocado a un lado, sino abajo del objeto.

Este arreglo tiene la gran ventaja de que la imagen se puede observar iluminando el holograma con la luz blanca de una lámpara incandescente común. Durante la reconstrucción se forma una multitud de rendijas frente a los ojos del observador, todas ellas horizontales y paralelas entre sí, pero de diferentes colores, cada color a diferente altura. Según la altura a la que coloque el observador sus ojos, será la imagen de la rendija a través de la cual se observe, y por lo tanto esto definirá el color de la imagen observada. A esto se debe el nombre de holograma de arco iris.

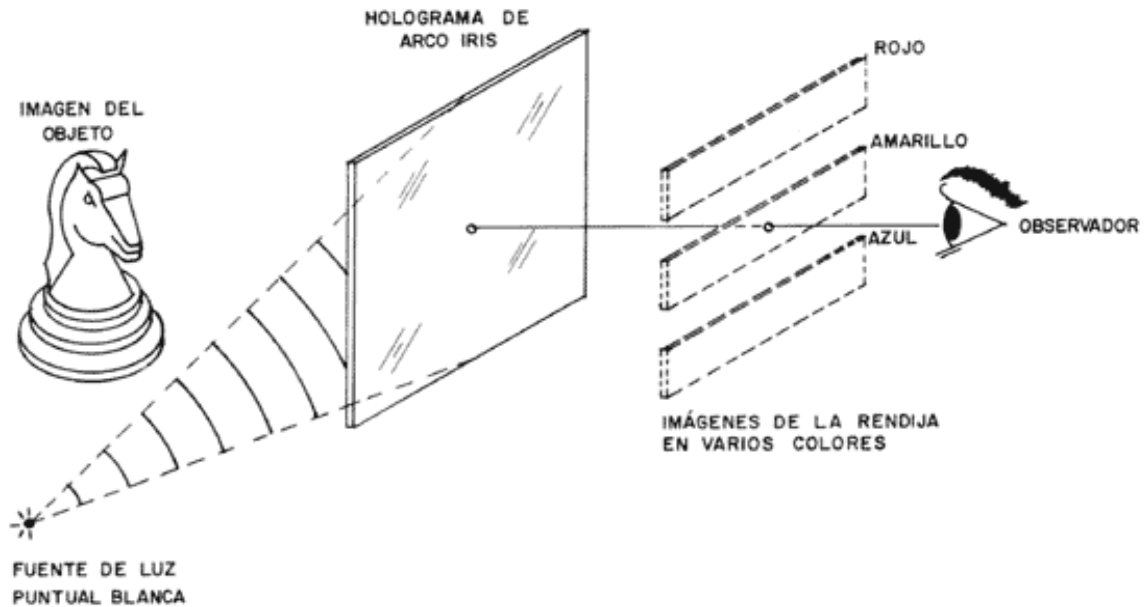


Figura 3. Formación de un holograma de arco iris.



Figura 4. Holograma prensado, fabricado por J. Tsujiuchi en Japón.

e) Hologramas de color. Si se usan varios láseres de diferentes colores tanto durante la exposición como durante la observación, se pueden lograr hologramas en color. Desgraciadamente, las técnicas usadas para llevar a cabo estos hologramas son complicadas y caras. Además, la fidelidad de los colores no es muy alta.

f) Hologramas prensados. Estos hologramas son generalmente de plano imagen o de arco iris, a fin de hacerlos observables con luz blanca ordinaria. Sin embargo, el proceso para obtenerlos es diferente. En lugar de registrarlos sobre una placa fotográfica, se usa una capa de una resina fotosensible, llamada Fotoresist, depositada sobre una placa de vidrio. Con la exposición a la luz, la placa fotográfica se ennegrece. En cambio, la capa de Fotoresist se adelgaza en esos puntos. Este adelgazamiento, sin embargo, es suficiente para difractar la luz y poder producir la imagen. Dicho de otro modo, la información en el holograma no queda grabada como un Sistema de franjas de

interferencia obscuras, sino como un sistema de surcos microscópicos. La figura 4 muestra un holograma prensado.

El siguiente paso es recubrir el holograma de Fotorresist, mediante un proceso químico o por evaporación, de un metal, generalmente níquel. A continuación se separa el holograma, para que quede solamente la película metálica, con el holograma grabado en ella. El paso final es mediante un prensado con calor: imprimir este holograma grabado en la superficie del metal, sobre una película de plástico transparente. Este plástico es el holograma final.

Este proceso tiene la enorme ventaja de ser adecuado para producción de hologramas en muy grandes cantidades, pues una sola película metálica es suficiente para prensar miles de hologramas. Este tipo de hologramas es muy caro si se hace en pequeñas cantidades, pero es sumamente barato en grandes producciones.

g) Hologramas de computadora. Las franjas de interferencia que se obtienen con cualquier objeto imaginario o real se pueden calcular mediante una computadora. Una vez calculadas estas franjas, se pueden mostrar en una pantalla y luego fotografiar. Esta fotografía sería un holograma sintético. Tiene la gran desventaja de que no es fácil representar objetos muy complicados con detalle. En cambio, la gran ventaja es que se puede representar cualquier objeto imaginario. Esta técnica se usa mucho para generar frentes de onda de una forma cualquiera, con alta precisión. Esto es muy útil en interferometría.

### **3.5.1 La Holografía de Exhibición**

Ésta es la aplicación más frecuente y popular de la holografía. Es muy conocida, por ejemplo, la exhibición que hizo una famosa joyería de la Quinta Avenida de Nueva York, donde por medio de un holograma sobre el vidrio de un escaparate se proyectaba hacia la calle la imagen tridimensional de una mano femenina, mostrando un collar de esmeraldas. La imagen era tan real que provocó la admiración de muchísimas personas, e incluso temor en algunas. Se dice que una anciana, al ver la imagen, se atemorizó tanto que comenzó a tratar de golpear la mano con su bastón, pero al no lograrlo, corrió despavorida.

Una aplicación que se ha mencionado mucho es la de la exhibición de piezas arqueológicas o de mucho valor en museos. Esto se puede lograr con tanto realismo que sólo un experto podría distinguir la diferencia.

Otra aplicación que se ha explorado es la generación de imágenes médicas tridimensionales, que no pueden ser observadas de otra manera. Como ejemplo, solamente describiremos ahora el trabajo desarrollado en Japón por el doctor Jumpei Tsujiuchi. El primer paso en este trabajo fue obtener una serie de imágenes de rayos X de una cabeza de una persona viva. Estas imágenes estaban tomadas desde muchas direcciones, al igual que se hace al tomar una tomografía. Todas estas imágenes se sintetizaron en un holograma, mediante un método óptico que no describiremos aquí. El resultado fue un holograma que al ser iluminado con una lámpara ordinaria producía una imagen tridimensional del interior del cráneo. Esta imagen cubre 360 grados, pues el holograma tiene forma cilíndrica. El observador podía moverse alrededor del holograma para observar cualquier detalle que desee. La imagen es realmente impresionante si se considera que se está viendo el interior del cráneo de una persona viva, que obviamente puede ser el mismo observador.

Otra aplicación natural es la obtención de la imagen tridimensional de una persona. Esto se ha hecho ya con tanto realismo que la imagen es increíblemente natural y bella. Sin duda ésta es la fotografía del futuro. Lamentablemente, por el momento es tan alto el costo, sobre todo por el equipo que se requiere, que no se ha podido comercializar y hacer popular.

Se podrían mencionar muchas más aplicaciones de la holografía de exhibición, pero los ejemplos anteriores son suficientes para dar una idea de las posibilidades, que cada día se explotan más.

### **3.5.2 La Holografía como Instrumento de Medida**

La holografía es también un instrumento muy útil, asociado con la interferometría, para efectuar medidas sumamente precisas.

La utilidad de la holografía proviene del hecho de que mediante ella es posible reconstruir un frente de onda de cualquier forma que se desee, para posteriormente compararlo con otro frente de onda generado en algún momento posterior. De esta manera es posible observar si el frente de onda original es idéntico al que se produjo después, o bien si tuvo algún cambio. Esto permite determinar las deformaciones de cualquier objeto con una gran exactitud, aunque los cambios sean tan pequeños como la longitud de onda de la luz. Para ilustrar esto con algunos ejemplos, mencionaremos los siguientes:

- a) Deformaciones muy pequeñas en objetos sujetos a tensiones o presiones. Mediante holografía interferométrica ha sido posible determinar y medir las deformaciones de objetos sujetos a tensiones o presiones. Por ejemplo, las deformaciones de una máquina, de un gran espejo de telescopio o de cualquier otro aparato se pueden evaluar con la holografía.
- b) Deformaciones muy pequeñas en objetos sujetos a calentamiento. De manera idéntica a las deformaciones producidas mecánicamente, se pueden evaluar las deformaciones producidas por pequeños calentamientos. Ejemplo de esto es el examen de posibles zonas calientes en circuitos impresos en operación, en partes de maquinaria en operación, y muchos más.
- c) Determinación de la forma de superficies ópticas de alta calidad. Como ya se ha comentado antes, la unión de la interferometría con el láser y las técnicas holográficas les da un nuevo vigor y poder a los métodos interferométricos para medir la calidad de superficies ópticas.

### **3.5.3. La Holografía como Almacén de Información**

La holografía también es útil para almacenar información. Esta se puede registrar como la dirección del rayo que sale del holograma, donde diferentes direcciones corresponderían a diferentes valores numéricos o lógicos. Esto es particularmente útil, ya que existen materiales holográficos que se pueden grabar y borrar a voluntad, de forma muy rápida y sencilla. Con el tiempo, cuando se resuelvan algunos problemas prácticos que no se ven ahora como muy complicados, será sin duda posible sustituir las memorias magnéticas o las de estado sólido que se usan ahora en las computadoras, por memorias holográficas.

### **3.5.4 La Holografía como Dispositivo de Seguridad**

Hacer un holograma no es un trabajo muy simple, pues requiere en primer lugar de conocimientos y en segundo lugar de un equipo que no todos poseen, como láseres y mesas estables. Esto hace que los hologramas sean difíciles de falsificar, pues ello requeriría, además, que el objeto y todo el proceso para hacer el holograma fueran idénticos, lo que obviamente en algunos casos puede ser imposible. Por ejemplo, el objeto puede ser un dedo con sus huellas digitales. Esto hace que la holografía sea un instrumento ideal para fabricar dispositivos de seguridad.

Un ejemplo es el de una tarjeta para controlar el acceso a ciertos lugares en los que no se desea permitir libremente la entrada a cualquier persona. La tarjeta puede ser tan sólo un holograma con la huella digital de la persona. Al solicitar la entrada al lugar con acceso controlado, se introduce la tarjeta en un aparato, sobre el que también se coloca el dedo pulgar. El aparato compara la huella digital del holograma con la de la persona. Si las huellas no son idénticas, la entrada es negada. De esta manera, aunque se extravíe la tarjeta, ninguna otra persona podría usarla.

Otro ejemplo muy común son los pequeños hologramas prensados que tienen las nuevas tarjetas de crédito. Estos hologramas, por ser prensados, son de los más difíciles de reproducir, por lo que la falsificación de una tarjeta de crédito se hace casi imposible. Si alguien con los conocimientos y el equipo quisiera falsificar estos hologramas lo podría hacer, pero su costo sería tan elevado que sería totalmente incosteable, a menos que lo hiciera en cantidades muy grandes a fin de que el costo se repartiera.

#### 4. TÉCNICAS DE TRANSMISION

Desde la creación de las llamadas redes de difusión y la tremenda popularidad de la que hoy goza Internet nos encontramos con una serie de necesidades, anteriormente inexistentes, que ahora se deben cubrir (información, comunicación, entretenimiento, etc.) Aquí es donde podemos encontrar la unión entre las redes de comunicación y los servicios multimedia; emisión de radio por Internet, retransmisión de conciertos en directo, charlas virtuales, etc. Todos estos servicios tienen una mayor demanda cada día y es misión de los sistemas informáticos hacerlos posibles.

El rendimiento de los procesadores crece de forma exponencial desde hace varias décadas, lo que permite realizar operaciones cada vez más complejas. Desgraciadamente, las redes de comunicación no crecen al mismo ritmo, de ahí que las técnicas de compresión hayan avanzado tanto en los últimos años y se hayan convertido en una herramienta de uso diario; cuando la cantidad de datos a transmitir crece mucho más rápido que el ancho de banda de nuestro enlace nos vemos obligados a utilizar técnicas alternativas que nos permitan realizar la transmisión en el menor tiempo posible. Los sistemas de transmisión de vídeo actuales suelen tener limitaciones importantes en cuanto al ancho de banda se refiere. Los equipos que sirven las secuencias deben tener una gran velocidad de transmisión para hacer llegar la información a los clientes que se conectan al mismo. De esta forma, el número de clientes capaces de acomodar se encuentra íntimamente ligado al ancho de banda disponible. Esta limitación es muy importante cuando hablamos de transmisiones de eventos multitudinarios (conciertos, deportivos, etc.) o de sistemas en tiempo real, donde la información no sólo debe llegar, sino que debe hacerlo en el menor tiempo posible.

Existen gran variedad de estándares de compresión de vídeo, en parte debido a cuestiones legales (patentes, código propietario) y en parte debido a que ciertos codecs funcionan mejor en ciertos casos.

Aquí vamos a ver la historia y el funcionamiento de los siguientes codecs Quesada L. (2006):

- H.261.
- MPEG-1.
- MPEG-2.
- H.263.
- MPEG-4.
- H.264/AVC.

##### 4.1 H.261.

H.261 es un estándar de codificación de vídeo diseñado por la International Telecommunication Unión en 1990 para transmitir vídeo sobre líneas ISDN a bitrates entre 40kbits/s y 2Mbits/s.

Compuesto por imágenes luma (de brillo) en formatos CIF (Common Intermediate Format) y QCIF

(Quarter CIF) con resoluciones 352x288 y 176x144, e imágenes chroma (de color) con resoluciones 176x144 y 88x72.

Fue el primer estándar de codificación de vídeo llevado a la práctica, y aunque fuera pionero y muy mejorable, todos los subsecuentes estándares de codificación de vídeo internacionales (desde MPEG-1 hasta incluso H.264) se han basado en su diseño. Adicionalmente, los métodos de desarrollo colaborativo del estándar conforman los básicos del proceso de estandarización hoy en día.

La unidad básica de procesamiento en H.261 se llama macrobloque. Cada imagen consiste de un array 16x16 de mediciones luma de 22x18 y dos arrays 8x8 de mediciones chroma de 11x9 (azul (+amarillo) y rojo (+verde)) (YCbCr). Se descompensa el color porque las distorsiones de luminosidad son mejor captadas por el ojo humano. La predicción de imágenes elimina la redundancia temporal, y se utilizan vectores de movimiento para que el codec compense los movimientos simples. Se utiliza una codificación de transformaciones usando una transformada de coseno discreta (DCT) (transformada de Fourier) de 8x8 para eliminar la redundancia espacial.

Se aplica cuantificación escalar para redondear los coeficientes de la transformada a la precisión escaneados en zig zag (por diagonales) y comprimidos usando un código de longitud variable para eliminar la redundancia estadística.

El control de BitRate de H.261 monitoriza el estado del buffer para ajustar el tamaño del paso del cuantificador ( $q$ ) según la siguiente fórmula:

$$q = 2[B/200*p] + 2$$

El estado del buffer es  $B$ .

El bitrate máximo es  $p \times 64\text{kbits}$ .

El tamaño máximo del buffer es  $B_s = p \times 6.4\text{kbits}$ .

Además, se llevan a cabo dos operaciones adicionales en el control del buffer para prevenir underflows u overflows:

- Si el nivel del buffer alcanza el punto de advertencia para overflow, los siguientes datos no son enviados al buffer para permitir que se descargue.
- Si el nivel del buffer alcanza el punto de advertencia para underflow, se utiliza una técnica de bit stuffing (se añaden bits basura) para que el decodificador pueda mantener la decodificación.

El estándar H.261 tan sólo especifica cómo decodificar el vídeo, de modo que se permitió a los diseñadores de codificadores realizar sus propios algoritmos de codificación, siempre que generaran datos que cualquier decodificador hecho conforme al estándar fuera capaz de decodificar.

De esta manera, también se permitió a los codificadores y decodificadores realizar cualquier tipo de preproceso y postproceso. Por ejemplo, una técnica efectiva de postproceso es la conocida como “deblocking filtering” que reduce la aparición de artefactos causados por la compensación de movimiento basada en bloques.

Las aplicaciones de videoconferencia generalmente transmiten imágenes de cabeza y hombros de gente con movimiento limitado y un fondo estático, por eso H.261 es adecuado para estas aplicaciones. Las mejoras introducidas en los estándares subsiguientes han resultado en mejoras significantes en la compresión con respecto a H.261, de forma que H.261 ahora está obsoleto, aunque sigue manteniéndose compatibilidad con él en ciertos sistemas de videoconferencia y para ciertos tipos de vídeo en Internet. En cualquier caso, H.261 conforma un hito histórico en el campo de la codificación de vídeo.

## 4.2 MPEG-1.

MPEG-1 define un grupo de estándares de codificación y compresión de audio y vídeo aceptados por MPEG (Moving Picture Experts Group). La calidad y el bitrate de MPEG-1 se equiparan a los de una cinta VHS. Consiste en varias partes:

- Sincronización y multiplexado de vídeo y audio.
- Codec de compresión para señales de vídeo no entrelazadas.
- Codec de compresión para señales de audio. El estándar define tres capas o niveles de complejidad (MP1, MP2, MP3).
- Procedimientos para comprobar la consistencia.
- Software de referencia.

Fue diseñado con el objetivo de alcanzar calidad de vídeo aceptable a 1.5Mbits/segundo y resolución de 352x240 a 29.97 frames por segundo o 352x288 a 25 frames por segundo.

La gran desventaja de MPEG-1 es que sólo soporta imágenes progresivas, que aunque disminuye el bitrate necesario y la capacidad de proceso del decodificador en cuestión, provoca artefactos cuando hay movimiento rápido. Aun así, MPEG-1 es el formato más compatible de la familia MPEG y se puede reproducir en casi cualquier ordenador y reproductor de VCD o DVD. De hecho, MPEG-1 se centra en almacenamiento de vídeo en CD-ROMs.

El control de bitrate de MPEG consiste en repartir el número total de bits disponibles entre los diferentes tipos de imágenes. Estos tipos son: Imágenes I: Imágenes codificadas sin predicción. Todos los macrobloques son codificados explícitamente. Imágenes P: Imágenes predichas. Los macrobloques pueden ser codificados con predicción hacia delante a partir de referencias a imágenes I o P anteriores, o pueden ser codificados explícitamente.

Imágenes B: Imágenes predichas bidireccionalmente. Los macrobloques pueden ser codificados con predicción hacia atrás, hacia adelante, hacia ambas, o codificados explícitamente.

MPEG sugiere que para las imágenes P se reserven de 2 a 5 veces el número de bits para las imágenes B, y que para las imágenes I se reserven hasta 3 veces el número de bits para las imágenes P, para obtener buenos resultados para escenas típicas. Aumentar el número de imágenes predichas podría mejorar la compresión a costa del tiempo de proceso de decodificación y de la tolerancia a errores. Aparte, no se puede utilizar predicción desde una imagen muy anterior o posterior, porque se obligaría al decodificador a mantener un buffer enorme en memoria y reduciría las capacidades en tiempo real del codec.

Se recomienda el algoritmo de cuantificado adaptativo, de forma que cada macrobloque es clasificado usando una dificultad de codificación  $c_l$  (fila, columna) y es asignado un paso del cuantificador  $Q$  (fila, columna), con un paso del cuantificador  $\Delta Q$  1 o 2 y un paso del cuantificador mínimo  $Q_{\text{mínimo}}$  arbitrario tal que:  $Q(\text{fila, columna}) = Q_{\text{mínimo}} + \Delta Q \cdot c_l(\text{fila, columna})$ . El codificador monitoriza el estado del buffer y ajusta  $Q_{\text{mínimo}}$  y  $\Delta Q$  para evitar overflows y underflows.

Este algoritmo de cuantificado se justifica de la siguiente manera: A más difícil de codificar es un macrobloque, más grande es el paso del cuantificador que se usará para codificarlo. De esta forma, para codificar cada macrobloque se utilizarán aproximadamente el mismo número de bits.



### 4.3 MPEG-2.

MPEG-2 es un estándar para la codificación genérica de imágenes en movimiento e información de audio asociada.

Fue desarrollado por MPEG (Motion Pictures Expert Group) con ayuda de ITU-T (International Telecommunication Unión). Aunque es utilizado para especificar el formato de las señales de televisión digital y de los formatos de DVD, no lo especifica completamente. Las distintas instituciones regionales lo adaptan a sus necesidades restringiendo y aumentando distintos aspectos del estándar, por lo que existen diferentes perfiles y niveles.

Con algunas mejoras, también son utilizados en la mayoría de transmisiones de High Definition Televisión (HDTV).

Incluye distintas partes (4 y 5 similares a MPEG-1):

- Parte 1 (de sistema): Define dos formatos de contenedor.
- Parte 2 (de vídeo): Similar a MPEG-1 aunque ofrece soporte para vídeo entrelazado. Ofrece mejor funcionamiento para bitrates superiores a 3Mbits/s.
- Parte 3 (de audio): Mejora el audio de MPEG-1, permite la codificación de audio de más de dos canales.
- Parte 6: Describe extensiones DSM-CC.
- Parte 7 AAC: Audio más eficiente, avanzado y más complejo de implementar, requiere mejor hardware.
- Parte 8: Extensión de vídeo de 10-bits.
- Parte 9: Extensión para interfaces de tiempo real.
- Parte 10: Extensiones de comprobación para DSM-CC.

Utiliza el algoritmo de control de bitrate TM5, que consiste en los siguientes pasos:

- Reserva un número de bits objetivo para la siguiente imagen dependiendo del tipo de imagen y de ponderaciones globales.
- Toma un valor de referencia de paso del cuantificador para los macrobloques a partir de la siguiente fórmula:

$$Q = 2[B \cdot 31 / r]$$

Donde  $r=2 \cdot (R/f)$ , R es el bitrate, f son las imágenes por segundo, y B es el grado de ocupación del buffer.

- Cuantificación adaptativa, el paso del cuantificador para un macrobloque es  $q=Q \cdot N$ , donde N es la actividad espacial normalizada para ese macrobloque.

### 4.4 H.263.

H.263 es un codec de vídeo originalmente diseñado por ITU-T en 1995/1996 como una solución de codificación que requiere un bitrate bajo, en un principio, para videoconferencia.

Fue desarrollado como una mejora evolutiva basada en la experiencia obtenida del desarrollo de H.261, MPEG-1 y MPEG-2.

Sustituye a H.261 ya que ofrece una mejora con respecto a este codec para cualquier bitrate. Más adelante fue mejorado en proyectos conocidos como H.263v2 o H.263v3. La mayoría del vídeo flash (como el utilizado en YouTube, Google vídeo, MySpace) es codificado en este formato.

Utiliza el algoritmo de control de bitrate TMN8.

+

El algoritmo TMN8 consiste en dos niveles:

- Nivel de imagen: Si el nivel del buffer excede un determinado umbral, la imagen es descartada, en caso contrario, a la imagen se le asigna un número de bits basado en el nivel de llenado del buffer.
- Nivel de macrobloque: Se adapta el paso del cuantificador para alcanzar el bitrate objetivo según el modelo logarítmico R-Q:

$$R = (1/2) \log_2 (2e^2 \sigma^2 / Q) \text{ si } \sigma^2 / Q^2 > 1/2e \text{ ó}$$

$$R = (e / \ln 2) (\sigma^2 / Q^2 \sigma^2) \text{ si } \sigma^2 / Q^2 \leq 1/2e$$

Donde Q es el paso del cuantificador buscado,  $\sigma$  es la desviación estándar del macrobloque y R es el bitrate buscado.

#### 4.5 MPEG-4.

MPEG-4 es un estándar usado primariamente para comprimir audio y vídeo. Fue presentado a finales de 1998 y es la designación para un grupo de estándares de codificación de audio vídeo y tecnología relacionada aceptado por MPEG (Moving Picture Experts Group).

Los usos para el estándar MPEG-4 son la web (streaming) y la distribución de datos multimedia en CD's y DVD's, videoconferencia y emisión de televisión. MPEG-4 incluye la mayoría de las características de MPEG-1 y MPEG-2 y otros estándares relacionados y añade nuevas características como soporte VRML para rendering 3D, ficheros compuestos orientados a objetos, Digital Rights Management, varios tipos de interactividad, y AAC como parte del estándar.

Con MPEG-4, una escena es vista como una composición de objetos de vídeo (VOs) con propiedades intrínsecas tales como una forma, movimiento, y textura. Las instancias de los VOs se llaman planos de VOs. La información de forma es binaria y se llama plano alpha.

MPEG-4 permite comprimir en gran medida los datos audiovisuales para su almacenamiento y transmisión, a la vez que respeta la calidad de video y audio. Este estándar crea representaciones codificadas de los datos de audio y video que forman la secuencia a través de la codificación basada en objetos. El estándar define una escena audiovisual como una representación codificada de objetos audiovisuales que tienen cierta relación en el tiempo y en el espacio. Estos objetos están organizados de forma jerárquica, en la base de ésta encontramos objetos de medios de comunicación primitivos, como: Imágenes estáticas (fondo fijo), objetos de vídeo (una persona hablando), objetos de audio (la voz asociada a una persona o la música de fondo). MPEG-4 estandariza varios de estos objetos y es capaz de representar los tipos de contenido natural y sintético.

Los objetos tienen algunas propiedades adjuntas como sus coordenadas espaciales, escala, localización, zoom, rotación, etc. Estas características permiten a un usuario reconstruir la secuencia original tras decodificar todas las capas de objetos. El usuario también puede manipular la imagen mediante operaciones sencillas sobre cualquiera de los parámetros de los objetos. Posee un protocolo de animación facial, donde un modelo tridimensional de una cara puede ser animado en tiempo real.

### 4.5.1 Códecs MPEG-4

Cuando apareció el estándar MPEG-4, algunos desarrolladores de software programaron códecs que cumplieran las especificaciones de dicho estándar. Los códecs más conocidos hoy en día son DivX, Xvid y H.264. A continuación analizaremos cada uno de éstos. Herrero et al (2004):

#### A. Códec DivX

DivX apareció en el verano de 1999 (aunque estas mismas siglas se utilizaron previamente para un sistema de pago por visión). Para visualizar videos con este códec, es necesaria su instalación. La versión más actualizada es el DivX 5. Entre sus características principales, nos podemos encontrar:

- Cuatro modos de velocidad de codificación (rendimiento/calidad): Muy rápida, rápida, estándar y lenta.
- Realzamiento psicovisual basado en DCT.
- Preprocesamiento: Filtros espaciales y temporales.

Este códec también permite elegir la cantidad de imágenes I, P y B y el intervalo máximo de imágenes clave, la compensación de movimiento global el tipo de cuantización (H.263 y MPEG-2) y entrelazado.

#### B. Códec XviD

Se originó a partir de DivX 4 y está basado en las mismas librerías de las que partió tanto DivX 4 como OpenDivX. Es de código abierto, por ello existen numerosos colaboradores mejorándolo desinteresadamente. Los perfiles que posee XviD permiten que la codificación sea compatible con los reproductores de mesa DivX. El perfil "AS" (Avance simple) limita la resolución a 352x288 y la velocidad de transmisión de bits permite hasta 384 Kbits/s. Los perfiles "ARTS" (Tiempo Real Avanzado de Streaming) limitan el vídeo a la misma resolución que el perfil "simple" pero el límite de velocidad de transmisión de bits es de 4 Mbits/s. Cada perfil tiene varios niveles con los que se puede variar la resolución/bitrate. Existe también la posibilidad de configurar libremente el códec sin utilizar ningún perfil. Las características que poseen estos perfiles son:

- Cantidad de Imágenes I, P y B. Se puede seleccionar la máxima cantidad de imágenes I, P y B seguidas. También se pueden seleccionar zonas donde aplicar un ajuste diferente de compresión a un grupo determinado de imágenes.
- Compensación de movimiento de cuarto de píxel. Ayuda a valorar el movimiento de forma más precisa, afina los detalles y mejora la compresión. Es útil para videos con baja resolución. No es compatible con los DVD-DivX de sobremesa.
- Compensación de movimiento global: Es una técnica que reduce la redundancia temporal en una secuencia de imágenes.
- Enmascaramiento de la luminancia o cuantización adaptable: XviD reduce un poco la calidad utilizando tasas de bits más bajas en las áreas más brillantes y en las más oscuras pues la pérdida de calidad en estas áreas es menos visible para el ojo humano.
- Cuantización entramada: Permite mayor compresión pero puede degradar la imagen y disminuir la velocidad.
- Tipo de cuantizador: H.263 (para tasas de bits bajas) y MPEG (para tasas de bits altas).

- Entrelazado: Trata a cada campo por separado, se utiliza cuando la fuente es entrelazada.

El Xvid posee la opción “precisión de búsqueda de movimiento” que otros códecs no tienen.

### C. Códec H.264

Este códec se basa en la norma H.264. Está pensado para utilizarse en aplicaciones de video que tengan baja tasa de bits (con media y alta definición en televisión), en flujos de video a través de Internet y en aplicaciones de alta definición en cine y DVD. El proceso de compresión pasa por 7 estados:

1. Estimación de movimiento: Además de comparar la imagen analizada con otras anteriores o posteriores buscando un patrón que indique el movimiento, este códec incorpora el concepto de referencia múltiple de imágenes. Se aplica en movimientos que se repiten periódicamente en la naturaleza, interpretación de movimientos y obstrucciones. Con esto se reduce el tamaño de almacenamiento.

2. Intra-Estimación: Se utiliza cuando el objeto que se trata es estático. Predice el nuevo bloque extrapolando los píxeles vecinos de bloques adyacentes en un grupo de direcciones diferentes predefinidas. La diferencia entre el bloque predefinido y el bloque actual es la información codificada.

3. Transformación: Los resultados de los estados 1 y 2 son transformados desde un dominio espacial hasta un dominio de frecuencias con el que trabaja el códec. Usa una DCT basada en un algoritmo de transformación de 4x4 enteros que reduce la aparición de macrobloques en el video y el “efecto mosquito” en los contornos.

4. Cuantización: Esta fase reduce la precisión del conjunto de números enteros obtenidos en la fase anterior cuantizándolos y eliminando los coeficientes con frecuencias altas sin variar la calidad perceptible. 5. Recorrido de filtros: En esta fase se utiliza un filtro llamado “de-bloking”. En los macrobloques 16x16, el filtro elimina los efectos producidos por las diferencias de tipos de estimación de los bloques adyacentes o por diferentes escalas de cuantización. En los bloques internos 4x4 el filtro elimina los efectos causados por la cuantización y por diferencias con los vectores de movimiento de los bloques adyacentes. Se consigue aumentar la nitidez de la imagen.

5. Codificación Entrópica: La codificación entrópica basa su eficiencia en la asignación de muy pocos bits a los símbolos que se usan con mucha frecuencia y un mayor número de bits a aquellos símbolos que aparecen pocas veces.

## 4.6 H.264/AVC.

H.264, MPEG-4 Part 10 o AVC (Advanced Video Coding) es un codec estándar para vídeo digital que consigue una muy alta compresión de datos. Quesada L (2006). Fue diseñado por el grupo de expertos en codificación de vídeo del ITU-T junto con MPEG.

El H.264 de ITU-T y el MPEG-4 Part 10 de MPEG son mantenidos conjuntamente para que sean idénticos. Reduce el bitrate necesario para una calidad de vídeo determinada a la mitad o menos. Proporciona suficiente flexibilidad para permitir al estándar ser aplicado a una gran variedad de aplicaciones (bajos y altos bitrates o calidades de vídeo).

El control de bitrate se divide en las siguientes partes:

- Control del nivel del GOP: Calcula los bits disponibles para el resto de imágenes en el GOP e inicializa el paso del cuantificador de la frame del refresco de decodificación instantánea (IDR).
- Control del nivel de imagen: En este nivel, se calcula el paso del cuantificador por el modelo cuadrático y se usa para llevar a cabo una optimización R-D para cada macrobloque de la imagen, mediante la fórmula:

$$R = \frac{a1 * \hat{S}}{Qstep} + \frac{a2 * \hat{S}}{Q^2step} - h$$

Donde h es el número de bits para la cabecera y los vectores de movimiento, a1 y a2 son los parámetros del modelo estimados por regresión lineal, y  $\hat{S}$  es la diferencia absoluta de la imagen actual i.

$\hat{S}_i$ , a su vez, es hallado así:

$$\hat{S}_i = b_1 * S_{i-1} + b_2$$

Donde b1 y b2 son los parámetros del modelo actualizados por la regresión lineal.

- Control del nivel de unidad básica: Una unidad básica se define como una imagen, corte, fila de macrobloques o conjunto de macrobloques. Se utiliza un modelo lineal para predecir la diferencia absoluta de la unidad básica actual en la imagen y se utiliza un modelo R-D cuadrático para calcular el paso del cuantificador, que es utilizado para optimizar R-D en la unidad básica. De esta manera se obtiene un buen equilibrio entre calidad de imagen y fluctuación de bitrate.

## 4.7 MPEG-7

MPEG-7 es una representación estándar de la información audiovisual. Permite la descripción de contenidos por palabras clave y por significado semántico (quién, qué, cuándo, dónde) y estructural (formas, colores, texturas, movimientos, sonidos). El formato MPEG-7 se asocia de forma natural a los contenidos audiovisuales comprimidos por los codificadores MPEG-1 (almacena y descarga archivos audiovisuales), MPEG-2 (televisión digital) y MPEG-4 (codifica audio y vídeo en forma de objetos), pero se ha diseñado para que sea independiente del formato del contenido.

El nuevo estándar ayuda a las herramientas de indexación a crear grandes bases de material audiovisual (imágenes fijas, gráficos, modelos tridimensionales, audio, discursos, vídeo e información sobre cómo esos elementos están combinados en una presentación multimedia) y buscar en estas bases de materiales manual o automáticamente.

### 4.7.1. El PostScript del multimedia

MPEG-7 es al multimedia lo que PostScript es al papel. Este último describe a un programa de textos el formato que debe tener la página; MPEG-7 hace lo mismo, pero sobre el contenido audiovisual. El MPEG-7 se basa en el popular lenguaje de metadatos XML, "que dominará Internet dentro de poco; habrá un montón de herramientas y eso favorecerá la interoperabilidad y la creación de aplicaciones", afirma Philippe Salembier, investigador de la Universidad Politécnica de Cataluña, que ha participado en el proyecto MPEG-7 durante los últimos tres años.

Sin embargo, una descripción en XML puede ser muy voluminosa. Es un problema para las aplicaciones en las que el espacio de almacenamiento o el ancho de banda de transmisión son insuficientes (discos con capacidad limitada, transmisión por módem, etcétera). Para estos casos se ha desarrollado el compresor BIM (Binary Format for MPEG-7).

En un escenario tipo, una aplicación genera la descripción MPEG-7 del contenido de, pongamos, un millón de películas; luego se pasa al formato XML, se almacena en servidores con discos de gran capacidad y ya está listo para su uso. Si la información de la descripción es demasiado grande para los servidores o si se tiene que mandar en un canal de transmisión con poco ancho de banda, el XML se compacta en un espacio hasta 100 veces menor con el codificador BIM. Al final del proceso se puede decodificar otra vez en XML y ya se pueden utilizar esos datos. "Además, el BIM es más robusto que el XML frente a los errores de transmisión", añade Salembier.

MPEG-7 incluye, además de la descripción de los contenidos, información sobre el tipo de compresión utilizada (JPEG, en dibujos; MPEG-2, en imágenes), las condiciones para acceder (derechos, precio), clasificación (adultos, por ejemplo), enlaces a otros materiales relevantes (para acelerar la búsqueda) y el contexto (final de los 200 metros femeninos de los Juegos Olímpicos de Verano de 2000).

El lenguaje MPEG-7 puede generar grandes ingresos al sector audiovisual (producción, posproducción, distribución y consumo). A juicio de Goldman Sachs, el mercado de herramientas informáticas para la gestión de contenidos pasará de los 378 millones de dólares (7.100 millones de pesetas) del año 2000 a los 4.500 millones de dólares (850.000 millones de pesetas) en 2005.

La industria también prepara el estándar MPEG-21, que propone la creación de un marco tecnológico para todo el proceso multimedia, explica José M. Martínez, científico de la Universidad Politécnica de Madrid, que también ha participado en el desarrollo del MPEG-7. El MPEG-21 debe abarcar desde los terminales utilizados al tipo de redes. También será importante para proteger legalmente los contenidos para todas las variedades de uso. En la primera fase, intentará definir las diferentes opciones tecnológicas de extremo a extremo en la cadena multimedia. Por ejemplo, se enviará un esquema MPEG-7 adecuado a un portátil.

## **5. TECNICAS DE VISUALIZACION 3D**

En los últimos años ha crecido sustancialmente el desarrollo de aplicaciones que utilizan la tecnología de realidad virtual. El objetivo de la realidad virtual es sumergir al usuario en un mundo artificial simulado por la computadora. Uno de los conceptos que caracteriza a la realidad virtual es la "tridimensionalidad". La sensación 3D se consigue mediante la proyección de gráficos estereoscópicos, es decir, se generan una imagen ligeramente diferente para el ojo izquierdo y otra para el ojo derecho del usuario.

Existen fundamentalmente dos tipos de sistemas estereoscópicos cuyo objetivo principal es proporcionar al usuario sensaciones de profundidad: sistemas estéreo activo y sistemas estéreo pasivo. En ambos métodos, el usuario debe llevar puestas unas gafas especiales. En el primero de los casos, las gafas incluyen componentes electrónicos, cosa que no ocurre en el segundo caso.

Los sistemas de estéreo activo son los que ofrecen una mayor calidad y generalmente, tienen también un costo superior. Son sistemas secuenciales, en los que las imágenes correspondientes al ojo izquierdo y derecho se alternan en rápida sucesión (típicamente 120 Hz.). En este tipo de sistemas se utilizan gafas con obturadores LCD (shutters glasses).

Los sistemas de estéreo pasivo más comunes utilizan proyectores. Las gafas no disponen de componente electrónico alguno, sino que llevan incorporados unos filtros análogos a los que se colocan en los proyectores. Los gráficos pueden ser generados por una sola computadora o con dos computadoras sincronizadas.

Para la visualización de modelos en tres dimensiones se han utilizado técnicas de visualización estereoscópicas entre las que destacan: Polarización, Conmutación y Anáglifo. Estos métodos permiten la separación de las imágenes para el ojo izquierdo y derecho cuya finalidad es proporcionar un efecto estereoscópico, es decir, un efecto tridimensional, en donde el usuario percibe profundidad cuando está observando una escena.

Las técnicas más comunes de visualización 3D son: visualización volumétrica, autoestereoscopia y estereoscopia. Escrivá M. et al (2006).

### 5.1 Visualización Volumétrica

La visualización volumétrica consiste en mostrar imágenes tridimensionales dentro de un volumen para ello se utilizan dispositivos volumétricos los cuales permiten que los objetos o imágenes puedan ser vistos por varios observadores al mismo tiempo desde cualquier orientación.

Los dispositivos Volumétricos son capaces de mostrar imágenes tridimensionales dentro de un volumen, en contraste con otros dispositivos, tales como una terminal grafica convencional, donde las imágenes se muestran sobre una superficie plana.

Puesto que las imágenes volumétricas ocupan físicamente un espacio 3D, se muestran automáticamente a uno o varios observadores de forma tridimensional. Esto es debido a que proporcionan una vista del objeto representado desde todos los puntos de vista. Los dispositivos volumétricos permiten que los objetos o imágenes puedan ser vistos por varios observadores al mismo tiempo desde cualquier orientación. Existen varios métodos para conseguir producir imágenes volumétricas. Estos métodos se pueden agrupar en dos categorías, dependiendo de si el volumen del dispositivo esta estático o en movimiento.

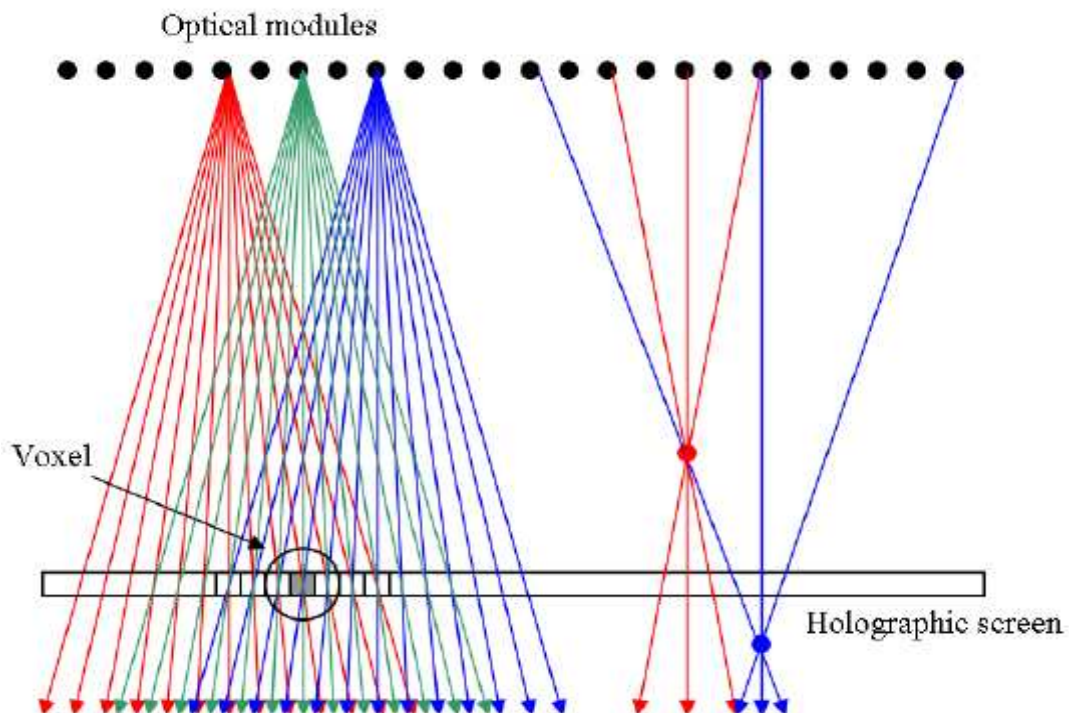


Figura 5: Una representación esquemática de como funciona el dispositivo de Holografía.

### 5.1.1 Dispositivos con Volumen Dinámico

En este tipo de dispositivos, el volumen se crea por el movimiento mecánico, vibratorio o rotatorio de una pantalla. La velocidad del movimiento de la pantalla debe ser lo suficientemente rápida para que esta no pueda ser percibida por el observador.

Los primeros dispositivos de este tipo se propusieron sobre el año 1940 como, por ejemplo, el de Parker and Wallis (1948). Desafortunadamente, la ausencia de los sistemas informáticos necesarios para el control de los dispositivos hizo imposible producir sistemas operacionales en aquella época.

Fue en la década de los 80-90, debido al aumento de potencia de los ordenadores, cuando mejoró la viabilidad de estos dispositivos. En este apartado describimos cuatro dispositivos con volumen dinámico, el de espejo oscilante, el de espejo varifocal, el Felix 3D y el Perspecta Spatial 3D.

### 5.1.2 Dispositivos de Espejo Oscilante

El espejo oscilante es un dispositivo que aparece en la década de los 60. Consiste en un espejo situado frente a un monitor CRT y que puede vibrar o moverse hacia adelante y hacia atrás rápidamente, (Figura 7).

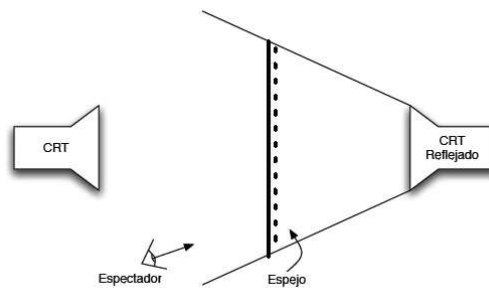


Figura 7: Esquema de funcionamiento de un espejo oscilante.

El espectador debe estar en el mismo lado del espejo que el monitor CRT, Así la imagen el monitor se refleja en el espejo. Cualquier punto dibujado en el monitor aparece a una profundidad distinta que depende de la posición del espejo en el instante en que el punto aparece en el monitor.

### 5.1.3 Dispositivos de Espejo Varifocal

El espejo varifocal es la técnica de volumen dinámico más empleada. El dispositivo contiene un espejo circular flexible anclado por los bordes. Se utiliza un altavoz de bajas frecuencias para cambiar la longitud focal del espejo (Figura 8). Controlando el altavoz y lo que se proyecta en el monitor, el espectador es capaz de ver los objetos con profundidad.



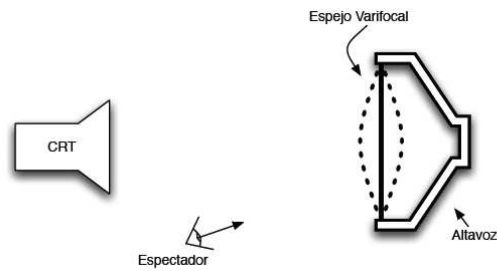


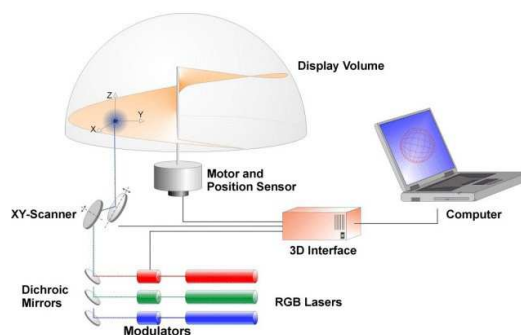
Figura 8: Esquema de funcionamiento del espejo varifocal.

### 5.1.4 Felix 3D

Los recientes avances en las tecnologías de proyección mediante diodos láser y, en particular de la tecnología DLP2 de Texas Instruments han hecho posible la fabricación y distribución de estos dispositivos de visualización. Están compuestos de dos partes: un proyector y una superficie de proyección que se monta sobre un eje y utiliza un motor para hacerla girar a gran velocidad.

Actualmente hay dos tecnologías de proyección que se están usando en estos dispositivos. Los dispositivos de Felix 3D y de Genex utilizan tres láseres: uno rojo, uno verde, y uno azul, mientras que Perspecta utiliza la tecnología DLP.

El visualizador Felix 3D se basa en una pantalla helicoidal semitransparente (figura 9(a)). La pantalla de proyección se hace girar rápidamente, proporcionando un medio de visualización volumétrica a través del cual se proyectan pulsaciones láser. Debido a la alta velocidad de rotación (unas 1.200 RPM) la pantalla en si se hace invisible al observador.



a) Principio de operación del visualizador Felix 3D.



b) Proyección volumétrica de una escalera en espiral en el visualizador.



c) Gráfico de trazado tridimensional de una macromolécula.

Figura 9: Funcionamiento del dispositivo Felix 3D.

Las imágenes en color se producen mediante la combinación de láseres rojo, verde y azul (RGB). A través de la modulación por separado de cada componente, es posible producir cualquier color (figura9).

### 5.1.5 Perspecta Spatial 3D

Actuality ofrece un dispositivo volumétrico que funciona de manera similar al Felix 3D, el Perspecta Spatial 3D. Este utiliza una pantalla vertical en lugar de una helicoidal, donde se proyectan imágenes a gran velocidad (ver figura 10). El ojo funde estas imágenes para crear una imagen 3D sin costuras. Dicha imagen se encuentra dentro del dispositivo volumétrico (figura 11).

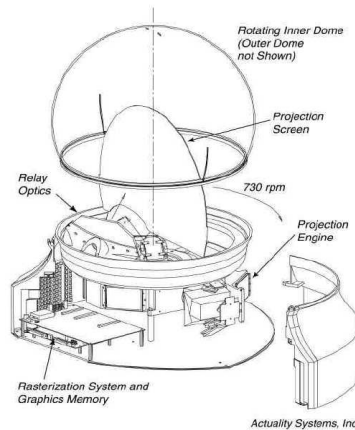


Figura 10: Esquema del dispositivo de visualización espacial Perspecta.

El inconveniente que tienen los dispositivos con volumen dinámico es que el usuario no puede realmente tocar la imagen. Mientras los dispositivos autoestereoscópicos lo permitían, los dispositivos volumétricos sólo lo permiten cuando el volumen es estático y no en todos los casos.

### 5.2. Dispositivos con Volumen Estático

Los dispositivos que pueden crear un volumen sin la necesidad de emplear el movimiento mecánico se llaman dispositivos con volumen estático. Estos dispositivos son más fiables, y el hecho de que la frecuencia de refresco de la imagen no venga determinado por el movimiento de la pantalla permite que el parpadeo de la imagen pueda ser reducido.



(a)



(b)

Figura 11: Imagen del dispositivo Perspecta Spatial 3-D

Los dispositivos de este tipo que se han propuesto hasta la fecha emplean un volumen gaseoso o cristalino para producir la imagen 3D o volumétrica. Entre ellos podemos encontrar el Helio Display y el FogDisplay. Además hay otro dispositivo, el DepthCube, que también produce imágenes volumétricas, pero dentro de una pila de LCDs.

Mientras los primeros permiten al usuario tocar los objetos 3D, este último sólo permite visualizarlos.

### 5.2.1. Helio display

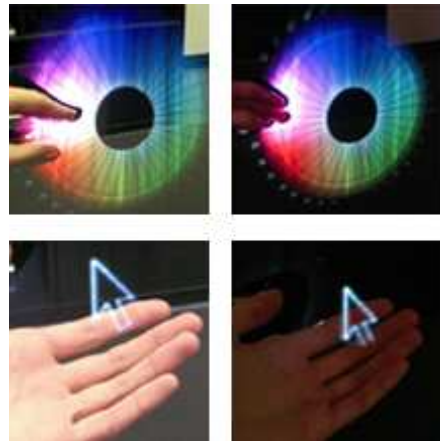


Figura 12: Imágenes visualizadas y manipuladas con Helio Display.

El Helio display es una tecnología interactiva que proyecta en el aire imágenes quietas o en movimiento que pueden ser manipuladas con la punta de los dedos. Estas imágenes son bidimensionales, y no son holográficas. IO2 Technology ha completado un prototipo totalmente operativo de este dispositivo.



Figura 13: Vista de un dispositivo Helio Display.

Este dispositivo puede mostrar imágenes sobre un área de 30 pulgadas de diagonal, incluyendo secuencias de vídeo. El dispositivo trabaja creando una nube de partículas microscópicas que hacen al aire imagen-amigable. Esta nube de partículas es proyectada e iluminada desde abajo, generando una imagen que flota en el aire.

## 5.2.2. FogScreen



Figura 14: El dispositivo volumétrico FogScreen.

FogScreen es una superficie de proyección incorpórea y penetrable que está formada por vapor seco proveniente del agua pura del grifo sin productos químicos. El vapor es fresco y seco al tacto.

FogScreen es fina y plana; por eso es capaz de generar una imagen de proyección de alta calidad. La imagen flota en el aire y puede ser observada por varios usuarios. Con sólo dos proyectores estándar se puede crear una imagen frontal y otra posterior a cada lado de la cortina de vapor. El espectador puede interactuar con la imagen y formar parte de ella.

## 5.2.3. Monitores con Varios LCDs

Como el Perspecta el dispositivo DepthCube de Light-Space usa tres chips DLP (ver figura 16(a)). Ahora bien, en vez de una sola pantalla de proyección este dispositivo tiene 20 pantallas de cristal líquido separadas una de otra unos 5 milímetros (figura 15). Las imágenes de los DLPs atraviesan las pantallas LCD produciendo imágenes a distintas profundidades. Estas imágenes se combinan ópticamente para formar una imagen 3D.

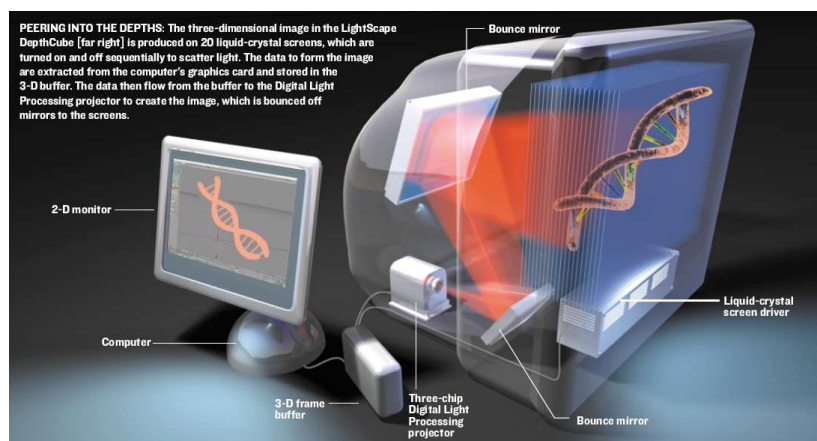


Figura15: Esquema del dispositivo DepthCube.

DepthCube es totalmente sólido y no se ve afectado por vibraciones como ocurre con los dispositivos de volumen dinámico. Otra ventaja de este dispositivo es su relativa compatibilidad con software tridimensional existente. Puesto que las imágenes del DepthCube se proyectan sobre planos 2D tienen una geometría cartesiana, haciendo al dispositivo compatible con el software que utilice la librería OpenGL. La visualización se hace mediante un software (llamado GL Interceptor) que en tiempo real extrae las imágenes generadas con OpenGL y su información de profundidad para crear las 20 imágenes que DepthCube necesita.

Existen otras empresas como por ejemplo Deep Video Imaging Ltd. que también fabrican este tipo de dispositivos. En la figura 16 se pueden ver los dispositivos fabricados por Light-Space y por Deep Video.



(a) Aspecto exterior de un DepthCube.



(b) Imagen del monitor MLD 3000

Figura 16: Aquí se puede apreciar los modelos tanto de Light-Space como de Deep Video.

### 5.3 Autoestereoscopía

Los dispositivos autoestereoscópicos muestran una imagen tridimensional a un espectador sin la necesidad de cristales o gafas. Además muchos de ellos permiten que varios usuarios obtengan una vista distinta (y correcta) del mismo objeto 3D.

Hay dispositivos autoestereoscópicos que presentan una imagen diferente a cada ojo, siempre y cuando el espectador este en una posición particular. La mayoría de estos mezclan en la pantalla dos 2 imágenes distintas, una para cada ojo. Estas técnicas tienen dos desventajas: el espectador debe permanecer en una posición fija, y cada ojo ve sólo la mitad de la resolución horizontal de la pantalla.

Los dispositivos holográficos y pseudo-holográficos muestran un campo de luz, permitiendo mostrar diferentes vistas simultáneamente.

Un campo de luz es una estructura 4D que almacena la radianza que fluye a lo largo de un conjunto de rayos, típicamente todos aquellos que parten de una pantalla 2D. Si las radianzas se visualizan correctamente, muchos espectadores pueden ver el mismo objeto simultáneamente.

Calcular y visualizar un campo de luz 4D requiere una gran capacidad de cálculo y almacenamiento. Para evitarlo se visualizan versiones 3D, donde los usuarios sólo ven distintas imágenes si se mueven horizontalmente.

De este modo se obtienen ilusiones 3D con paralaje horizontal. Es decir, el objeto sólo exhibe profundidad y auto-occlusión en la dimensión horizontal. Los campos de luz 4D permiten visualizar paralaje completo, horizontal y vertical.

A continuación se da un repaso a las tecnologías autoestereoscópicas y a dispositivos en fase de prototipo o producción que se pueden encontrar.

#### 5.3.1. Estereogramas

El estereograma es una ilusión óptica basada en la capacidad que tienen los ojos de captar imágenes desde distintas perspectivas. Estas perspectivas diferentes son procesadas por el cerebro de tal forma que parecen ser una imagen tridimensional.

Aunque los estereogramas se utilizan desde hace mucho, se han vuelto muy populares en la actualidad gracias a los RDS1 (figura 17) creados con programas de ordenador.



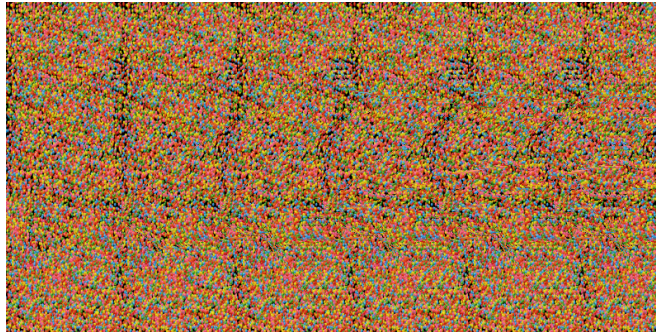


Figura 17: Este estereograma muestra un tiburón

Para ver un estereograma es necesario desenfocar la vista de la imagen, de tal manera que ambas perspectivas sean captadas. Algunos recomiendan mirar al infinito, es decir, fijar la vista en un objeto distante y sin desenfocar, tratar de mirar la imagen. Otros prefieren fijar la vista en un dedo mientras lo acercan lentamente hacia la imagen, o tratar de observar el reflejo de la imagen en un cristal. Depende de cada persona y de su condición visual.

Los estereogramas son una forma barata de representar imágenes 3D visibles por varios observadores. Además, permiten mostrar animaciones.

### 5.3.2. Estereogramas holográficos

La holografía es una técnica avanzada de fotografía, que consiste en crear imágenes que por ilusión óptica parecen ser tridimensionales.

Para esto se utiliza un rayo láser, que graba microscópicamente una película fotosensible. Ésta, al recibir la luz desde la perspectiva adecuada, proyecta una imagen en tres dimensiones.

Un estereograma holográfico es una combinación de la holografía y del par estéreo que graba una serie de imágenes 2D en una película holográfica. Esta película se puede doblar y formar un cilindro, de tal forma que el espectador puede caminar alrededor del cilindro y ver la escena desde cualquier punto de vista. Desde cualquier punto, el ojo izquierdo verá una vista de la escena y el ojo derecho otra.

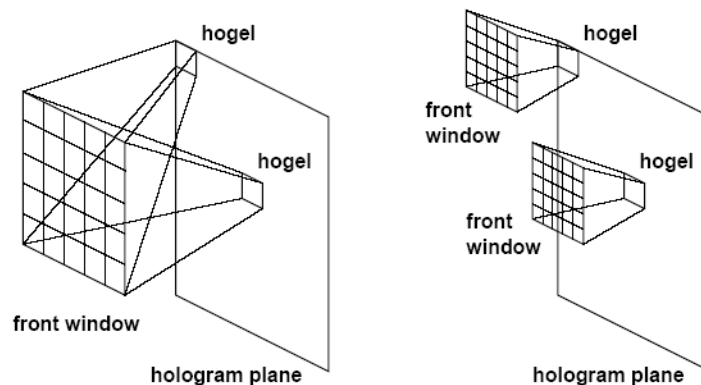


Figura 18: Hoguels en estereogramas holográficos de dos pasos (izquierda) y un paso (derecha). Al iluminar los estereogramas las imágenes almacenadas en cada hoguel se visualizan en las ventanas frontales.

Durante algún tiempo los dispositivos holográficos convencionales no han tenido mucho éxito por sus limitaciones con respecto a color, a ángulo de visión, al material, y al tamaño final de la imagen. Incluso con la proliferación de la técnica del estereograma holográfico en la década de los 80, seguía existiendo la mayoría de estas limitaciones.

Zebra Imaging, Inc. [20] invirtió en el desarrollo de los estereogramas holográficos de un solo paso. Ofrece

una tecnología para imprimir estereogramas holográficos a todo color con un ángulo de visión bastante grande (hasta 110) y de tamaño ilimitado y paralaje completo.

La tecnología del estereograma holográfico de Zebra Imaging se basa en crear una matriz de elementos holográficos de 1 o 2 milímetros de lado llamados hoguets. Como los píxeles de imágenes digitales 2D, la matriz de hoguets se puede utilizar para formar imágenes completas de cualquier tamaño. Cada hoguet es una grabación holográfica en una película de foto-polímero (Figura 18).

El resultado es la aparición de una escena 3D que cruza la superficie del estereograma. Dicha escena tiene apariencia 3D para múltiples usuarios y permite atravesarla siempre que no se cruce el plano de la superficie. Las desventajas de los estereogramas holográficos son su alto coste y la dificultad de iluminarlos.

### 5.3.3. Dispositivos de Barrera de Paralaje

Una barrera de paralaje consiste en una serie de finos cortes verticales en un material opaco. La barrera se coloca cerca de una imagen que se ha registrado en tiras verticales. Si las tiras en la imagen se han muestreado con la frecuencia correcta para los cortes de la barrera de paralaje, y el espectador está a la distancia apropiada de la barrera, esta ocultará los píxeles de la imagen apropiados de la imagen al ojo derecho e izquierdo respectivamente y el espectador percibirá una imagen autoestereoscópica (ver figura 19).

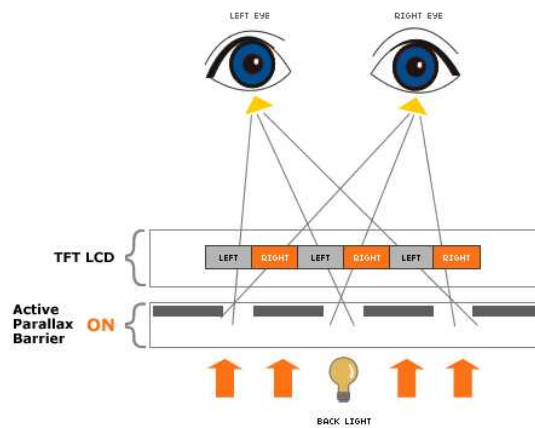


Figura 19: Esquema de funcionamiento de una barrera de paralaje.

Se pueden preparar las imágenes para que sean vistas desde diferentes puntos de vista. Cuando el espectador cambie de posición, la barrera redirigirá la vista a las nuevas imágenes que se han de visualizar (una para cada ojo). Esto permite además que varias personas puedan ver a la vez la escena mostrada. El inconveniente es que el número de vistas está limitado por la óptica y, por lo tanto, al moverse ligeramente sobre un mismo punto se produce un efecto de parpadeo incómodo.

La impresión en láser de alta resolución ha permitido producir imágenes de calidad muy alta: la barrera se imprime en una cara de un material transparente y la imagen en la otra. Esta técnica fue iniciada en la década de los 90 y actualmente está siendo utilizada por Sanyo en sus dispositivos.

Varios fabricantes tienen prototipos o productos que utilizan esta tecnología: Sanyo Electric Co., Ltd., Sharp Corporation (figura 20) o Sony Corporation. En septiembre de 2002, Sharp anunció la creación de un consorcio (<http://www.3dc.gr.jp/english>) con los fabricantes de dispositivos 3D. El consorcio está destinado al avance de estos dispositivos y tiene unos 100 miembros entre los que destacan Itochu Corporation, NTT Data Corporation, Sanyo Electric Co., Ltd., Sharp Corporation y Sony Corporation.



(a) Actius AL3DU.



b) Actius RD3D.



(c) LL-151-3D TFT.

Figura 20: Dispositivos de barrera de paralaje de Sharp.

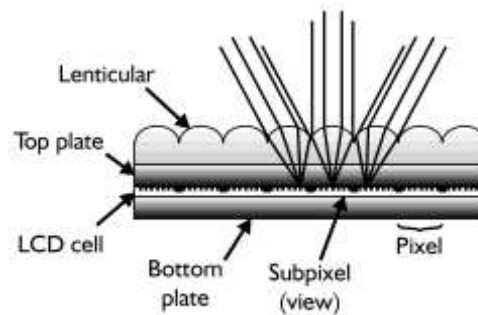
### 5.3.4. Lenticulares

Este tipo de dispositivo es muy similar a las barreras de paralaje. Sobre una pantalla se sitúa una hoja lenticular (ver figura 21(a)) formada por una serie de lentes verticales semicilíndricas hechas normalmente del plástico. La hoja lenticular se diseña de tal forma que la luz paralela que llega a la hoja es enfocada sobre los píxeles de la pantalla.

Cuando la pantalla visualiza una imagen, cada lente se encarga de la información de la imagen que se encuentra detrás de ella y dirige la luz en diferentes direcciones. De este modo aparecen distintas imágenes en función de la posición desde donde se mire al lenticular. Los lenticulares se pueden utilizar también con un monitor u otro dispositivo de visualización 2D para producir una imagen tridimensional dinámica. Empresas como 3DZ venden este tipo de filtros (figura 21(a)) para monitores. Como las barreras de paralaje, los lenticulares tienen solamente paralaje horizontal. Existe otro tipo de dispositivo que usa lentes esféricas en vez de cilíndricas para producir imágenes con paralaje completo (figura 22). Empresas como Stereographics venden dispositivos usando esta tecnología (figura 23).



(a) Montaje del lenticular



(b) Esquema del lenticular

Figura 21: En la figura puede verse como se colocaría un lenticular sobre un monitor y el esquema de funcionamiento del lenticular.



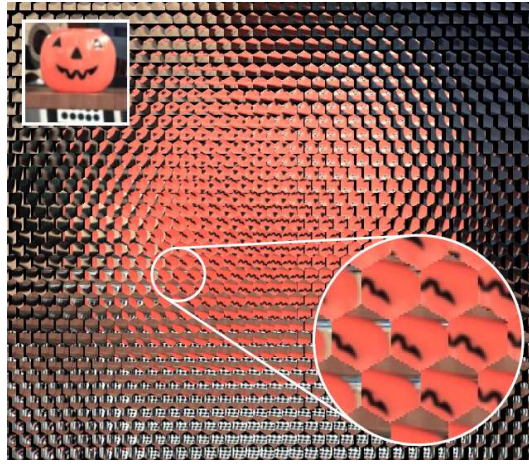


Figura 22: Lenticular con paralaje completo.

### 5.3.5. El Dispositivo de Holográfica

Holográfica ha desarrollado un dispositivo holográfico escalable multi-usuario. El dispositivo se basa en tecnología de proyección trasera y utiliza una pantalla holográfica. Su aproximación reconstruye todos los rayos de luz que están presentes en la visión 3D. Así presenta una imagen virtualmente continua a espectadores que son libres de moverse dentro de un área grande delante de la pantalla sin perder de vista la imagen 3D mostrada por el dispositivo. Para hacer esto, el dispositivo utiliza una matriz de micro-displays y una pantalla holográfica (Figura 24).



(a) SynthaGram 404



(b) SynthaGram 204

Figura 23: Dispositivos de Stereographics.

### 5.3.6. The Light Field Display

Ruigang Yang et al (2005) presentan un dispositivo autoestereoscópico que se compone de una matriz de proyectores digitales y una pantalla de proyección con una hoja de microlentes (Figura 24) similar a un lenticular. Los proyectores se utilizan para generar una matriz de píxeles de intensidad y color controlada que se proyectan sobre la pantalla. Cada lente transmite los rayos en diversas direcciones delante de la pantalla. En esencia, lo que hace este dispositivo es simular un campo de luz. Los dispositivos que acabamos de describir permiten, en la mayoría de los casos, visualizar una representación del campo de luz. Su funcionamiento consiste en visualizar varias imágenes desde una misma superficie 2D de visualización. Este tipo de dispositivos tienen la ventaja de que no son invasivos y además permiten múltiples espectadores.

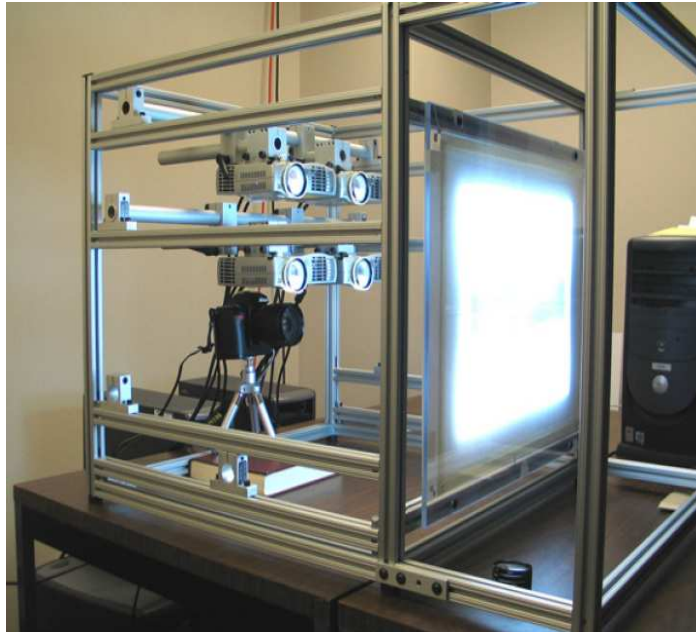


Figura 24: Prototipo de Light Field Display con 4 proyectores.

#### 5.4. Estereoscopia

El mecanismo de visión humana es estereoscópico por naturaleza. A través de la visión binocular somos capaces de apreciar las diferentes distancias y volúmenes en el entorno que nos rodea. Nuestros ojos, debido a su separación, obtienen dos imágenes con pequeñas diferencias entre ellas. Estas diferencias se denominan disparidad (ver figura 25).

Nuestro cerebro procesa esas diferencias entre ambas imágenes y las interpreta de forma que percibimos una sensación de profundidad en los objetos que nos rodean. Este proceso se denomina estereopsis. La distancia interpupilar más habitual es de 65 mm, pero puede variar desde los 45 a los 75 mm.

Durante las últimas décadas se ha investigado la creación de dispositivos estereoscópicos. Los objetos mostrados con estos dispositivos aparecen realmente delante del espectador. Para ello se usan, por ejemplo, gafas de conmutación o polarizadas. Estas gafas bloquean una de las dos imágenes para cada ojo. Además, existen otras técnicas que junto con las gafas se describen a continuación.

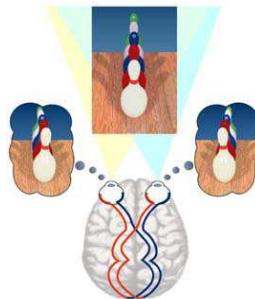


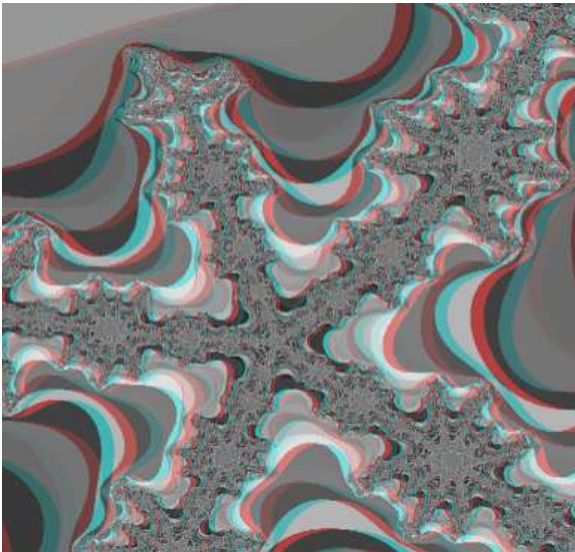
Figura 25: Nuestro cerebro recibe dos imágenes ligeramente diferentes.

##### 5.4.1. Anáglifos

En monitores convencionales una de las técnicas más utilizadas son los anáglifos. Las gafas que se necesitan son muy baratas y cualquier persona puede disponer de ellas. Las gafas utilizan filtros de

colores para separar las dos imágenes. Si vemos a través de un filtro rojo, los colores verde o azul se ven como negro. Si utilizamos un filtro verde, azul o cian, el rojo parece negro. A partir de este principio podemos mezclar dos imágenes en una foto o vídeo y utilizar lentes con filtros de color para separar las imágenes y ver el efecto estereoscópico.

Por convención, el filtro de color rojo se coloca frente al ojo izquierdo. El color del otro filtro depende del medio que se vaya a utilizar para visualizar la imagen. Si es un medio impreso se utiliza el azul. Para imágenes proyectadas y para vídeo se usa un filtro verde, porque es más brillante. Con estos filtros, la imagen aparece en blanco y negro. Otra variante utiliza un filtro rojo y otro cian. Con estos filtros la imagen 3D se puede ver en color. Puede verse un ejemplo de un anáglifo en la figura 26(a) y las gafas necesarias para verlo en la figura 26(b).



(a) Un fractal en anáglifo (imagen cortesía de Rafael Mullor).



(b) Gafas para ver anáglifos.

Figura 26: Anáglifos.

Las imágenes o vídeos producidos en anáglifo se pueden proyectar sin necesidad de equipo especial (como en las películas “Spy kids” y “El niño tiburón”). Y también se pueden distribuir en VCD o DVD para verse sin problemas en equipos domésticos. El problema es que el público tarda un poco en acostumbrarse a los filtros y los adultos no los puede utilizar más de 10 o 15 minutos seguidos.

Los niños pueden tolerarlo durante más de 1 hora, lo que dura una película de largometraje.

#### 5.4.2 ChromaDepth

ChromaTek Inc ofrece visualización estereoscópica a un precio similar a los anáglifos. Su sistema ChromaDepth™ se basa en la desviación que producen los diferentes colores del espectro. En un prisma, la luz se desvía ligeramente dependiendo de su longitud de onda: más desviación en el rojo, menos en el azul. La información de profundidad se codifica por colores. Las gafas especialmente diseñadas para ver estas imágenes disponen de unos cristales transparentes con microprismas. Cuando la imagen (ver figura 27), denominada CyberHologram™, se observa con las gafas HoloPlay™ (para imágenes de ordenador) o C3D™ (para imágenes impresas), la imagen 2D se convierte en tridimensional.

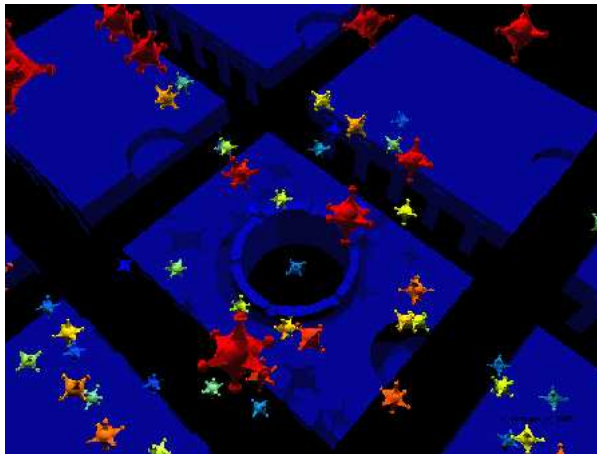


Figura 27: Imagen en ChromaDepth.

La desventaja de este sistema es la pérdida de información cromática. La ventaja sobre el anáglifo es que las imágenes pueden verse también en 2D.

#### 5.4.3. Gafas de Lentes Polarizadas

En la mayoría de los casos los anáglifos y el método de ChromaTek no permiten representar el color correctamente. Por ello se utilizan las gafas de lentes polarizadas que requieren una mayor inversión por necesitar dispositivos de visualización especializados. La técnica funciona en base a un fenómeno de la física llamado polarización de la luz. Si se proyecta luz polarizada en una dirección y la vemos con un filtro polarizado, colocando el filtro a una inclinación de 90 grados respecto a la luz original, toda la luz será bloqueada. Por ello podemos proyectar dos imágenes, una polarizada en un sentido y la otra a 90 grados y utilizar dos filtros polarizados para que cada ojo vea una imagen distinta.

Los filtros son relativamente baratos, el inconveniente es que sólo funciona con sistemas de proyección. Además, requiere generalmente dos proyectores o un proyector especialmente modificado, además de una pantalla que no despolarice la luz. También hay que considerar que los filtros polarizados oscurecen la imagen y se necesitan proyectores muy luminosos.

Por último, existe un problema con los proyectores actuales de vídeo, de cristal líquido (LCD), ya que estos polarizan la luz para funcionar. De este modo, al colocar los filtros polarizadores, se pierde todavía más luminosidad. Aun así, este método permite ver las imágenes en color y produce menos cansancio visual.

#### 5.4.4. Gafas de Conmutación

Hasta ahora hemos visto gafas pasivas donde estas no realizan ninguna tarea. Existe una técnica de autoestereoscopia que utiliza gafas activas, las gafas de conmutación (shutterglasses). Las gafas de conmutación (Figura 28(a)) consisten en cristales de cristal líquido que son capaces de oscurecerse por completo y no dejar pasar la luz. Las gafas van alternando rápidamente la apertura y el cierre del LCD delante de cada ojo (figuras 28(b) y 4(c)). Esto, junto con la proyección alternativa de las imágenes de cada ojo, permite que cada uno vea una imagen diferente, consiguiendo así la ilusión 3D (Figura 28).





(a) Gafas de conmutación.



(b) Imagen del ojo izquierdo.

(c) Imagen del ojo derecho.

(d) Imagen estéreo sin las gafas de conmutación.

Figura 28: Gafas de conmutación e imágenes generadas para utilizarlas.

Las gafas de conmutación son la opción más cara de las cuatro presentadas en esta sección. Requieren la sincronización de las gafas con el dispositivo de visualización y por ello precisan de hardware adicional. No obstante, son las que producen mejores resultados.

Los dispositivos estereoscópicos han tenido buena aceptación en entornos profesionales. Aun así no han sido tan aceptados por diseñadores gráficos, arquitectos o aficionados a los juegos por ordenador. Esto es debido a que los usuarios rechazan el uso de cualquier dispositivo invasivo que limite su agudeza visual, por ejemplo, las gafas. Estos estudios han motivado la investigación de dispositivos no invasivos que sean capaces de producir imágenes estereoscópicas sin que el espectador necesite llevar un equipamiento especial.

## 6 CONCLUSIONES

En este trabajo encontramos un estudio de cada una de las diferentes técnicas de adquisición, transmisión y visualización de escenas 3d que se han ido desarrollando a través de los tiempos después de múltiples esfuerzos y dedicación de diferentes grupos de investigación. La necesidad de resolver problemas en concreto ha llevado al desarrollo de diferentes técnicas pretendiendo encontrar la mejor solución para cada uno de ellos; esto ha contribuido a que hoy en día encontremos diversas opciones, entre las que podemos escoger, que se adecuen mas a nuestro intereses.

La visión, tal como la observamos en la naturaleza, es intrínsecamente activa, está íntimamente ligada a la acción. El objetivo final de la visión activa es dotar a un sistema autónomo móvil de la capacidad de percepción en tiempo real, para que éste pueda realizar tareas en entornos complejos y dinámicos. En cuanto a las aplicaciones de este área de la visión artificial, indicar que,

fundamentalmente, están enmarcadas dentro de los sistemas de robots guiados de manera visual, tales como navegación autónoma, inspección, vigilancia, detección-seguimiento de móviles, etc., donde el problema del procesamiento en tiempo real es particularmente relevante.

Después de haber hecho una investigación con respecto a las técnicas de adquisición, transmisión y visualización de escenas 3D se determinó:

Que las técnicas de adquisición menos utilizada en los sistemas de visión 3D son la telemetría láser y el control de parámetros ópticos y hasta el momento se dispone de poca información sobre estos métodos. Sin duda alguna el método más interesante y al cual se le están encontrando más aplicaciones hoy en día es la holografía, su inconveniente es que requiere de equipos muy costosos y conocimientos muy amplios, lo cual impide hasta el momento, que se pueda comercializar y hacer popular.

Que el codec estándar para vídeo digital que consigue una muy alta compresión de datos es el H.264/AVC, es compatible con el codec MPEG-4 además proporciona suficiente flexibilidad lo cual le permite ser aplicado a una gran variedad de aplicaciones. La intención del proyecto H.264/AVC fue la de crear un estándar capaz de proporcionar una buena calidad de imagen con tasas binarias notablemente inferiores a los estándares previos (MPEG-2, H.263 o MPEG-4 parte 2), además de no incrementar la complejidad de su diseño.

Que la técnica de visualización 3D más usada es la estereoscópica, ya que, imita el mecanismo de la visión humana, la principal desventaja de esta técnica es que requiere el uso de un equipamiento especial, tales como gafas de obturación y anaglifos, esto limita la agudeza visual. Por otra parte, la técnica de autoestereoscopia utiliza dispositivos que no requiere de ningún tipo de cristales o gafas y además muchos de ellos permiten que varios usuarios obtengan una vista distinta del mismo objeto 3D; su principal desventaja es que el espectador debe permanecer en una posición fija. Y por último las técnicas de visualización volumétrica utilizan dispositivos volumétricos los cuales permiten que los objetos o imágenes puedan ser vistos por varios observadores al mismo tiempo y desde cualquier orientación.

Como trabajo futuro se propone realizar un estudio más profundo sobre las técnicas de visualización autoestereoscópicas y volumétricas con el fin de lograr incursionar en el tema de realizar una visualización tridimensional sin necesidad de cristales o gafas que generan una visualización reducida y cansancio en los ojos y el impedimento de ser visto por varios usuarios a la vez.

## 7. ANEXO 1: ESTADO DEL ARTE

### ADQUISICION

FECHA	AUTORES	TITULO	DESCRIPCIÓN	METODO
2007	Joji Takei, Shingo Kagami, Koichi Hashimoto	3000 fps 3d shape measurement using a high speed camera projector system	Sistema de medida de forma 3d con frame de alto ratio. Consiste en un proyector DMD (Digital Micromirror Device) y una cámara de alta velocidad, ambos operando sobre los 6000 fps. Introduciendo un algoritmo de procesamiento de imágenes de detección y tracking, los patrones de la luz proyectada permite medidas de imágenes 3d frame por frame	Luz Estructurante
2007	Idaku Ishii, Kenkichi Yamamoto, Kensuke Doi	High speed 3D image acquisition using coded structured light projection	Propone un nuevo método de proyección de luz estructurada codificada que puede seleccionar, no solo, una codificación temporal sino también una codificación espacial adaptable para obtener imágenes tridimensionales a una rata alta de frame.	Luz Estructurada
2007	Jiajun Zhu, Greg Humphreys, David Koller	Fast Omnidirectional 3D scenes Acquisition UIT An Array Of Stereo Cameras	Se describe el diseño e implementación de un sistema, basado en una cámara, de adquisición de escenas 3D.El sistema usa una matriz de cámaras móviles digitales SLR (Single Lens Reflex) off the shelf para	Par estereo

			<p>capturar pares estereo de alto rango dinámico (HDR) sobre campos de 360° visión horizontal. Aumentando los pares estereo con patrones de luz proyectada y aplicando técnicas de reconstrucción estereo, se podrá generar alta resolución, rangos de imágenes exactas del entorno con perfecta correspondencia en la textura de color HDR.</p> <p>Este enfoque con frecuencia funciona mejor que los escáneres de láser para áreas con propiedades reflectivas cambiantes tales como bajo albedo, brillo o translucidez. Además la textura de la superficie capturada a partir de la visión estereo es perfectamente registrada con la superficie resultante.</p>	
2007	Michael Waschbüsch, Stephan Würmlin, Daniel Cottin, Markus Gross	Point Simple 3D Video Of Real World Scenes	<p>Presenta un enfoque de punto de muestreo para capturar cintas de video 3D y posteriormente volver a hacer escenas del mundo real. El sistema de adquisición esta compuesto de bloques de video 3D colocados esparcidamente. Los bloques contienen un proyector de bajo costo, dos cámaras a escala de grises, y una cámara a color con alta resolución. Para mejorar la profundidad del cálculo se confía en los patrones de la luz estructurada. La textura de las imágenes y los patrones de visión aumentada de la escena son adquiridos simultáneamente por el tiempo de proyección multiplexado de patrones complementarios y la exposición de las cámaras sincronizadas. Los mapas de profundidad de alta resolución son extraídos</p>	Luz Estructurada



			usando algoritmos de profundidad de estereo realizado en la adquisición de los patrones de las imágenes.	
2006	Sang Yun, Dongbo Min and Kwanghoon Sohn	3D Scene Reconstruction System with hand Held Stereo Cameras	Describe un sistema de reconstrucción de escena 3D que crea modelos 3D con múltiples partes de imágenes estereo adquiridas por un dispositivo hand held. Propone un algoritmo que consta de 2 partes. En la primera parte, se obtiene el mapa de profundidad con el estereo matching y la geometría de la cámara en cada vista. El algoritmo esta basado en un método de adaptación de ventana con un marco de trabajo jerárquico. En la segunda parte, se usa la característica SIFT para estimar el movimiento de la cámara. El algoritmo LMEDS reduce el efecto outliers en este proceso.	Par Estereoscopico
2006	S. Ledesma, C Heman, I. Sidelnik	Hologramas Digitales De Fase Con información Estereoscópica Para Visivo Tridimensional.	Se ha desarrollado una forma de representar imágenes estereoscópicas a través de hologramas de Fourier que requieren un solo paso en su generación. El holograma retiene la información de fase de la transformada de Fourier de la escena de entrada modificada y contiene además una fase lineal para lograr el efecto estereoscópico.  La escena modificada consiste en la escena original a la que se agrega una fase aleatoria para garantizar una mejor distribución de la información en la fase de la transformada de Fourier. Para implementar el holograma se ha utilizado un modulador de luz espacial	Holografía

			operando en distintas configuraciones. Se muestran resultados experimentales donde se aprecia la reconstrucción estereoscópica y se discuten además las limitaciones del dispositivo.	
2005	Bahram Javidi, Fellow and Seung Hong	Three Dimensional Holographic Image Sensing And Integral Imaging Display	Se propone una reconstrucción por sensado holográfico y por imagen integral 3D computacional/óptica. Se demuestra experimentalmente que a través de técnicas de imagen integral, es posible reconstruir una escena 3D la cual ha sido obtenida por medio de hologramas digitales. El enfoque propuesto tiene la ventaja de la alta resolución del sensado holográfico y la visualización robusta de imagen 3D integral. Se demostró que con los hologramas digitales una secuencia de imágenes holográficas digitales puede ser generada a la distancia de interés y que por medio de geometría óptica estas son computacionalmente convertidas a imágenes elementales usando un array pinhole virtual.	Holografía
2005	Marcelo Bernandes Vieira, Luis Velho, Asla Sa	A Camera Projector System For Real Time 3D Video	Este sistema consiste en un dispositivo de adquisición, procesamiento de datos y modelos de visualización. Esto genera video 3D en tiempo real (30fps) a partir de una gran variedad de escenas. El dispositivo de adquisición emplea estereo activo y está compuesto de cámaras de video y un proyector calibrados y sincronizados. El módulo de procesamiento de datos extrae información de profundidad del código de	Luz estructurante

			luz estructurante. El modulo de visualización hace video 3D usando la geometría inducida por el código de color usado. En este artículo se enfatiza en el aspecto de captura de video 3D del sistema usando una cámara/proyector.	
2005	Katsuya Kondo, Syoji Kobashi and Yukata Hata	A Real Time Analysis Of 3D Scene From Monocular Images By Observing Known Background	Propone un método en tiempo real para analizar escenas tridimensionales de objetos a partir de serie de imágenes monoculares. Una cámara simple observa algunos puntos sobre el objeto tal que la posición 3D de los puntos puede ser estimada por un filtro de Kalman extendido. Se aplica este método a dos aplicaciones en tiempo real. Uno es adquirir la geometría 3D de un objeto, y el otro es estimar la posición 3D del objeto. Este enfoque no necesita datos del modelo a priori del objeto y logra la estimación de la geometría 3D de la posición.	Multiples cámaras
2003	X. Wu and T. Matsuyama	Real Time Active 3D shape Reconstruction For 3D Video	Se enfoca en la reconstrucción en tiempo real de la forma 3D de un objeto en movimiento en una área amplia. Para registrar eficientemente el cambio dinámico de la forma de un objeto 3D se requiere la construcción en tiempo real de la forma 3D y un tracking activo. Para la reconstrucción de la forma 3D se propone un método de intersección de volumen basado en un plano paralelo. Para obtener la forma de un objeto 3D en movimiento en un área amplia, se introduce un tracking activo al objeto por múltiples cámaras. Para realizar la reconstrucción activa de la forma 3D, se aumenta el método de intersección de	Multiples cámaras

			volumen de tal manera que puede ser aplicada a esas imágenes capturadas por cámaras activas.	
2003	Salah Derrouich, Kiichiro Izumida, Kenji Murao	The Use Of CNN Models And Vertical Rectification For a Direct Trigonometric Recovery Of 3D Scene Geometry From a Stream Of Images	Se presenta una recuperación trigonométrica directa de la geometría de una escena 3D a partir de una secuencia de imágenes. Estas imágenes, son rectificadas verticalmente usando salidas giroscópicas para minimizar la razón relativa de desplazamiento entre cada frame. Para que coincida esta secuencia de imágenes rectificadas verticalmente en tiempo real se pide prestado la fuerza del CNN (Cellular Nonlinear Network) y los resultados de coincidencia son trigonómicamente procesados por el rango de estimación 3D.	Múltiples cámaras
2002	Stephan Wurmlin, Edouard Lamboray, Oliver Stadt	3D Video Recorder	Es un sistema capaz de registrar, procesar y poner en marcha video tridimensional a partir de múltiples puntos de vista. Primero se registra video 2D a partir de varias cámaras de video digital sincronizadas y se almacena imágenes preprocesadas en el disco. Una etapa de procesamiento offline convierte estas imágenes en una estructura jerárquica tridimensional y almacena este video 3D en el disco. Se muestra como se puede compensar la calidad de video 3D con el rendimiento del proceso e idear una compresión eficiente y esquemas codificados para la representación del video 3D.	Múltiples cámaras
2000	M Pollefeys, R Koch, M vergauwen, L Van Gool	Automated reconstruction of 3d scenes from sequences of images	Presenta un sistema que recupera un modelo de superficie 3D de una secuencia de imágenes tomada con una cámara de	Cámaras de perspectiva llena, no requiere

			consumación off-the-shelf. El usuario adquiere las imágenes moviendo la cámara libremente alrededor del objeto. el modelo 3D obtenido es una versión a escala del objeto original y también se obtiene el modelo de la superficie de textura.	modelos y calibración a priori.
1998	Andrew Fitzgibbon and Andrew Zisserman	Automatic 3D Model Acquisition And Generation Of New Images From Video Sequences	La escena que tiene que ser modelada es registrada, desde diferentes puntos de vista, por un video cámara. La posición relativa y la orientación de la cámara y sus parámetros de calibración serán automáticamente recuperados de los datos de la imagen, después por los algoritmos. Por lo tanto, no hay necesidad de medidas de la escena o procesos de calibración alguna. Tampoco hay restricciones en rango, es tan fácil modelar un objeto pequeño como modelar un edificio completo.	Cámara móvil
1997	Huaibin Zhao and J. Aggarwal	3D Reconstruction of an Urban Scene From Synthetic Fish-Eye Images	El propósito es generar un modelo 3D de una escena urbana con suficiente detalle y exactitud por medio de un par estéreo o una secuencia de imágenes tomadas por un lente fish-eye de 180°. Se desarrolla un algoritmo de triangulación teórica para calcular las coordenadas 3D de un punto característico conociendo su mapeo en un par estereo. Se demostró experimentalmente la eficacia del algoritmo.	Par Estereoscopico
1995	Nelson L. Chang and Avideh Zokhor	Arbitrary View Generation For Three Dimensional Scenes From Uncalibrated Video Cameras	Este artículo se enfoca en la representación y generación de vistas arbitrarias de escenas tridimensionales. Esta representación	Multiples cámaras

			<p>consiste en densos mapas de profundidad en muchos puntos de vista preseleccionados a partir de una secuencia de imágenes. Además en vez de usar múltiples cámaras estacionarias calibradas, se deriva los mapeos de profundidad a partir de secuencia de imágenes capturadas por una cámara descalibrada. Una vez los mapas de profundidad son calculados en los puntos de vista preseleccionados, la intensidad y la profundidad en estos lugares son usadas para reconstruir vistas arbitrarias de la escena 3D.</p> <p>El enfoque propuesto consiste en escanear una video cámara a través de varias trayectorias de la escena para generar secuencia de imágenes para ser usadas en la construcción de los mapas de profundidad en varios lugares preespecificados, llamados “referente frames” usando sus frames vecinos capturados. Una vez, la profundidad y la intensidad en los frames de referencia son guardados como una representación compacta de la escena. Esta representación es entonces usada para reconstruir vistas arbitrarias localizadas dentro o fuera de las trayectorias escaneadas.</p>	
1995	Adrian Hilton and G. Gonçaves	3D Scene Representation using a Deformable Surface	Se presenta un método basado en superficie deformable (DS) para construir representaciones de escenas 3D reales a partir de múltiples gamas de imágenes. El modelo DS utiliza superficies físicas y restricciones de arista para reconstruir una representación a partir de los datos del ruido	Múltiples cámaras

			3D. la representación construida identifica regiones de superficie consistente y discontinuas las cuales se ajustan al dato 3D dentro de tolerancias conocidas. La fusión de múltiples gamas de imágenes permite una reconstrucción de la escena 3D completa. Una representación multiescala es obtenida lo cual proporciona un orden de reducción de la magnitud de los datos.	
1990	L. Hong and D. Brzakovic	An Approach to 3D Scene Reconstruction From Noisy Binocular Image Sequences Using Information Fusion	El sistema de visión descrito en este paper reconstruye escenas 3D a partir de secuencia de imágenes binoculares ruidosas. Primero, el sistema establece todas las posibles coincidencias entre los píxeles característicos y el primer par de imagen binocular, y asigna un valor confidencial a una posible coincidencia. Cada posible coincidencia es asociada con un volumen 3D las posibles coincidencias son usadas para producir proyecciones de volúmenes 3D asociados en el resto de pares de imágenes binoculares. Estas proyecciones son utilizadas para limitar la búsqueda de posibles coincidencias. Las nuevas coincidencias son usadas para actualizar, valores confidenciales un algoritmo de fusión óptimo. Finalmente, los píxeles que coincidieron junto con los valores de alta confidencialidad son usados para reconstruir una escena 3D.	Imágenes binoculares
	Szymon Rusinkiewicz, Olaf Hall Holt, Marc Levoy	Real Time 3D Model Acquisition	Se propone un nuevo sistema de adquisición de modelo 3D que permite al usuario rotar un objeto con la mano y ver un continua actualización del modelo mientras el objeto	Visión Activa

			es escaneado. Este lazo de realimentación permite al usuario encontrar y reparar huecos en el modelo en tiempo real, y determinar cuando el objeto es realmente cubierto.	
	Vitor Sequeira, Joao Gonçalves, Isabel Ribeiro	3D Environment modelling using laser range sensing	Describe una técnica para construir un modelo volumétrico 3D de un entorno desconocido basado en datos adquiridos por un Láser Range Finder de un robot móvil.	Telemetría Láser

#### TRANSMISION

FECHA	AUTORES	TITULO	DESCRIPCION	METODO
2007	Panagiotis Papadimitrio and Vassilis tsaoussidis	SSVP: A Congestion Control Scheme For Real Time Video Streaming	Presenta un nuevo protocolo end to end llamado Scalable Streaming Video Protocol (SSVP), el cual opera sobre UDP y es optimizado por aplicaciones unicast de video streaming. SSVP emplea un control de congestión AIMD y adapta la tasa de envío ajustando adecuadamente la IP6.	
2007	Sung Yeol Kim, Jongeun Cha, Seung Lee	3DTV System using Depth Image-Based video In The MPEG-4 Multimedia	Se presenta un sistema 3DTV usando una nueva representación basado en profundidad de imagen (DIBR). Después de obtener	H.264



		Framework	imágenes con profundidad a partir de cámaras multivista o cámaras de rango profundo, se descompone la profundidad de imagen en tres imágenes de capa disjunta y una imagen descriptora de capa. Después, se combinan las imágenes descompuestas con imágenes a color para generar una nueva representación de escenas 3D dinámicas, llamada video con profundidad 3D. Este es comprimido por un codificador H.264/AVC y enviado a clientes sobre redes IP.	
2007	Nukhet Ozbek, Burak Gorkemli, A Murat and Turhan Tunali	Adaptive Streaming Of scalable Estereoscopic Video Over DCCP	Se propone un Nuevo modelo adaptable de streaming que utiliza DCCP (Datagram Congestion Control Protocol) para hacer stream de video estereoscopico sobre internet para transporte de 3DTV. El modelo asigna el canal disponible de ancho de banda, el cual es calculado por el CDP. La tasa de video es adaptada a la tasa del CDP para cada grupo de imágenes (GoP) por medio de la extracción de capas de un bitstream multivista scalable.	DCCP
2006	A. Aggoun and M. tabit	Data Compresión Of Integral Images For 3D TV	Presenta un enfoque del problema de comprimir una cantidad significativa de datos requeridos para presentar video integral 3D. El algoritmo esta basado en la transformada discreta del coseno 3D (3D DCT). Este toma ventaja de la alta correlación entre las subimágenes generadas por el arreglo macrolente y toma ventaja de la representación de los puntos de vistas de la imagen unidireccional para decrementar el numero de bits requeridos para codificar los vectores de movimiento.	Transformada discreta del coseno

2005	Baochun Bai, Pierre Boulanger and Janelle Harms	An Eficiente Multiview Video Compresión Scheme	Se presenta aun esquema de compresión de video multivista utilizando MPEG-2 LIKE, el cual también clasifica streams de video dentro del principal stream y el stream secundario. Un streamj es escogido como el stream principal y es codificado usando solo un algoritmo MPEG-2 LIKE basado en MCP. El stream secundario es comprimido usando DCP a partir de múltiples vecinos cercanos y MCP.	MPEG-2 LIKE
2005	Jianwei Niu, Rui He, Jianping Hu	MPEG-4 video Encoder Based On DSP-FPGA Techniques	Diseña e implementa un encoder de videdo MPEG-4 basado en técnicas coordinadas DSP-FPGA. El modulo FPGA toma las tareas de adquisición de video, separación de YUV y los datos de funciones de entrada y salida, mientras que el DSP esta dedicado a la compresión de video MPEG-4 para utilizar la memoria on chip DSP.	MPEG-4
2003	Hayrettin Genel and Y. Murat Erten	End to End QoS control scheme For Video Transmission Over Internet	Se implementa un framework integrado de transmisión de video end to end usando un bajo ratio de bit de video codificado (H.263+) y el protocolo de red estándar en tiempo real (RTP-RTCP) y esquemas de transmisión apropiados utilizando aplicaciones de nivel de calidad de técnicas de servicio de control.	H.263
2001	Haitao Zheng and Jill Boyce	An Improved UDP Protocol For Video Transmission Over Internet-To-Wireless Networks	Propone un Nuevo protocolo de datagrama de usuario completo (CUDP) el cual utiliza un canal de información de error obtenido de links físicos para ayudar a la recuperación del error en el nivel de paquetes. Propone varios niveles de paquetes basados en código con máxima distancia separable (MDS), un sistema de	MPEG-CODE

			codificación de control de errores y obtener formulas analíticas para estimar las perdidas equivalentes de frame de video para diferentes versiones del protocolo de datagrama de usuario (UDP). Se valida el paquete de codificación propuesto y el protocolo CUDP usando MPEG-CODE bajo varias perdidas de paquetes de Internet y perfiles de canal inalámbrico.	
1999	Jill M. Boyee	Packet Loss Resilient Transmission Of MPEG Video Over The Internet	Se propone un método el cual mejora la calidad del video recibido con respecto al video MPEG transmitido sobre redes IP con pérdida de paquetes. En el método de alta prioridad de protección (HIPP), el reenvío de la corrección de error codificado Rced Salomón es aplicado solo con una alta prioridad de dato de video para proporcionar protección a la calidad de video recibido con baja over head. El método HIPP no requiere soporte de red para la priorización y puede ser usado con encoders y decoders de video estándar.	MPEG
1998	Bor Prihavec And Franc Solina	User Interface for video Observation Over the Internet	Se ha construido el Internet video Server (IVS) y su extensión de visión global el cual genera un imagen panorámica de todo el entorno, y su interface de usuario, el usuario gana un mejor entendimiento de la posición observada y un control más intuitivo de la cámara. Para ver toda la escena usando solo IVS se necesita al menos 100 movimientos de cámara. Por medio de la extensión de visión global el número de movimientos de la cámara requeridos para localizarla y dirigirla al algún objetivo puede ser	

			disminuido a 1.	
		Representation, Coding, And Rendering Of 3D Video Objects With MPEG-4 and H.264/AVC	Se presenta un sistema completo para la extracción, representación, codificación y renderizado interactivo de video de objeto 3D. La representación de datos esta basada en modelos de mallas 3D y mapeos de textura usando texturas de video. La extracción de la geometría esta basada en algoritmos de forma a partir de la silueta. Los modelos de voxeles resultantes son convertidos en mallas 3D que son codificados usando herramientas de MPEG-4 SNHC. Las texturas de video correspondientes son preprocesadas tomando en cuenta la forma del objeto y codificarlo usando un codificador H.264/AVC	H.264/AVC

#### VISUALIZACION

FECHA	AUTORES	TITULO	DESCRIPCION	METODO
2006	Ana Claudia M.T.G. Oliveira, Fátima L.S. Nunes, Larissa Pavarini,	Virtual reality framework for medical training: implementation of a deformation class using java	La deformación de objetos es un factor importante y una necesidad para que el entrenamiento y la simulación médica obtengan mayor realismo. Es por ello que Ana Cláudia M. T. G. de Oliveira propuso el desarrollo de un sistema de estereoscopia basado en la técnica de los anáglifos y enfocado a la deformación de los objetos cuando existe una colisión. La técnica de deformación en objetos tridimensionales fue puesta en práctica con la ayuda de la API de	Técnica de Anaglifo

			<p>Java 3D y su desarrollo basado en el método masa-resorte, teniendo como objetivo proveer una clase que se utilizará en las herramientas de entrenamiento médico. El método de los anáglifos fue elegido para poner el módulo de estereoscopia en práctica. El proyecto estereoscópico tiene un grupo de clases y métodos jerárquicos que pueden separar las imágenes correspondientes a las visiones del ojo izquierdo y derecho, atribuyendo los colores respectivos de los lentes usados para la visualización. El marco desarrollado contiene apenas las clases para la detección, la estereoscopia y la deformación de objetos cuando se presenta una colisión, pero se necesita cubrir todas las características pertinentes en el dominio de un tema como lo es el entrenamiento médico.</p>	
2005	Leonardo C. Botega, Ana Claudia M.T.G. Oliveira, Larissa Pavarini, Fátima L.S.Nuñes	Implementación de módulo de estereoscopia utilizando anáglifos para herramientas de realidad virtual para tratamiento médico	Leonardo C. Botega, también utilizó visión estereo pasiva, el objetivo de este trabajo fue implementar un módulo de estereoscopia, que fuera capaz de proporcionar interacción y sensaciones realistas para lograr mayor inmersión en las aplicaciones de entrenamiento médico con técnicas de anáglifo. Para lograr su objetivo se construyó un atlas virtual de anatomía para cáncer de mama. La aplicación se desarrolló con el API de Java3D que fue la encargada de realizar la carga de los objetos virtuales en la representación. Los objetos fueron modelados con la herramienta 3D	

			<p>Estudio Max y posteriormente exportados y anexados al atlas virtual. La importación de los objetos para el API de Java3D puede ser un problema, debido a la dificultad encontrada para modificar las cualidades en los modelos importados; como la transparencia y color. Otro problema encontrado es que la herramienta no funciona correctamente con los tonos rojos y azules por lo que se tuvo que hacer coincidir con los colores de los lentes adaptándose a las condiciones que el Anaglifo requiere.</p>	
2003	W.Kresse D.Reiners and C. Knaope	Color consistency for digital multiprojector stereo display systems	<p>En la actualidad se han diseñado e implementado diversos sistemas de visión estereoscópica, en su gran mayoría han empleado proyectores digitales de LCD, a pesar de las ventajas que ofrecen este tipo de proyectores presentan una desventaja significativa como lo es la consistencia del color. Tomando en cuenta las ventajas que aportan los proyectores de LCD, W. Kresse, desarrolló el HEyeWall y la cueva digital. El HEyeWall es la primera aplicación con capacidad de despliegue estéreo y la cueva digital es la primera cueva con proyectores digitales. En este trabajo describe un método para la corrección del color y del brillo de los sistemas de visualización de multiprojector con el propósito de hacer uniforme el color en las pantallas. La ventaja principal de esta tecnología es que existe una buena separación de imagen.</p>	Heye Wall

			Además no hay requisitos especiales para las pantallas.	
2002	Simon Stegmaier, Thomas Ertl, Dirc Rose	A case study on the applications of a generic library for low-cost polychromatic passive stereo	Simón Stegmaier desarrolló un software genérico por medio de una biblioteca de OpenGL basado en la visión estéreo pasiva policromática de alta calidad y de bajo costo. Por medio de la precarga de la biblioteca se agregan capacidades estéreas. En cuanto a la configuración, los pares estéreos son controlados por tres parámetros importantes: los ojos del espectador, la separación que existe entre los ojos y la anchura de la exhibición. Una limitación de este trabajo es que la biblioteca, pues solamente puede ser aplicada a los programas que se ligan dinámicamente a las bibliotecas de OpenGL.	

## 8 BIBLIOGRAFIA

Actuality Systems, Inc, <http://www.actuality-systems.com/>

Aggoun, A., and Tabit M. (2006). Data compression of Integral images for 3D TV. School of Engineering and Design, Brunel University, Uxbridge, UK.

Arrebola, F., Camacho, P., Sandoval, F. (2000). Sistemas de visión activa una aproximación a la visión biológica.

Bajcsy, R.(1988). Active perception. Proceeding of the IEEE, vol. 76, nº 8:996-1005.

Baochun, Bai, Pierre, Boulanger and Janelle, Harms.(2005). An efficient multiview video compression scheme. IEEE International Conference on Multimedia and Expo: 836-839.

Bernandes, Vieira, Marcelo, Velho, Luis, Sa, Asla , A.(2005). Camera projector system for real time 3D video. Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops - Volume 3:96.

Botega, Leonardo, C., Oliveira, Ana, Claudia, Pavarini, Larissa, Nuñez, Fátima(2005). Implementación de módulo de estereoscopia utilizando anaglifos para herramientas de realidad virtual para tratamiento médico.

Boyce, Jill, M.(1999). packet loss resilient transmission of MPEG video over the internet. Signal Processing: Image Communication, Volume 15, Number 1, pp. 7-24.

Chang, Nelson, L. and Zokhor, Avideh.(1995). Arbitrary view generation for three dimensional scenes from uncalibrated video cameras. Department of Electrical Engineering and Computer Sciences University of California, Berkeley, CA,USA.

ChromaTek Inc., <http://www.chromatek.com/>.

Derrouich, Salah, Izumida, Kiichiro, Murao, Kenji. (2003). The use of CNN models and vertical rectification for a direct trigonometric recovery of 3D scene geometry from a stream of images. Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers, 2003

Felix 3D, <http://www.felix3d.com/>.

Fitzgibbon, Andrew, and Zisserman, Andrew.(1998). Automatic 3D model acquisition and generation of new images from video sequences. Dept. of Engineering Science, University of Oxford,19 Parks Road, Oxford OX1 3PJ, UK.

FogScreen Inc., <http://www.fogscreen.com/>

Genel, Hayrettin, and Murat, Erten, Y.(2003). End to end QoS control scheme for video transmission over internet. IEEE Computer Society Washington, DC, USA.

Genex Technologies, Inc., <http://www.genextech.com/>.



Granero, L., De Gracia, Vicente.(2002). Técnicas de digitalización tridimensional basada en luz estructurada. Departamento de Desarrollo de Producto Unidad de Ingeniería de Producto Sección de Digitalización Tridimensional AIDO. Instituto Tecnológico de Óptica, Color e Imagen Paterna.

Haitao, Zheng, and Jill, Boyce.(2001). An improved UDP protocol for video transmission over internet-to-wireless networks.

Herrero, Ruth, Lloret, Jaime, Jiménez, José. (2004). Estudio de codecs de compresión MPEG-4 para su aplicación a la videovigilancia. Departamento de Comunicaciones Universidad Politécnica de Valencia.

Hilton, Adrian, and Gonçalves, G.(1995). 3D scene representation using a deformable surface. Physics-Based Modeling in Computer Vision, Proceedings of the Workshop on Cambridge, MA, USA.

Hong, L., and Brzakovic, D. (1990). An approach to 3D scene reconstruction from noisy binocular image sequences using information fusion. Computer Vision, Proceedings, Third International Conference on Osaka, Japan, pp. 658-661.

IO2 Technology LLC, <http://www.io2technology.com/>

Javidi, Bahram, Fellow, and Seung, Hong.(2005). Three Dimensional Holographic Image Sensing and Integral Imaging Display. Journal of display technology, vol. 1, no. 2.

Jiménez, L., Aracil, R., Peñin, L., Ferre, M.(199). Torres, F.(199). Modelado de entornos 3D para planificación de trayectorias de robots mediante visión estereoscópica y control activo de parámetros ópticos. Div. Ingeniería de sistemas y Automática (DISAM), E.T.S.I. Industriales, Universidad Politécnica de Madrid.

Kondo, Katsuya, Kobashi, Syoji, and Hata, Yukata, A.(2005). Real time analysis of 3D scene from monocular images by observing known background. Proceedings of International Symposium on Intelligent Signal Processing and Communication Systems, pp 5-8.

Kresse, W., Reiners, D., and Knaepe, C.(2003). Color consistency for digital multiprojector stereo display systems. International Conference Proceeding Series; Vol. 39 Proceedings of the workshop on Virtual environments, pp 271-279.

Ledesma, S., Heman, C., Sidelnik, I. (2006). Hologramas digitales de fase con información estereoscópica para visión tridimensional. Laboratorio de Procesado de Imágenes, Departamento de Física, FCEyN, UBA.

LightSpace Technologies Inc., <http://www.lightspacetech.com/> March 2005.

Marr, D. (1982). LA VISIÓN. Ed. Freeman, 1982.

Medioni, G., Nevatia, R.(1985). Segment-based stereo matching computer vision, Graphics, and Image Processing, 31:2-18.

Oliveira, Ana, Claudia, Nuñez, Fátima, Pavarini, Larissa.(2006). Virtual reality framework for medical training: implementation of a deformation class using java. Proceedings of the 2006 ACM international conference on Virtual reality continuum and its applications, pp 347-351.

- Ozbek, Nukhet, Gorkemli, Burak, Murat A., and Tunali, Turhan(2005). Adaptive Streaming Of Scalable Estreoscopic Video Over DCCP.
- Panagiotis, Papadimitrio, and Tsaoussidis, Vassilis.(2007). SSVP: A congestion control scheme for real time video streaming. *Computer Networks: The International Journal of Computer and Telecommunications Networking*, pp 4377-4395.
- Parker, E., and Wallis, P., A. (1948). Three-dimensional cathode-ray tube displays. *J. IEEE*, 1948, pp.371-390.
- Pollefeys, M., Koch, R, Vergauwen, M., Van, Gool, L.(2000). Automated reconstruction of 3d scenes from sequences of images. *ISPRS Journal of Photogrammetry and Remote Sensing*, Volume 55, Number 4, November 2000, 251-267.
- Prihavec, Bor, and Solina, Franc.(1998). User interface for video observation over the internet. *Journal of Network and Computer Applications*, 21, 219–237.
- Quesada, L.Compresión de Video y Control de BitRate; <http://www.elzeta.com>.
- Ruigang Yang, Shunnan Chen, Xinyu Huang, Sifang Li, Liang Wang, and Christopher Jaynes. Towards the Light Field Display. *IEEE VR 2005 Workshop on Emerging Display Technologies*.
- Rusinkiewicz, Szymon, Hall, Holt, Olaf, Levoy, Marc.(2002). Real time 3D model acquisition. *ACM transactions on Graphics*, pp: 438 - 446.
- Sanyo Electric Co. Ltd., <http://www.sanyo.com/>.
- Sequeira, Vitor, Gonçalves, Joao, Ribeiro, Isabel. (1995). 3D environment modelling using laser range sensing. *Robotics and Autonomous Systems*, 16(1), pp 114-127.
- Shii, Idaku, Yamamoto, Kenkichi, Doi, Kensuke. (2007). High speed 3D image acquisition using coded structured light projection. *International Conference of Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ* pp.925-930
- Smolic A, Muller K., Merkle P.(2004). Representation, coding, and rendering of 3D video objects with MPEG-4 and H.264/AVC. *IEEE 6th Workshop on Multimedia Signal Processing*, pp 379-382.
- Stegmaier, Simon, Ertl, Thomas, Rose, Dirc(2002). A case study on the applications of a generic library for low-cost polychromatic passive stereo. *IEEE Visualization 2002 Oct. 27 - Nov. 1, 2002, Boston, MA, USA*.
- Stereographics Ltd, <http://www.stereographics.com/>.
- Takei, Joji, Kagami, Shingo, Hashimoto, Koichi.(2007). 3000 fps 3d shape measurement using a high speed camera projector system. *Proceedings of the IEEE/RSJ International Conference of Intelligent Robots and Systems San Diego, CA, USA*.
- Waschbüsch, Michael, Würmlin, Stephan, Cottin, Daniel, Gross, Markus.(2007). Point simple 3D video of real world scenes. *Elsevier Science Inc. New York, USA, Volumen 22*, pp 203-216.
- Würmlin, Stephan, Lamboray, Eduard, Staadt, Oliver.(2002). 3D video recorder. *Computer Science Department, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland*.

Wu, X., and Matsuyama, T.(2003). Real time Active 3D shape reconstruction for 3D video. Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis.

Yeol, Kim, Sung, Cha, Jongeun, Lee, Seung.(2007). 3DTV system using depth image-based video in the MPEG-4. Multimedia Framework.

Yun, Sang, Dongbo, Min, and Kwanghoon, Sohn. (2006). 3D Scene Reconstruction System with hand Held Stereo Cameras, 2006.

Zhao, Huaibin, and Aggarwal, J.(2000). 3D Reconstruction of an Urban Scene from Synthetic Fish-Eye Images. Proceedings. 4th IEEE Southwest Symposium pp 219-223.

Zhu, Jiajun, Humphreys, Greg, Koller, David.(2007). Fast Omnidirectional 3D scenes Acquisition Using an Array of Stereo Cameras, 2007.

3DZ, <http://www.3dz.co.uk/>.