ESCUELA POLITÉCNICA NACIONAL

ESCUELA DE INGENIERÍA

ESTUDIO DEL SISTEMA DE TELEVISIÓN ESTEREOSCÓPICA COMO UNA APLICACIÓN DE LA TELEVISIÓN DIGITAL

PROYECTO PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO EN ELECTRÓNICA Y TELECOMUNICACIONES

WILLIAM RAMIRO PEÑAHERRERA HERRERA
FAN ALÍ VALVERDE VALAREZO

DIRECTOR: ING. TANIA PEREZ RAMOS

Quito, Febrero 2002

DECLARACIÓN

Nosotros, William Ramiro Peñaherrera Herrera y Fan Alí Valverde Valarezo, declaramos que el trabajo aquí descrito es de nuestra autoría; que no ha sido previamente presentada para ningún grado o calificación profesional; y, que hemos consultado las referencias bibliográficas que se incluyen en este documento.

La Escuela Politécnica Nacional, puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley, Reglamento de Propiedad Intelectual y por la normatividad institucional vigente.

William Peñaherrera H.

Fan Alí Valverde V.

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por William Ramiro Peñaherrera Herrera y Fan Alí Valverde Valarezo, bajo mi supervisión.

Ing. Tania Pérez Ramos
DIRECTORA DE PROYECTO

AGRADECIMIENTO

Agradezco principalmente a Dios por haberme dado la fortaleza necesaria para superar las pruebas que hasta hoy me ha puesto la vida. A mis compañeros y amigos por el apoyo moral que me han brindado durante toda mi carrera, sin el cual me hubiese sido más difícil el culminar la misma. A mis tíos que siempre me han respaldado de una u otra forma y un agradecimiento muy especial a mi mejor amiga, mi madre, que me ha apoyado y lo sigue haciendo en todo sentido, gracias a sus sacrificios y abnegación me ha legado la mejor de las herencias, una buena educación que me servirá para defenderme en esta vida.

Gracias a la Politécnica Nacional y a todos mis maestros por permitirme llegar hasta aquí y hacer realidad uno de mis sueños.

Gracias a todos.

William.

DEDICATORIA

El presente trabajo significa la culminación de una etapa más y quiero dedicarlo a mi madre quien no solo me ha dado la vida sino también me ha enseñado a vivirla de la mejor manera. Para ti mi María.

William

AGRADECIMENTO

Una vez culminado este trabajo le agradezco a Dios por mantenerme con vida hasta la etapa final del mismo. A mis padres ya que sin su sacrificio y apoyo incondicional hubiera sido imposible culminar esta etapa de mi vida; a mis hermanos, esposa, familiares y amigos que de una u otra manera me apoyaron.

Un agradecimiento especial a la Ing. Tania Pérez, ya que debido a su acertada dirección esta tesis es una realidad.

Fan Alí

DEDICATORIA

Este trabajo es dedicado a mis padres, Elena y Orlando. A mis hermanos, esposa, sobrinos y especialmente a mi princesa querida Paula Anahí.

Fan Alí

RESUMEN

Como es bien conocido la televisión es un medio masivo de comunicación que ha llegado a formar parte de nuestra cultura y de nuestro que hacer diario, debido a que es una fuente de información y entretenimiento que se encuentra masivamente difundido.

Con el avance que ha tenido la televisión digital en estos últimos tiempos y poniendo de manifiesto el inminente cambio de analógico a digital que tarde o temprano se llevara a cabo en nuestro país, nos hemos propuesto realizar el estudio de una de las aplicaciones importantes de la televisión digital, como lo es la televisión estereoscópica.

El presente trabajo comprende un estudio de las técnicas de transmisión de las señales televisión estereoscópica, de sus estándares y tendencias mundiales, así como de los fundamentos para la compresión de este tipo de señales.

Este proyecto esta orientado a profesionales y estudiantes que se encuentren de una u otra forma relacionados con el ámbito del video y que tengan interés por la obtención de imágenes tridimensionales.

Este trabajo se divide en seis capítulos que tienen como contenido fundamental lo siguiente:

Con los fundamentos teóricos de digitalización de la señal de televisión, como son el muestreo, cuantización y codificación se pretende dar una introducción a la Televisión Digital y sus principales sistemas se describen en el capítulo I. A continuación se describirán los conceptos principales e historia de la

estereoscopia, antecedentes de la Televisión Estereoscópica, y los principales métodos de visualización.

En el Capítulo II se detalla los proyectos de mayor importancia de la Televisión Estereoscópica, así como sus principales aplicaciones.

En el capítulo III se describen los procesos utilizados en la generación de la señal de Televisión Estereoscópica, los cuales se los ha divido de la siguiente forma: captación de la imagen, compresión de las señales digitales estereoscópicas y despliegue de las imágenes.

Los principales estándares para la transmisión de señales estereoscópicas que pueden ser imágenes fijas o secuencia de imágenes se exponen en el capítulo IV.

El capítulo V describe una pequeña pero significativa muestra de los productos existentes en el mercado, tratando de en lo posible incluir precios de los equipos y sus características técnicas mas representativas.

En el capítulo VI se ponen a consideración algunos de los aspectos más importantes a los que se ha llegado al término de este trabajo y se realizan algunas sugerencias para estudios posteriores.

PRESENTACIÓN

La televisión desde sus inicios ha ido adquiriendo mayor capacidad de difusión y aceptación por parte de los usuarios hasta convertirse en lo que es hoy en día, un auténtico medio de masas.

Hasta hace poco, la totalidad de transmisiones televisivas se las realizaba en forma analógica pero con el avance tecnológico y la digitalización de las señales por sus conocidas ventajas, como uso eficiente del espectro radioeléctrico lo que se traduce en aumento del número de canales, alta resolución y calidad, están haciendo que esta sea la tendencia a nivel mundial.

En nuestro país la transformación de lo analógico a lo digital deberá ir de a poco, completando un ciclo que va de los elementos de producción de la imagen (cámaras, gráficas, sonido, edición) a su posterior transmisión (antenas satelitales, cableado de fibra óptica), hasta que finalmente llegue al televisor familiar.

Con la llegada de la televisión digital se hacen posibles una serie de facilidades como: los servicios interactivos, tele banca, servicios de Internet, etc.

La Televisión Estereoscópica al ser una de las tantas aplicaciones de la televisión digital es de vital importancia, debido a que siempre ha existido un gran interés por parte del ser humano en ver las imágenes de una manera mas real y vívida, ya que la estereoscopia es una técnica que comenzó a desarrollarse hace mas de 150 años, que merced a las posibilidades informáticas y tecnológicas en cuanto a

tratamiento de la imagen y transmisión de la misma puede desarrollarse como una herramienta de visualización muy poderosa, no solo en televisión comercial sino también en otras aplicaciones como realidad virtual, medicina, ingeniería molecular, topografía y estudio de terreno, investigación espacial, video bajo demanda, telecompras, etc.

Aunque la estereoscopia precedió a la fotografía, no se había podido transmitir antes una imagen estereoscópica para televisión, debido a las limitaciones del ancho de banda que presentaba la televisión analógica, ahora con todas las facilidades de la televisión digital y gracias a los nuevos estándares de compresión digital, es posible la transmisión de imágenes estereoscópicas (tridimensionales).

CONTENIDO

CAPÍTULO I

1. FUNDAMENTOS TEÓRICOS

1.1 LA TELEVISIÓN DIGITAL

La digitalización de la señal analógica es una tendencia debido a las ventajas que presenta una señal digital en su manejo y procesamiento con respecto a la señal analógica, principalmente en aspectos tales como: calidad de transmisión independiente de la distancia debido a la regeneración, transmisión de la información independiente de su naturaleza, facilidad de adaptación a nuevas tecnologías y medios de transmisión, entre otras.

Se entiende por digitalización de la señal a la transformación de una señal analógica a un código binario (unos y ceros) denominados bits, cuya agrupación de 8 bits forman 1 byte, pudiéndose distribuir por lo tanto 2^8 (256) valores parciales.

El cambio de técnica de transmisión y almacenamiento de analógico a digital en televisión se debe al mejor manejo de la señal, ya que la posibilidad de transformar tanto señales acústicas como visuales, hace que se anulen las diferencias entre audio y video. Además la digitalización universal de datos promete fusionar los instrumentos que se utilizan en telecomunicaciones, informática y televisión; llegando en el futuro a los hogares por vías de transmisión totalmente digitales.

Hasta ahora la televisión convencional (analógica) se rige básicamente en tres sistemas de televisión a color, conocidos como: PAL (**P**hase **A**lternating **L**ine) que se utiliza en España, Italia, Inglaterra y Alemania; SECAM (**S**équentiel **C**ouleur **A**. **M**émoire) que se utiliza en Francia , Rusia y algunos países de Europa Oriental; y el sistema NTSC (**N**ational **T**elevision **S**istems **C**ommittee) utilizada en Canadá, Estados Unidos, México, Japón Y algunos países de América del Sur.

Durante algún tiempo se pensaba que si la televisión analógica era sustituida por la digital, la televisión digital fracasaría debido a que no sería rentable por el hecho de necesitar un gran ancho de banda, pero con los avances en técnicas de compresión y manejo de señal se ha reducido mucho el requerimiento de ancho de banda, dando como resultado que la televisión digital sea un hecho en la actualidad.

1.1.1 DIGITALIZACIÓN DE LA SEÑAL DE TELEVISIÓN

Como se conoce para digitalizar una señal analógica se deben seguir tres pasos que son: Muestreo, Cuantización y Codificación.

Muestreo .- es el procedimiento mediante el cual se toman muestras de la amplitud de una señal analógica a determinados intervalos de tiempo. La frecuencia de toma de muestras deberá cumplir el criterio de Nyquist, el mismo que dice: "la frecuencia de muestreo debe ser por lo menos del doble de la máxima frecuencia contenida en la señal, $f_m > 2\ f_{señal}$".

El muestreo de una señal de video se lo define para las tres señales que componen una señal de video, siendo estas, la señal de blanco y negro conocida como luminancia **(Y)** y el color denominado crominancia **(C)**, estando este último formado por la crominancia de color diferencia rojo **(Cr)** y crominancia de color diferencia azul **(Cb)**. Una muestra simple (con las tres componentes) es llamado elemento de cuadro, píxel o pel.

La frecuencia de muestreo de la señal de luminancia es de 13,5Mhz y la de crominancia es de 6,75 Mhz que corresponde a la mitad del valor de la anterior. Se debe decir además que existen varias clases o formatos de muestreo que se especifican con la notación : **Y:Cr:Cb** . Donde cada letra establece la proporción de muestras de la señal a la que corresponde, es decir si encontramos la notación 4:2:2, indica que por cada 4 muestras de la *componente de luminancia* (**Y**)se

toma 2 muestras de crominancia de color diferencia rojo (**Cr**) y 2 muestras de color diferencia azul (**Cb**).

Los formatos de muestreo mas utilizados son **4:4:4** (recomendado para gráficos complejos y post - producción), **4:2:2** (recomendado para operaciones regulares de los estudios de televisión), **4:1:1**(recomendado para circuitos cerrados de televisión) y **4:2:0** (que se recomienda para transmitir señales de televisión).

Cabe anotar que el formato de muestreo **4:2:0** no significa que no exista muestra de color diferencia azul **(Cb)**, sino que por cada 4 muestras de la componente **Y** se toman 2 muestras de **Cr** y 0 de **Cb**, luego en la próxima línea por cada 4 muestras de **Y** se toman 0 de **Cr** y 2 de **Cb,** con lo cual cada componente de crominancia es muestreada a un cuarto del de luminancia.

Cuantización: al muestrear la señal se obtienen diferentes niveles de voltaje, a los cuales mediante este proceso deberá asignárseles un determinado número de bits para cada muestra, es decir cuantificar la muestra. El número de bits utilizado establece el número de niveles de cuantización determinando así la exactitud con que una muestra puede ser representada. Para señales de televisión, en video se utilizan 8 o 10 bits (es decir 256 0 1.024 niveles) , mientras que para audio se utilizan 16 o 20 bits (osea 65.536 o 1'048.576 niveles) , esto se debe a que la sensibilidad del oído exige por lo menos una resolución de 16 bits. El conjunto de muestreo y codificación es conocido como conversión A/D.

Codificación: El proceso de codificación tiene por objetivo el minimizar el número de bits que se necesitan para representar la información de video y audio de una transmisión.

Los procesos de codificación se basan en la eliminación, en mayor o menor grado, de la información redundante o ajena a la imagen (o sonido) de la señal a transmitirse. Esa supresión ocasiona inevitablemente efectos secundarios de codificación, debiéndose determinar un algoritmo de codificación tal que los efectos secundarios permanezcan virtualmente imperceptibles bajo condiciones

de observación definidas.

Aunque con algunas imágenes muy críticas, que tienen un alto contenido de partes en movimiento, quizá no se consiga ese objetivo y aparecerán efectos secundarios visibles en la imagen decodificada, tales como una menor resolución de los detalles finos, información diagonal y, especialmente, representación del movimiento dinámico.

En las transmisiones de alta calidad es preciso contar con una velocidad binaria lo suficientemente grande para poder conseguir en la práctica una imagen no degradada en condiciones de recepción nominales, para un alto porcentaje del contenido de imagen previsto en las aplicaciones de radiodifusión

La mayoría de los expertos concuerdan actualmente que para transmitir una señal de televisión de alta definición, cuya calidad sea virtual y subjetivamente transparente con respecto a la señal de estudio, bastaría con unos 110-120 Mbps para la codificación de la señal de imagen. La gran mayoría de imágenes (incluidas las representaciones de movimiento muy críticas) no tendrían *efectos* secundarios de codificación perceptibles.

1.1.2 SISTEMAS PARA TELEVISIÓN DIGITAL

En la actualidad los sistemas utilizados para televisión digital son: DVB (**D**igital **V**ideo **B**roadcasting **S**ystem) que fue adoptado por los países de la Comunidad Económica Europea, Australia, Nueva Zelandia, Singapur e India; el sistema ATSC (**A**dvanced **T**elevision **S**ystems **C**omittee) adoptados en Canadá, Estados Unidos , Argentina en América del Sur, Corea del Sur, Taiwán y China Oriental; Sistema ISDB (Integral **S**ervice **D**igital **B**roadcasting) que rige en Japón y promete ser un duro contendiente para los dos sistemas anteriores.

1.1.2.1 Sistemas DVB

El proyecto DVB (**D**igital **V**ideo **B**roadcasting) comprende a 170 organizaciones de 21 países, interesadas en estandarizar a nivel mundial los mecanismos de difusión de televisión y servicios asociados. Los participantes son departamentos gubernamentales, reguladores, operadores, difusores y fabricantes. Es el estándar utilizado en Europa y como tal adoptado oficialmente por el Instituto Europeo para Normalización de las Telecomunicaciones (ETSI).

En la tecnología DVB se utiliza el sistema MPEG-2 como método de compresión de audio y video; además proporciona técnicas de modulación y métodos de codificación para corrección de errores en sistemas por satélite, por cable y terrestres; también DVB proporciona formatos de inserción de datos al canal de transmisión y receptores de 6, 7 y 8 MHz. La figura 1.1 muestra en diagrama de bloques como se halla estructurado el sistema DVB.

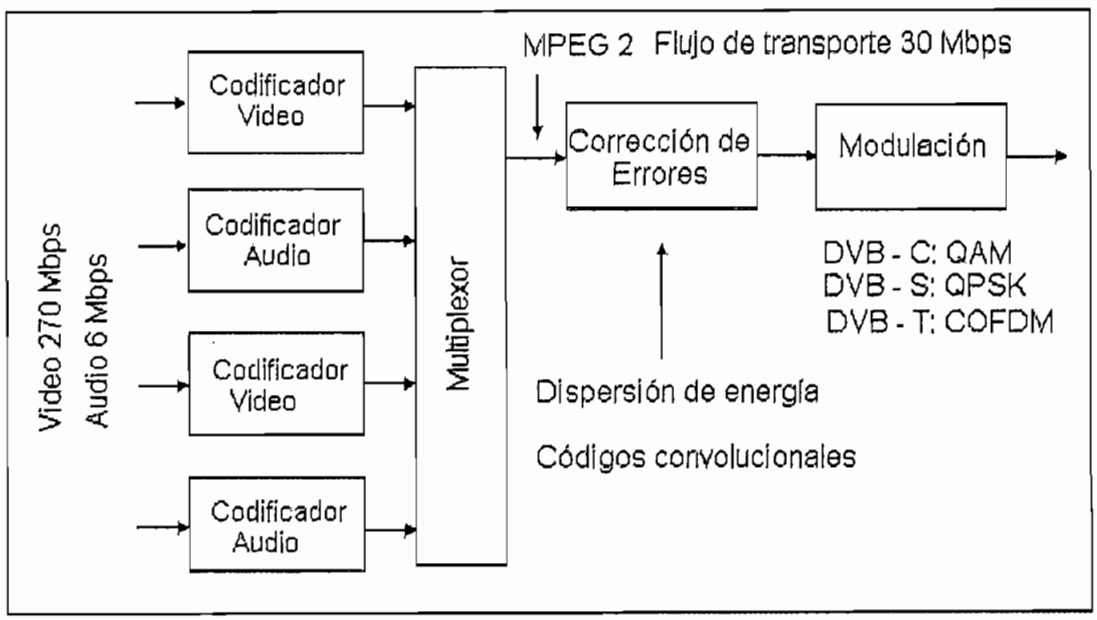


Figura 1.1 Diagrama de bloques del sistema DVB.

Por la existencia de varios medios de transmisión, el DVB bajo la supervisión del ETSI ha desarrollado varios estándares de video digital, tales como:

1.1.2.1.1 DVB-S (Difusión de video digital por satélite)

El sistema DVB-S (**D**igital **V**ideo **B**roadcasting by **S**atellite) permite un incremento de la capacidad de transmisión de televisión digital vía satélite utilizando técnicas de compresión basadas en el estándar MPEG-2. Para esta transmisión se adopta la codificación QPSK (**Q**uadrature **P**hase **S**hift **K**eying) con velocidad de transmisión variable de 18.4 a 48.4 Mbps.

Los sistemas de transmisión pueden llevar combinaciones flexibles de audio y video MPEG-2 y otros datos, constituyendo canales que son a continuación multiplexados. Usa un estándar de enmascaramiento (scrambling) disponible (Common Scrambling Algorythm) que controla el acceso a esta información, evitando problemas de piratería. En la figura 1.2 se muestra un típico sistema de difusión de video digital por satélite.

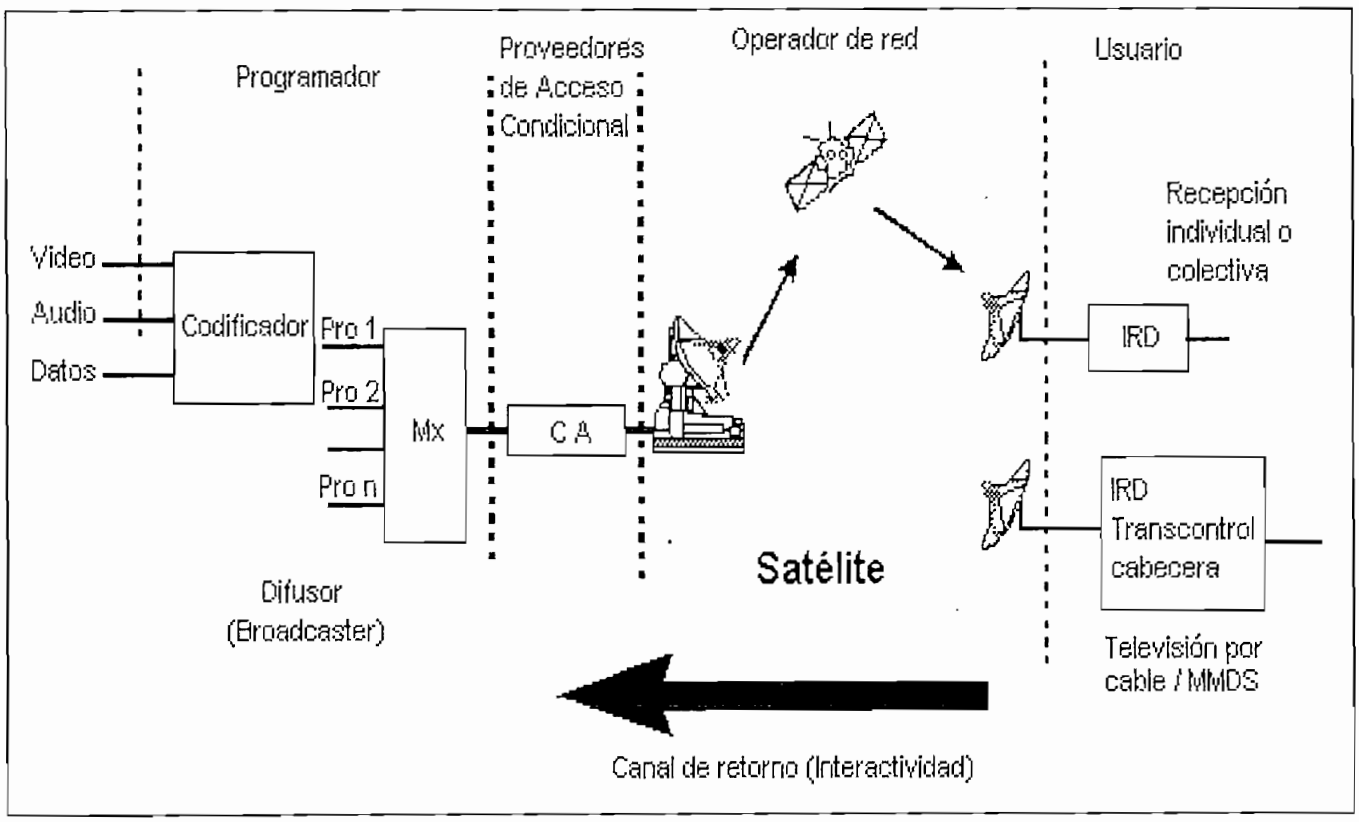


Figura 1.2 Sistema de difusión de video digital por satélite (DVB-S)

1.1.2.1.2 DVB-T (Difusión de video digital terrestre)

Se basa en la utilización de la tecnología de modulación COFDM (**C**oded **O**rthogonal **F**requency **D**ivision **M**ultiplexing) que divide la información a transmitirse entre un cierto número de portadoras (modo "2k" con 1705 portadoras y modo "8k" con 6817) cada una modulada individualmente con una tasa binaria baja. Se protege la información a transmitirse mediante códigos FEC (**F**orward **E**rror **C**orrection), además se introduce un intervalo de guarda que se inserta entre símbolos consecutivos para evitar la interferencia intersimbólica y proteger a la señal frente a los ecos (propagación multitrayecto). Se escogió esta modalidad de modulación debido a que los estudios llevados a cabo demostraron que este diseño rinde buenos resultados en zonas con gran densidad de obstáculos, donde pueden producirse reflexiones de ondas con trayectorias múltiples de propagación. Combinando los parámetros antes mencionados se obtienen 60 modos de operación, con capacidades binarias entre 5 y 32 Mbps.

Las especiales características de este estándar permiten ofrecer un elevado grado de inmunidad frente a ecos o propagación multitrayecto, de hecho si el eco cae dentro del intervalo de guarda incluso puede beneficiar a la señal.

Además permiten la introducción de redes de frecuencia única (SFN, **S**ingle **F**requency **N**etworks) donde todos los transmisores están sincronizados en término de bit, frecuencia y tiempo, es decir todos emiten lo mismo a la vez y en la misma frecuencia. En la figura 1.3 se muestra un sistema de difusión de video digital terrestre.

Las ventajas en términos de eficiencia espectral son evidentes. En donde antes se emitía un único programa analógico utilizando para ello 9 frecuencias, ahora se podrán emitir 9 tramas, una por canal, conteniendo cada trama un número de programas según el modo DVB-T seleccionado. A modo de ejemplo 4 programas de televisión se codificarán por trama, resulta que se tendría 36 programas utilizando el mismo espectro.

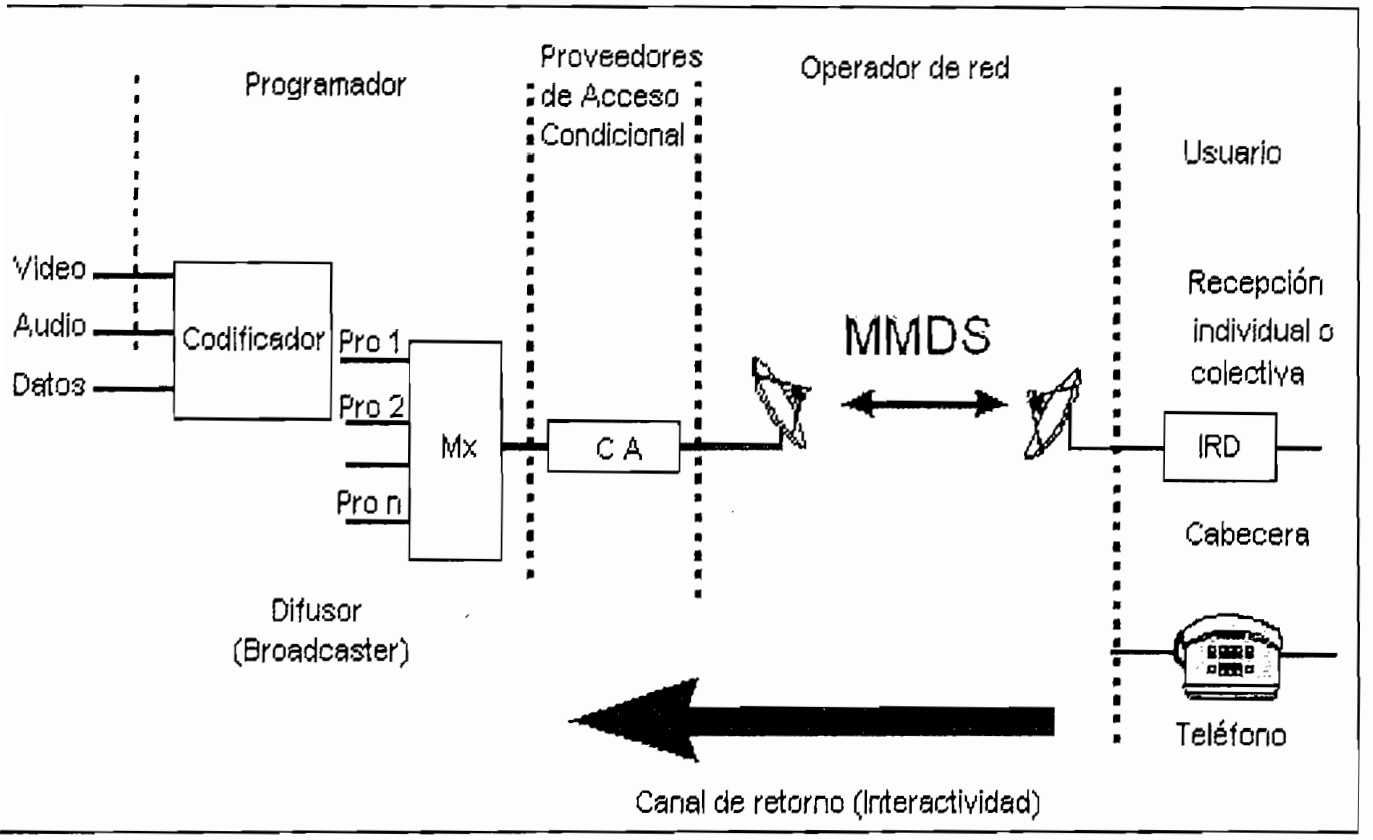


Figura 1.3 Sistema de difusión de video digital terrestre (DVB-T)

1.1.2.1.3 DVB-C (Difusión de video digital por cable)

El sistema de red de cable tiene funcionamientos semejantes al DVB-S, la única diferencia radica en que el DVB-C se basa en la modulación QAM(Modulación de amplitud en cuadratura) en lugar de la técnica de modulación QPSK que utiliza el sistema satelital.

El sistema se centra en 64-QAM, pero los sistemas de niveles más bajos como 16-QAM y 32-QAM también pueden usarse, teniendo en cuenta la capacidad del sistema contra la robustez de los datos. Los sistemas de niveles altos, como 128-QAM y 256-QAM también son posibles de utilizarse, dependiendo de la capacidad del cable para cubrir el margen de codificación. La figura 1.4 presenta un esquema de sistema de difusión de video digital por cable.

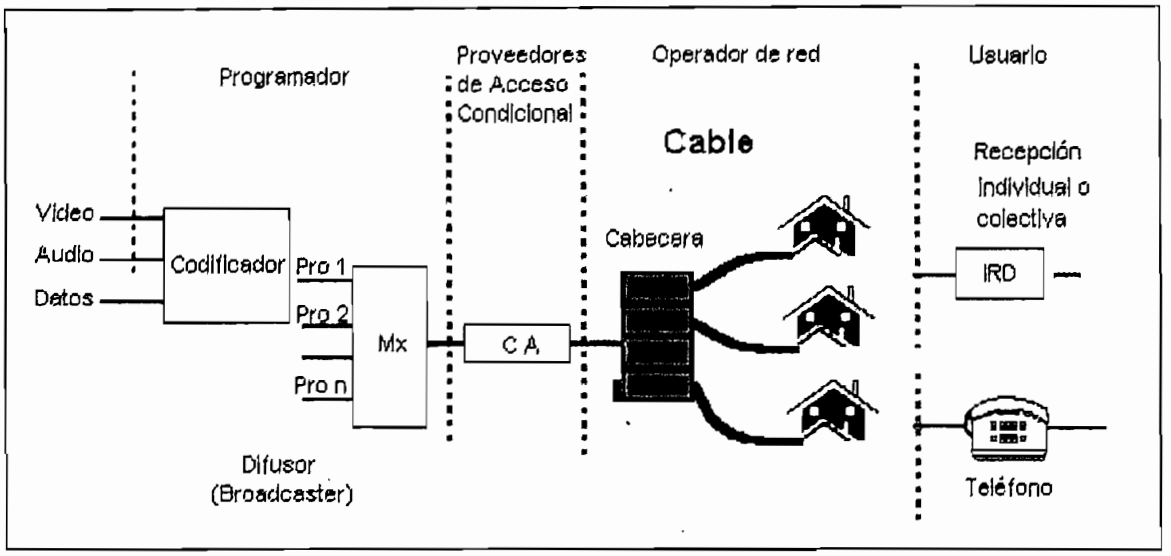


Figura 1.4 Sistema de difusión de video digital por cable (DVB-C)

1.1.2.1.4 DVB-MC/S (Difusión de video digital multipunto por microondas)

El sistema **DVB-MC** utiliza frecuencias de microonda por debajo de 10 GHz, se aplica básicamente a la cobertura directa sobre las casas de los usuarios. Es basado fundamentalmente en el DVB-C, y permite que un receptor común sea usado para transmisiones por cable y transmisiones por microonda.

La norma **DVB-MS** usa frecuencias aproximadamente sobre los 10 GHz para la distribución directa en hogares de los espectadores. Es basado en DVB-S, y puede ser recibido por un receptor DVB-S equipado con un conversor de frecuencia.

1.1.2.2 Sistema ATSC.

ATSC fue formado por la unión del Comité en Coordinación Inter-social (JCIC) para establecer normas técnicas voluntarias para los sistemas de televisión avanzados, incluyendo televisión digital de alta definición (HDTV). Este estándar de televisión digital describe un sistema diseñado para transmitir video y audio de alta calidad, además de datos por encima de los 6 Mhz por un solo canal. El sistema puede entregar con facilidad sobre los 19 Mbps de throughput en un

canal terrestre de difusión y sobre los 38 Mbps en un canal de cable de televisión. Esto significa que la resolución de la codificación fuente de video puede ser hasta 5 veces mas alta que la televisión convencional (NTSC).

El sistema ATSC tiene como objetivo aumentar al máximo la información que pasa por el medio de transmisión, minimizando la cantidad de bits exigida para representar la imagen de video y audio asociados, esto se consigue representando video, audio y fuentes de datos con tan pocos bits (tren de bits) como sea posible, conservando el nivel de calidad requerido.

Aunque los subsistemas de transmisión RF descritos en esta norma son diseñados específicamente para aplicaciones terrestres y por cable, el objetivo es que el video, audio y servicio de transporte multiplexado pueden ser usados en otras aplicaciones.

De acuerdo a la norma de televisión digital A/53[1], el sistema ATSC se subdivide en tres subsistemas, como se muestra en la figura 1.5, siendo estos:

1. codificación y compresión de fuente
2. Servicio de multiplexación y transporte
3. Transmisión RF

1.1.2.2.1 Codificación y compresión de fuente

Este subsistema se refiere a métodos de reducción de velocidad de transmisión, conocido también como compresión de datos, aplicados a video audio y trenes de bits que incluyen control de datos, control de datos de acceso condicional, datos asociados con servicios de programas de audio, video, y servicios de programas independientes.

[1] ATSC, "A/53: Digital Televisión Estándar" ,pag 17-19.

El propósito de la codificación es minimizar el número de bits necesarios para representar la información de audio y video. El sistema de televisión digital emplea MPEG-2 para compresión de video y el estándar de compresión de audio digital AC-3, para la compresión de audio.

1.1.2.2.2 Transporte y multiplexación de servicios

Trata sobre la división de trenes de datos digitales en paquetes de información, los tipos de paquetes, y los métodos mas adecuados para la multiplexación de paquetes de audio y video.

El sistema de televisión digital emplea MPEG-2 para el transporte de paquetes, y la multiplexación de video, audio y señales de datos para sistemas de difusión digital. Transportar la información en paquetes permite que los bits sean separados en tamaño fijo, y así poder aplicar métodos de corrección de errores, multiplexación y conmutación de trenes de bits, sincronización de tiempo, etc., así como permitir la compatibilidad con mecanismos de transporte que usan el Modo de Transferencia Asincrónico (ATM).

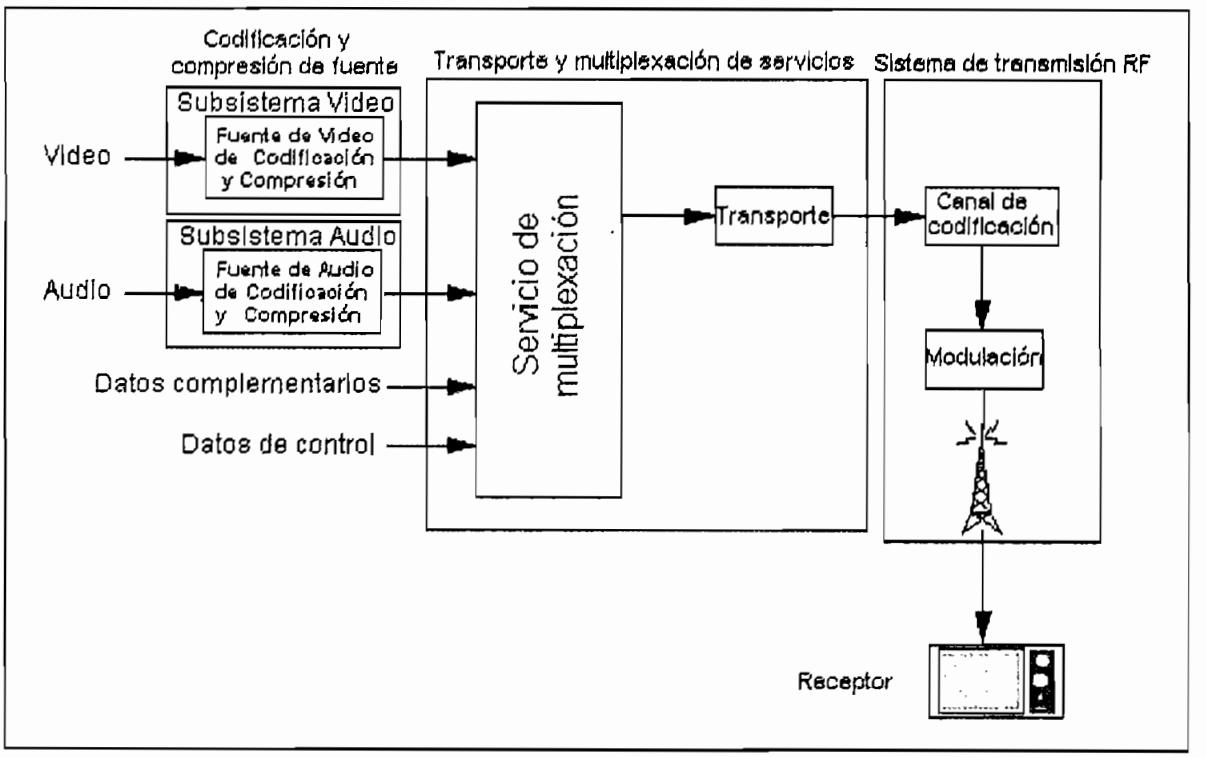


Figura 1.5 Modelo de Difusión de TV digital terrestre para el sistema ATSC.

1.1.2.2.3. Transmisión RF:

Este subsistema tiene que ver con la codificación y modulación. El codificador toma el tren de bits de datos y agrega información adicional que usa el receptor para reconstruir la señal recibida.

El sistema de modulación usa la técnica de banda lateral vestigial (VSB), que ofrece dos modos de operación: un modo 8 VSB para difusión terrestre y el modo 16 VSB para una velocidad alta de datos.

1.2 LA ESTEREOSCOPÍA.

1.2.1 PRINCIPIOS DE LA ESTEREOSCOPÍA.

La palabra estereoscopía viene del griego *estereós* y *skopeín* que significa "observación sólida", lo cual da a entender que se trata de una observación tridimensional con cierto nivel de profundidad.

Nuestro mecanismo natural de visión es estéreo, es decir, somos capaces de apreciar, a través de nuestros dos ojos, las diferentes distancias y volúmenes en el entorno que nos rodea. Debido a la separación existente entre los ojos, obtenemos dos imágenes con pequeñas diferencias entre ellas, esto se denomina *disparidad*. El cerebro procesa las diferencias entre ambas imágenes y las interpreta de forma que percibimos la sensación de profundidad, lejanía o cercanía de los objetos que nos rodean. Este proceso se denomina ***estereopsis***.

La vista humana es capaz de determinar distancias de hasta unos cien metros gracias a la visión ligeramente distinta que percibe cada uno de los ojos de la escena observada (paralaje) un ejemplo se muestra en la figura 1.6. La distancia mas común entre las pupilas es de 65 mm, pudiendo variar desde los 45 a los 75 mm.

En la estereopsis intervienen diversos mecanismos. Cuando observamos objetos muy lejanos, los ejes ópticos de nuestros ojos son paralelos. Cuando observamos un objeto cercano, nuestros ojos giran para que los ejes ópticos estén alineados sobre él, es decir, *convergen.* A su vez se produce la acomodación o enfoque para ver nítidamente el objeto. Este proceso conjunto se llama ***fusión***. No todo el mundo tiene la misma capacidad de fusionar un par de imágenes en una sola tridimensional. Alrededor de un 5% de la población tiene problemas de fusión.

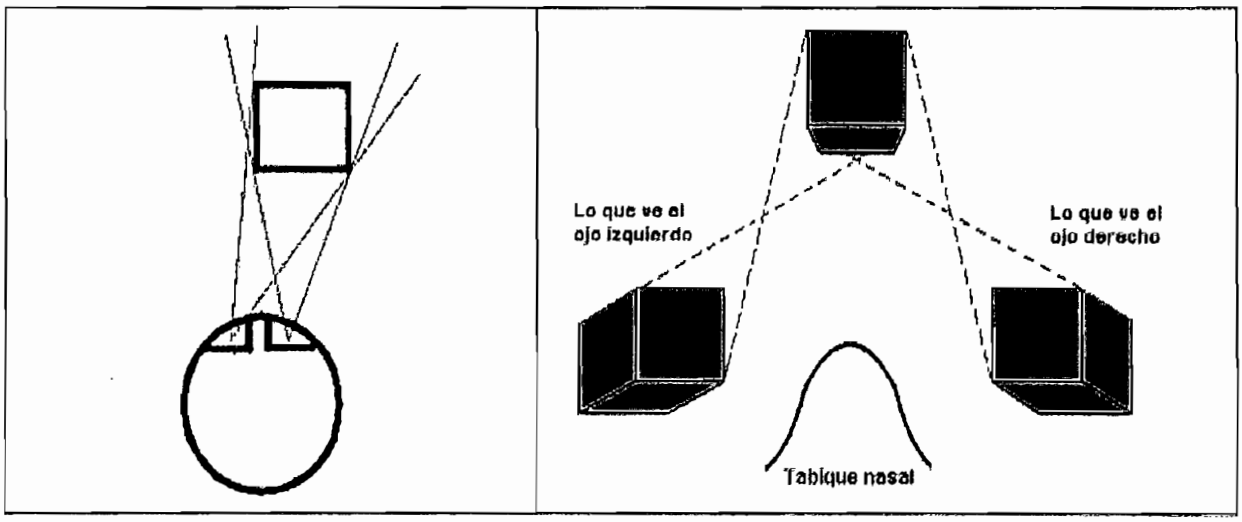


Figura 1.6. Estereopsis visual.

La ***agudeza estereoscópica*** es la capacidad de discernir, mediante la estereopsis, detalles situados en planos diferentes y a una distancia mínima. Hay una distancia límite a partir de la cual no somos capaces de apreciar la separación de planos, y que varía de una persona a otra. Así, la distancia límite a la que dejamos de percibir la sensación estereoscópica puede variar desde unos 60 metros hasta cientos de metros.

Un factor que interviene directamente en esta capacidad es la separación Ínterocular. A mayor separación entre los ojos, mayor es la distancia a la que apreciamos el efecto de relieve. Esto se aplica por ejemplo en los prismáticos, donde mediante prismas se consigue una separación ínter ocular efectiva mayor que la normal, con lo que se logra apreciar en relieve objetos distantes que en condiciones normales no seríamos capaces de separar del entorno.

También se aplica en la fotografía aérea, en la que se obtienen pares estereoscópicos con separaciones de cientos de metros y en los que es posible apreciar claramente el relieve del terreno, lo que con la visión normal y desde gran altura sería imposible. El efecto obtenido con una separación interocular mayor que la habitual es el de que los objetos parecen más pequeños de lo normal (liliputismo), y la técnica se denomina ***hiperestereoscopía***.

El efecto contrario se consigue con la ***hipoestereoscopía***, es decir, con la reducción de la distancia interocular, imprescindible para obtener imágenes estereoscópicas de pequeños objetos (macrofotografías), o incluso obtenidas por medio de microscopios.

1.2.1.1 Sistemas de visión

La percepción de profundidad de la visión humana es un proceso complejo y sofisticado que responde a más de diez factores, los cuales, unidos a la imagen en dos dimensiones que se proyecta sobre la retina del ojo, permiten ver el espacio en tres dimensiones. Estas percepciones de profundidad se pueden clasificar en **monoculares** y en **binoculares**.

1.2.1.1.1 Percepción monocular

Es aquella igual para ambos ojos y tiene la misma efectividad si son vistas sólo por uno de ellos. Entre éstas se puede citar :

- ***perspectiva lineal,*** es la reducción progresiva del tamaño de la imagen a medida que la distancia al objeto aumenta.
- ***tamaño de la imagen,*** a medida que la imagen de un objeto es mayor, éste parece más cercano.
- ***perspectiva de superficie,*** se refiere a lo borroso que los objetos se perciben con la distancia.

- ***matices y sombras,*** que es la impresión de convexidad dada por el efecto de que la mayor parte de la iluminación proviene de arriba.

- ***gradiente de textura,*** esto es un tipo de perspectiva lineal que corresponde al grado de rugosidad de un objeto uniforme a medida que se va alejando.

- ***paralaje de movimiento,*** es el resultado del cambio de posición de un objeto en el espacio, sea por causa del movimiento mismo del objeto, o bien a causa del desplazamiento de la cabeza del observador.

- ***acomodación,*** es el ajuste de la distancia focal del cristalino. el cristalino del ojo puede hacerse más plano o más convexo según la necesidad de enfocar el objeto observado. Se hace plano cuando se enfocan objetos distantes y se hace convexo para aquellos más próximos. El cerebro procesa estos cambios determinando referencias aproximadas de distancias.

1.2.1.1.2 Percepción binocular

Es la percepción que resultan de ver con los ojos desde puntos de vista ligeramente diferentes y son fundamentalmente dos:

- ***disparidad binocular,*** que es la diferencia entre las imágenes de un mismo objeto proyectadas sobre la retina de nuestros ojos. Nuestro cerebro procesa las diferencias entre ambas imágenes y las interpreta de forma que percibimos la sensación de profundidad, lejanía o cercanía de los objetos que nos rodean. Este proceso se denomina estereopsis y es el más importante indicador de profundidad.

- ***convergencia,*** es la capacidad de hacer converger el eje óptico de los dos ojos sobre un mismo objeto. Así, cuando deseamos ver los objetos con claridad los ejes ópticos de nuestros ojos giran automáticamente de tal forma que la imagen del objeto quede situado sobre las áreas más sensibles de la retina y el cerebro determina las posiciones relativas.

1.2. 2 HISTORIA DE LA ESTEREOSCOPÍA.

Se consideran como pioneros en este tema a los famosos Euclides y el genial Leonardo da Vinci, quienes ya estudiaron el fenómeno de la visión binocular, también el famoso astrónomo Kepler llevo a cabo estudios relacionados con la estereoscopia. Siendo el físico escocés, Sir Charles Wheatstone, quien en junio de 1838 describió con cierto rigor el fenómeno de la visión tridimensional, construyendo luego un aparato con el que se podía apreciar en relieve dibujos geométricos, llamado Estereoscopio.

Años más tarde, en 1849, Sir David Brewster diseñó y construyó la primera cámara fotográfica estereoscópica, con la que obtuvo las primeras fotografías en relieve, construyendo posteriormente un visor con lentes para observarlas.

En 1862 Wendell Holmes, construyo un modelo de estereoscopio de mano, que se hizo muy popular a finales del siglo XIX, con el que se podían ver fotografías estereoscópicas montadas sobre cartón.

Durante los años 30, existe un resurgir de la estéreo fotografía a raíz de la aparición de la cámara 3D, con película de 35 mm. Como la Realist o la famosa ViewMaster que facilitaban al aficionado este tipo de imágenes.

En los años 50 se intentó la explotación comercial de las películas 3D, pero con escasa incidencia en el mercado cinematográfico. Además, algunas de las películas que se realizaron presentaban problemas de visión, por falta de conocimiento de toda la problemática que conlleva una película estereoscópica, lo que ocasionaba molestias visuales que hicieron que una parte del público rechazara este tipo de cine.

Experimentos con video anáglifo (gafas de colores) fueron numerosos y se difundieron ya en el año de 1953. La difusión del sistema anaglifo se continuó haciendo esporádicamente, dando lugar a la aparición ocasional de casetes anáglifos y videodiscos, pero esta técnica al emplear el método Pulfrich o gafas

prismáticas imposibilita una alta calidad y una visión confortable con video, siendo mejor con displays de computadora.

No seria sino hasta los años 80 cuando se consiguen los resultados mas espectaculares con los sistemas de gran formato de película para obtener imágenes de alta resolución en pantallas gigantescas, tras grandes inversiones en investigación y medios.

Para los años 90, los avances de la informática permiten presentar imágenes 3D en monitores de ordenador y utilizarlas para presentaciones en diseño asistido por computador.

1.3 ANTECEDENTES DE LA TELEVISIÓN ESTEREOSCÓPICA

La televisión estereoscópica no es una técnica del futuro, ya hoy en día se usa en países donde la televisión digital esta muy difundida, tales como: España, Japón, Estados Unidos y en Sudamérica en el vecino país de Argentina, en donde se ha experimentado ya con este tema.

Los sistemas estereoscópicos en general nacen de la necesidad del ser humano en visualizar las imágenes con profundidad y poder tener una idea real de distancia de los objetos, siendo el cerebro el que funde las dos imágenes que percibe cada ojo.

La televisión estereoscópica realmente fue imaginada por los pioneros de la TELEVISION a principios de los años 1920, despertando gran interés desde los primeros experimentos realizados con este nuevo medio, de tal forma que pioneros electrónicos tales como Hammond, Logie Baird, Lee DeForest, Zworykin y otros describen en sus patentes dispositivos de 3DTV.

Así por ejemplo en Agosto 10 de 1928 John L. Baird, en su laboratorio expuso ante otros científicos y representantes de la prensa, su sistema de televisión estereoscópica, el cual consistía en un aparato de transmisión que contiene un

disco perforado como se muestra en la figura 1.7A con dos espirales, el primer espiral comienza con un arreglo de orificios alrededor de una mitad de la circunferencia del disco, la segunda espiral ocupa la otra semicircunferencia con un arreglo similar, separadas una de la otra alrededor de 65 mm., que es la distancia aproximada que existe entre los ojos humanos. Detrás del disco, cuando este es montado en el transmisor como se muestra a la izquierda de la figura 1.7, está un arreglo de una fuente intensa de luz. En el frente del disco y alineado con la fuente luminosa, se colocan unos lentes en una relación tal que los puntos de transmisión luminosos ocasionen que el objeto se vea en una forma transversal.

El arreglo es duplicado de tal manera que cada espiral tiene su lente y fuente de luz; así, se obtienen dos puntos luminosos transversales del objeto alternadamente y dos imágenes son transmitidas, una para el ojo izquierdo y otra para el ojo derecho.

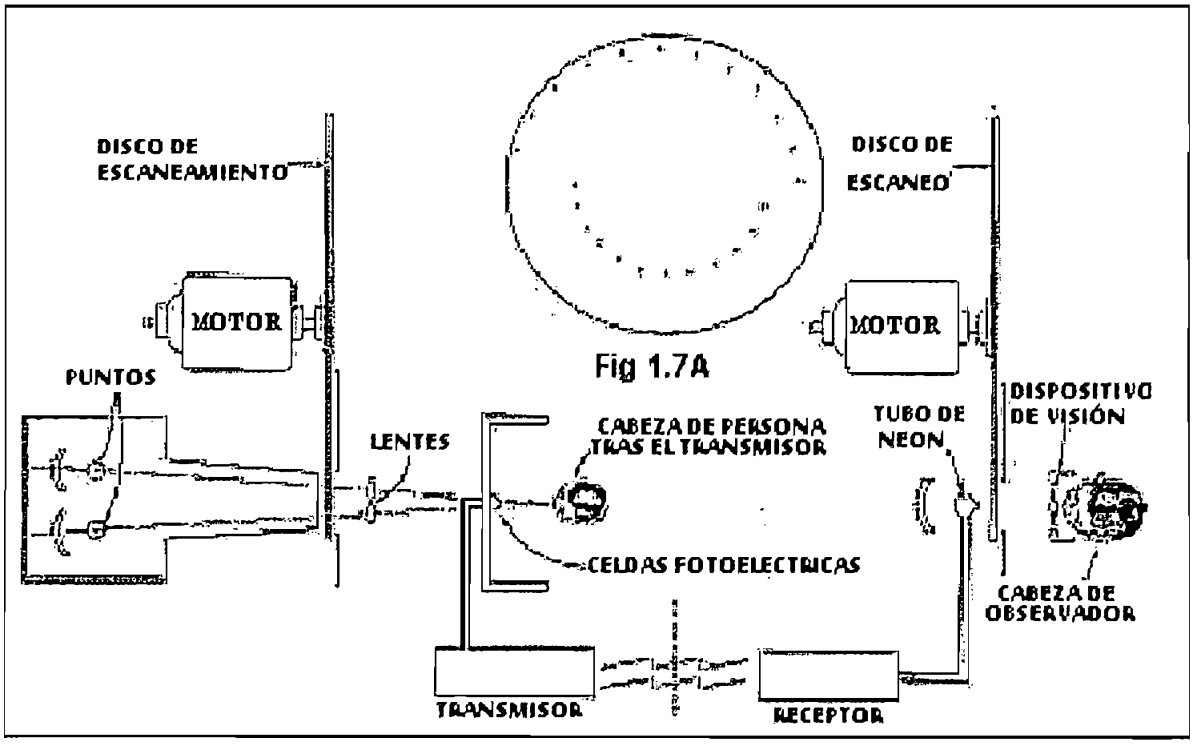


Figura 1.7 Sistema Baird de Televisión Estereoscópica

En la estación receptora se utiliza un dispositivo similar, como se muestra en la parte derecha de la figura 1.7. Un disco con el mismo arreglo de orificios corre

sincronizadamente con el disco trasmisor; pero detrás del disco receptor esta un tubo de neón arreglado como una televisión ordinaria.

El tubo de neón, sin embargo cubre ambas espirales y las ilumina alternadamente; de esta manera en la pantalla receptora aparecen dos imágenes separadas aproximadamente media pulgada. Una de estas corresponde al objeto como sería visto por el ojo derecho y la otra como lo vería el ojo izquierdo. Estas imágenes son entonces visualizadas a través de un visor estereoscópico, consistente de dos prismas, los cuales causan la convergencia y mezcla de las dos imágenes en una, similar a un visor estereoscópico para fotografías.

Es interesante notar que el dispositivo visor estereoscópico es realmente innecesario ya que se puede tener la capacidad de hacer que las imágenes se mezclen sin el uso de los prismas, con solo observar las imágenes fija y concentradamente , de tal forma que el ojo izquierdo sobreponga la imagen en la izquierda y el ojo derecho sobreponga en el otro. Este es en efecto el método usado por la mayoría de expertos en estereoscopia.

Aunque al parecer en la actualidad Logie Baird ha sido considerado el primero en construir dispositivos que funcionen. El primer dispositivo comercial fue el sistema dual de tubo de rayos catódico Dumont`s que apareció en los años 50.

1.4 MÉTODOS PARA VISUALIZACIÓN ESTEREOSCÓPICA

Para visualizar una imagen que presenta el efecto estereoscópico, se han ideado varios métodos, pero siempre teniendo en cuenta el principio de que cada ojo debe ver solamente la imagen que le corresponde. Al hacer una clasificación de los sistemas existentes, diremos que hay de dos tipos:

a.) Los que requieren de algún dispositivo especial, como son: *sistema anaglifo, sistema entrelazado, sistema polarizado, visores estereoscópicos.*

b.) Los que posibilitan ver una imagen prescindiendo de visores especiales, de tal forma que nuestra visión se adapte al estereopar, para poder captar la profundidad de la imagen, estos son: *visión cruzada, visión relajada, y displays auto estereoscópicos.*

1.4.1 SISTEMA ANAGLIFO.

Un anaglifo es el resultado de formar pares estereoscópicos (para dar imágenes tridimensionales) a partir de los positivos que se tiñen de diferente color, generalmente verde y rojo. Las dos imágenes se copian sobre el mismo papel ligeramente fuera de registro, o se montan por separado en un visor especial . En ambos casos deben observarse a través de filtros de colores complementarios a los de la imagen que debe observar cada ojo.

Si vemos a través de un filtro rojo, los colores verde o azul se ven como negro y si utilizamos un filtro verde, azul o cyan, el rojo parece negro, este es el principio utilizado para mezclar dos imágenes y al observarlas con filtros de color poder separar las dos imágenes. Un filtro de distinto color se pone en cada ojo, pudiendo combinarse los colores de la siguiente manera: , rojo-azul, rojo-verde, rojo-cyan. Gracias a esto cada ojo ve una imagen ligeramente distinta a la del otro, produciéndose la ilusión tridimensional cuando nuestro cerebro funde ambas imágenes.

Por convención, el filtro rojo se coloca del lado izquierdo. El color del otro filtro depende del medio que se va a utilizar. Para impresión se acostumbra a utilizar el azul. Para video o proyección el filtro es verde, que es mas brillante. Con estos filtros, la imagen parece estar en blanco y negro. Sin embargo uno tarda en acostumbrarse a los filtros.

La otra variante antes mencionada de filtro rojo-cyan, se utiliza si la imagen no esta muy saturada, por lo que se puede hacer una separación de color de la imagen, conservando el componente rojo de la imagen izquierda y los componentes verde y azul de la imagen derecha. De esta manera se puede

conservar el color de la imagen. Sin embargo la diferencia de luminosidad de las dos imágenes puede resultar muy cansada después de un tiempo.

Si la imagen es demasiado saturada en color, es posible que algunos elementos no se vean en una de las imágenes, por lo que es necesario bajar la saturación de color de la imagen. Este método tiene como ventaja el bajo costo de las gafas, y su desventaja radica en la perdida cromática. En la figura 1.8 se muestra un ejemplo de estas gafas.

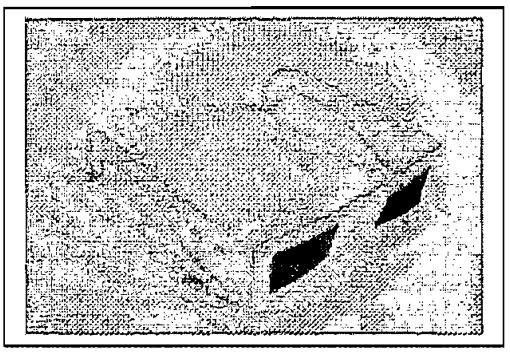


Figura 1.8 Gafas Anaglifas.

1.4.2 SISTEMA ENTRELAZADO

Usa el barrido de la pantalla como método de separación del estereopar. Con este sistema se presentan en secuencia y alternativamente las imágenes izquierda y derecha, sincronizadamente con unas gafas dotadas con obturadores de cristal líquido conocida como gafas shuttle (o también denominadas **LCS**, **L**iquid **C**rystal **S**hutter glasses o **LCD**, **L**iquid **C**rystal **D**isplay glasses) las cuales pueden tener cable o usar dispositivo infrarrojo; de forma que cada ojo ve solamente su imagen correspondiente.

A una frecuencia elevada, el parpadeo es imperceptible. Este sistema es utilizado en monitores de computador, TV y cines 3D de última generación. La figura 1.9 muestra un ejemplo de gafas LCD.

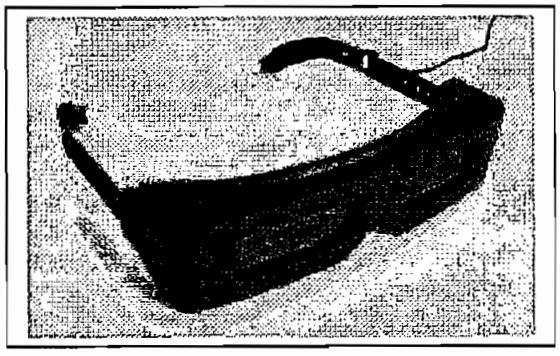


Figura 1.9 Gafas LCD.

La ventaja de este método radica en que no se pierde croma. La desventaja esta en que son mas costosas que las gafas anaglifas o polarizadas.

1.4.3. SISTEMA POLARIZADO

El sistema polarizado utiliza una técnica que funciona en base a un fenómeno de la física llamado polarización de la luz. Como sabemos la luz se transmite por ondas, pudiendo ser estas horizontales o verticales, la luz emitida por una lámpara es en todas direcciones , existiendo filtros que pueden eliminar las ondas en una dirección o polaridad, la luz resultante se denomina luz polarizada. Si se proyecta luz polarizada en una dirección y la vemos con un filtro a una inclinación de 90 grados respecto a la luz original, toda la luz será bloqueada.

Por lo tanto se puede proyectar dos imágenes, una polarizada en un sentido y la otra 90 grados y utilizar dos filtros para que cada ojo vea una imagen distinta. Los filtros son relativamente baratos y no presenta perdida cromática, teniendo como inconveniente que solo funcionan con sistemas de proyección, que generalmente requiere dos proyectores o un proyector especialmente modificado y una pantalla especial (reflejante) además de un entorno lo mas oscuro posible. A continuación en la figura 1.10 se muestra un par de gafas utilizadas en el método polarizado.

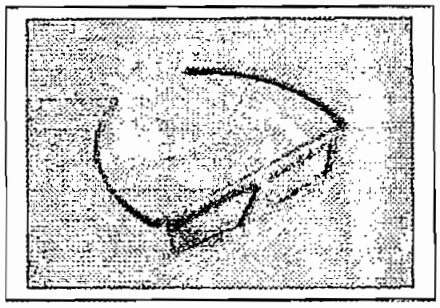


Figura 1.10 Gafas polarizadas.

Este método es ideal para audiencias grandes pudiéndose ver las representaciones a color y su principal inconveniente es que los filtros polarizados obscurecen la imagen por lo que se necesitan proyectores muy luminosos, existiendo un problema adicional con los proyectores actuales de video de cristal liquido, ya que estos polarizan la luz para funcionar, por lo que al colocar los filtros polarizadores la perdida de luminosidad es aun mayor. La pantalla además no debe romper la polarización de la luz, y este tipo de pantalla es difícil de conseguir en tamaños grandes.

1.4.4 VISORES ESTEREOSCÓPICOS

Estos visores se fundamentan en el principio de *Charles Wheatstone,* proyectando en forma paralela estereopares, (imagen izquierda–ojo izquierdo, imagen derecha-ojo derecho), ejemplos de esto son el famoso estereoscopio *Wheatston*, o el estereoscopio *Brewster,* así como el tan conocido View Master (Juguete que usa discos de cartón con fotografías estereoscopias). Algunos ejemplos de esto se muestra en la figura 1.11.

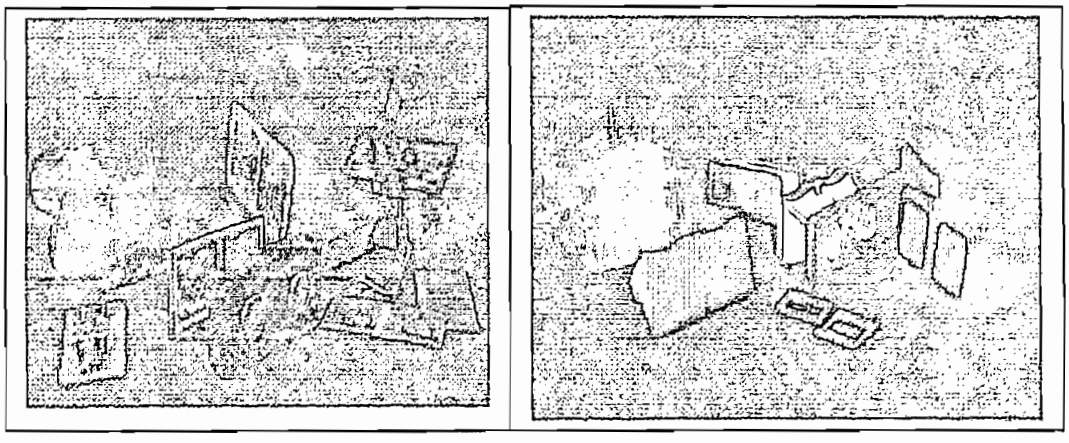


Figura 1.11 Visores estereoscópicos.

1.4.5 SISTEMA HMD (Head Mounted Display)

Un despliegue montado en la cabeza (HMD) es un casco estereoscópico y constituye un caso mas sofisticado de visor estereoscópico. Este sistema porta dos pantallas y los sistemas ópticos para cada ojo, de forma que la imagen se genera en el propio dispositivo. Su principal uso hasta ahora ha sido la Realidad Virtual, a un costo muy elevado y de forma experimental, aunque al bajar de precio aparecen otras aplicaciones, como los videojuegos. Los tipos más difundidos de HMD son: HMD con LCD, HMD proyectado y HMD con CRT (Tubos de Rayos Catódicos) pequeño.

El sistema HMD con LCD utiliza la tecnología de despliegue de cristal líquido (LCD) para mostrar las escenas, este sistema es mas claro con respecto a los otros tipos de HMDs, sin embargo la resolución y el contraste es bajo debido a que los cristales son polarizados para controlar el color de un píxel, lo cual le crea un retardo en la formación de la imagen por lo que se puede llegar a juzgar mal la posición de los objetos.

En el HMD proyectado el casco utiliza fibra óptica para transmitir la escena a la pantalla, este método es similar al CRT con la diferencia de que el fósforo es iluminado por la luz transmitida a través de la fibra óptica, donde cada fibra controla una celda con varios píxeles. El casco proyectado proporciona mejor resolución y contraste que el despliegue de LCD, esto significa que se puede ver una imagen con mucho mayor detalle. La desventaja de este dispositivo es que es caro y complicado de fabricar.

En el sistema HMD con CRT pequeño el casco utiliza dos tubos de rayos catódicos que se posicionan en el lado del casco, utilizando espejos para reflejar la escena hacia el ojo. A diferencia de el casco proyectado, el fósforo es iluminado por un rayo de electrones y no por cables de fibra óptica. El casco con CRT es muy similar al casco proyectado, sin embargo, este tipo de casco es más pesado que la mayoría de los otros tipos de casco debido a los componentes electrónicos que le son

agregados lo que provoca la generación de grandes cantidades de calor haciendo que quien lo utilice se sienta incómodo debido al peso y el calor. La figura 1.12 muestra un sistema HMD.

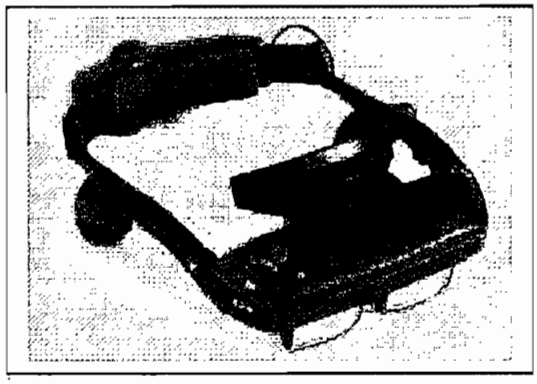


Figura 1.12 Visor HMD con LCD.

1.4.6 VISIÓN RELAJADA

Los ojos observan cada uno su imagen correspondiente, manteniendo sus ejes ópticos paralelos, es decir, como si mirásemos al infinito, la figura 1.13 trata de ilustrar esto.

Sólo puede usarse este método con imágenes no superiores a 65 milímetros entre sus centros. Es el método usado para ver las imágenes de los libros con estereogramas de puntos aleatorios ("ojo mágico"). La ventaja de este método reside en el hecho de prescindir de dispositivos especiales. Siendo su desventaja que para algunas personas resulta difícil la relajación visual.

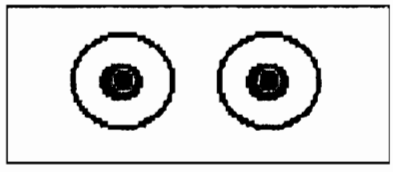


Figura 1.13 Visión Relajada

1.4.7 VISIÓN CRUZADA

Este método es similar a la visión relajada, pero consiste en que las imágenes se observan cruzando los ejes ópticos de los ojos. En la figura 1.14 se muestra una ilustración de aquello.

El par estéreo se presenta invertido, es decir, la imagen derecha está situada a la izquierda y viceversa. Para ayudarnos podemos mirar un lápiz situado entre nuestros ojos y las imágenes. Este método debe usarse con imágenes de dimensiones superiores a 65 milímetros entre sus centros, aunque la imagen virtual aparece más pequeña. Las características en cuanto a ventajas y desventajas son las mismas que para la visión relajada.

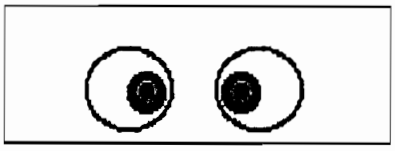


Figura 1.14 Visión Cruzada.

1.4.8 MONITORES AUTOESTÉREO

Se están desarrollando prototipos de monitores que no precisan gafas especiales para su visualización. Todos ellos emplean variantes del sistema lenticular, es decir, micro lentes dispuestas paralela y verticalmente sobre la pantalla del monitor, que generan una cierta desviación a partir de dos o más imágenes (normalmente de 2 a 5). La figura 1.15 trata de ilustrar este tipo de monitores.

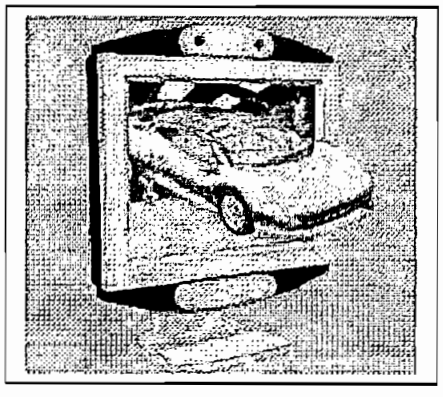


Figura 1.15 Monitor autoestereoscópico

1.5 MÉTODOS PARA SIMULAR EL EFECTO 3D.

Además de los métodos de visualización estereoscópica, se han ideado métodos que simulan el efecto estereoscópico, es decir sistemas que no son propiamente un sistema de visualización estéreo, ya que no se parte de un par de imágenes sino de una única imagen 2D animada. Estos sistemas son: El sistema Cromatek, el Sistema Dinámico conocido también como Sistema Pulfrich y el sistema VISIDEP.

1.5.1 SISTEMA CROMATEK.

Este sistema utiliza lo que se conoce como rejilla de difracción. La rejilla de difracción parece una mica común y corriente, pero funciona de manera semejante a un prisma de cristal, la luz que la atraviesa, se desvía de manera distinta según su color. Cuando uno usa una mica de difracción en un ojo, los objetos parecen tener una profundidad distinta según su color.

En un dibujo el azul se vera siempre en el fondo, el amarillo en medio y el rojo mas cerca. Las imágenes preparadas para este sistema pueden verse de manera normal y sólo con los lentes aparecen en 3D, incluso en imágenes que no fueron diseñadas para 3D, pero que se elaboraron con colores intensos. El inconveniente es que la selección de colores es limitada y no funciona bien con fotografías.

1.5.2 SISTEMA DINÁMICO.

Este sistema se basa en el llamado efecto *Pulfrich*, descubierto en 1922 por un médico alemán de nombre *Carl Pulfrich.* El efecto Pulfrich se fundamente en un dato fisiológico de nuestro cerebro, este dato indica que el cerebro tarda un poco en procesar las imágenes. Si las imágenes están obscuras el cerebro tarda un poco más.

Así para simular un efecto estereoscópico, se observa una imagen en movimiento horizontal sobre un plano y se pone un filtro en un solo ojo, logrando así que en la estereopsis el cerebro perciba la misma imagen pero con una pequeña diferencia

de posición horizontal, lo que genera el efecto estereoscópico ya que al ocupar mas tiempo el cerebro en procesar la imagen, esta parecerá estar en una posición o ángulo distinto con respecto al mismo objeto fijo observado directamente. En la figura 1.16 aparecen unas gafas utilizadas en este sistema.

El inconveniente de esta técnica es que se requiere que todo el tiempo exista movimiento, sin embargo la imagen puede verse de manera normal si no se utilizan los filtros.

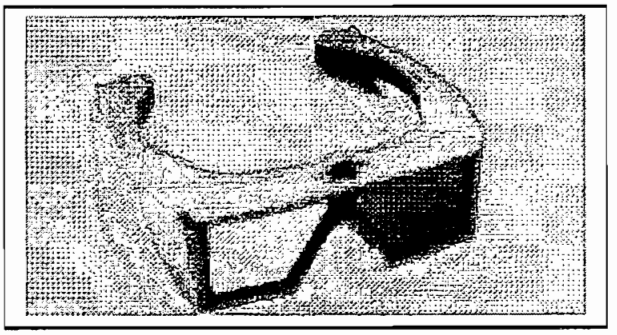


Figura 1.16 Gafas utilizadas en el Sistema Dinámico.

1.5.3 SISTEMA VISIDEP

Este sistema, desarrollado por universidades de Carolina del Sur en Estados Unidos, hace posible ver imágenes tridimensionales, sin necesidad de lentes especiales, cámaras, proyectores o algún tipo de efecto especial en el televisor.

Dando además la posibilidad de que todas las personas, incluso los de visión pobre y visión en un solo ojo, puedan apreciar las imágenes en tres dimensiones.

VISIDEP (***V**isual **I**mage **D**epth **E**nhancement **P**rocess*) que en español vendría a ser el "Proceso de resaltamiento de la profundidad de las imágenes visuales" produce imágenes que tienen profundidad realista y llenura, en lugar de las exageradas imágenes que parecen brincar afuera de la pantalla.

Este sistema se basó en el estudio de cómo una persona visualmente dañada percibe profundidad, el resultado del estudio fue que una persona tuerta percibe

profundidad moviendo su cabeza y comparando una secuencia visual de marcos desde ángulos diferentes, aunque cabe anotar que este concepto era concebido como imposible por algunos fisiólogos.

VISIDEP simula la óptica de una persona con un solo ojo; usando equipo de video convencional, más un dispositivo de codificación especial que produce un despliegue, tiempo-secuencia de imágenes capturadas desde dos puntos de diferente vista, en un simple canal.

La técnica de sistema entrelazado utiliza gafas shutter para hacer que cada ojo vea los marcos correspondientes a cada ojo, en cambio VISIDEP confía en la habilidad del cerebro en fundir imágenes presentadas rápidamente.

Una vez codificada la imagen, puede ser reproducida por cualquier simple cámara convencional de video, cine o proyector de diapositivas. La profundidad pasa a la pantalla en lugar de afuera hacia el publico, haciendo que la imagen parezca mas viva.

CAPÍTULO II

2 PROYECTOS Y APLICACIONES

2.1 DETALLE DE LOS PROYECTOS EXISTENTES DE TELEVISIÓN ESTEREOSCÓPICA.

Los proyectos existentes en la actualidad han sido desarrollados en Japón, Estados unidos y Europa principalmente. La actividad de estos proyectos ha atraído a numerosas Instituciones y Compañías que se encuentran interesadas en un amplio rango de actividades relacionadas con la televisión tridimensional, principalmente en tres áreas de la visión estereoscópica como son: la psico-óptica, dispositivos tecnológicos y el procesamiento de las señales.

El conocimiento de factores de la visión humana es esencial para el diseño y la realización económica de cámaras 3D, para la generación de imágenes en computadora, transmisión de señales así como el diseño de pantallas. Los primeros intentos, realizados con éxito, involucran ayudas visuales como gafas anaglifas, lentes polarizados entre otras.

Los desarrollos de última generación apuntan a métodos auto estereoscópicos, en los que no se necesita la ayuda de lentes. Con los métodos de interpolación de procesamiento de la señal se evita el cansancio e incomodidad de la vista. Una pregunta futura será si es que se hará y cómo se logrará que la holografía pueda ser incluida en los sistemas de televisión tridimensional.

A continuación se detallan algunos de los proyectos más conocidos dentro del campo de creación de imágenes estereoscópicas.

2.1.1 PROYECTO COST 230

El proyecto COST (European **CO**operation in the **S**cientific and **T**echnical field) nace a principios de la década de los 70' y representa la primera forma de colaboración científica sistemática del continente europeo en este campo.
A partir de 1991 se crea el proyecto COST 230 (Stereoscopic television – standards, technologies and signal processing), que investiga las posibilidades de una imagen espacial real basadas en métodos estereoscópicos, la cual se divide en tres grupos específicos de trabajo investigativo, que enfocan los siguientes aspectos:

- ***Factor humano en la televisión estereoscópica (3DTV):*** aspecto psico-óptico de la visión binocular, teoría de la producción de TV estereoscópica y metodología de evaluación de la calidad de la imagen estereoscópica.

- ***Tecnología en la televisión estereoscópica:*** dispositivos de adquisición de imágenes estereoscópicas, dispositivos de grabación, dispositivos de mezcla y edición.

- ***Formación y transmisión de la señal de televisión estereoscópica:*** técnica de codificación, interpolación y síntesis de imágenes virtuales.

2.1.1.1 Factor humano en el proyecto COST 230.

El factor humano en la realización de un sistema de televisión estereoscópico es muy importante para el proyecto COST 230, ya que al querer imitar el sistema de percepción de imágenes tridimensionales se debe tener un buen conocimiento de la psico-óptica de la visión binocular, que viene a constituir la forma en que nuestro cerebro actúa con las imágenes que captan nuestros ojos para producir una imagen en tres dimensiones.

2.1.1.1.1 Métodos de evaluación

Material específico para pruebas estereoscópicas fue producido tanto con cámaras estereoscópicas experimentales como con modelos de prueba generados en computadora, utilizándolo para investigar parámetros de cámaras y displays así como aparatos para compresión de imágenes estereoscópicas.

Con este material se simularon errores de cámara mediante el desplazamiento específico entre las imágenes de los ojos izquierdo y derecho, consiguiendo así establecer los límites permisibles para esta clase de distorsión en una secuencia de pares estereoscópicos que contienen diferentes magnitudes de estos desplazamientos. Como consecuencia de esto se esperaba una inmediata reacción basada en estimulo-respuesta así como efectos que ocurrirían debido a una prolongada exposición frente al display.

Como las investigaciones de los factores humanos no están relacionados a una tecnología especifica de display, un rango de diferentes tipos de displays estereoscópicos se usaron, dependiendo de lo que parecía ser mejor para cierta tarea y que estaba disponible en el laboratorio. Así los displays experimentales incluyen presentación secuencial en monitores que presentan 100 campos por segundo (visto a través de gafas electro ópticas de obturador), proyección de video estereoscópico con estándar y equipo de HDTV, así como slides estéreo y películas estereoscópicas de 35 mm (vistas a través de gafas polarizadas).

Las personas que intervinieron en las pruebas fueron de una gran variedad de ambientes educacionales y ocupacionales, que presentaban una agudeza visual normal (mejor que 1 minuto de arco $[minarc]^2$ de agudeza monocular y capacidad de discriminación de profundidad), generalmente personas sin experiencia en imágenes 3D.

Los procedimientos experimentales se basan en la recomendación de la UIT-R BT. 500-7 (Metodología para la evaluación subjetiva para la calidad de imágenes

de TV). De esta manera se utilizó una escala de cinco grados de calidad de cuadro (que va desde "Excelente" hasta "mala") y degradación visual (desde "imperceptible" hasta "muy molesta") como se muestra en la tabla 3.1, para evaluar efectos psico-ópticos primarios de errores de cámara y parámetros de display, el método de elección entre opciones predeterminados se utilizó para establecer umbrales de visibilidad para una interferencia específica.

ESCALA DE 5 NOTAS	
Calidad	**Degradación**
5 Excelente	5 Imperceptible
4 Buena	4 Perceptible pero no molesta
3 Aceptable	3 Ligeramente molesta
2 Mediocre	2 Molesta
1 Mala	1 Muy molesta

Tabla 2.1 Escalas de calidad y degradación de la UIT-R

Para evaluar los efectos relevantes psicológicos que no están dentro de los procedimientos psico-físicos tradicionales, se aplicaron métodos de aproximación de valoración multidimensional recientemente desarrollados. De acuerdo con estos métodos, respuestas transitorias (efectos inmediatos de un cuadro) fueron recolectadas con una palanca de mando durante la presentación del cuadro. Adicionalmente medidas de presentación posterior como cuestionarios de evaluación o entrevistas fueron aplicadas para recoger una diferenciada y bien considerada reflexión de experiencias y actitudes.

2.1.1.1.2 Beneficios subjetivos específicos de sistemas avanzados de 3DTV.

Se anticipa que la televisión estereoscópica (al igual que la HDTV) podría eventualmente cubrir un rango de aplicaciones domésticas y de negocios como la difusión de televisión, video - telefonía y video conferencia. Por consiguiente se considera importante examinar las ventajas especificas de 3D en diferentes campos de aplicación desde el punto de vista del usuario.

---

[2] **minarc**: minuto de arco. sesentava parte de un grado.

En lo concerniente a la *difusión de televisión estereoscópica* basados en los métodos subjetivos de evaluación de 3D versus 2D (HDTV), se obtiene de un resumen de las pruebas realizadas, una clara preferencia por las presentaciones estereoscópicas debido a la sensación intensa y satisfactoria, al mostrar a los participantes escenarios idénticos para 2D y 3D, siendo elegidas como más atrayentes e interesantes (por el 88% de participantes) dando un claro favoritismo a la televisión estereoscópica (3D).

Por otro lado el estudio muestra también factores limitantes, a pesar de que algunos efectos especiales y probablemente molestos (como por ejemplo el salto de imágenes fuera de la pantalla o espacio exagerado estereoscópico) habían sido evitados, los sujetos de prueba indicaron una cantidad significativa de micro eventos 3D desagradables, siendo mayor el número de estos conforme aumentaba el tiempo de exposición frente al display.

En cuanto a la distancia del observador con respecto al display (un parámetro que influye directamente en el requerimiento espacial de resolución del display) en un mismo porcentaje fueron preferidas tanto pantallas grandes de HDTV como las de 3DTV mostrando que ambos sistemas pueden ser compatibles. Determinándose así que es aconsejable una distancia entre 3 y 4 veces el alto del cuadro.

En otro experimento fueron investigadas las ventajas de imágenes 3D versus imágenes 2D en una aplicación de *video conferencia*, esperando que algunos aspectos de una conversación cara a cara que se pierden con equipo de televisión convencional se puedan transmitir con técnicas estereoscópicas apropiadas.

Se desea obtener imágenes tridimensionales de tamaño real de los conferencistas y sus ambientes con adecuada resolución espacial con el ajuste de la perspectiva individual y paralaje de movimiento natural. Además de un contacto real entre los ojos de los conferencistas, ya que en los sistemas convencionales la cámara esta ubicada sobre el display como se muestra en la

figura 2.1, provocando así que exista un ángulo de defasaje entre el eje de enfoque de la cámara y la línea de enfoque entre los interlocutores.

Esta desviación impide que los interlocutores tengan un real contacto entre sus ojos. Para eliminar la desviación angular dos cámaras se colocan, una al lado derecho y otra al izquierdo del display (2D o 3D) para formar una base estereoscópica.

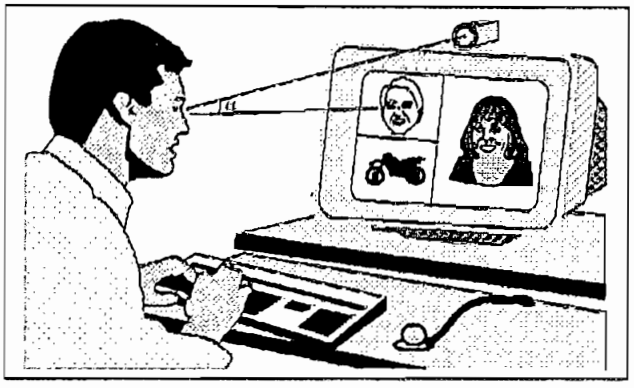


Figura 2.1 Sistema de videoconferencia convencional.

El análisis de imágenes tridimensionales y técnicas de síntesis se desarrollaron para construir imágenes de cámaras virtuales las cuales pueden aparentar estar puestas delante de los ojos del conferencista. Las desviaciones horizontal y vertical pueden ser corregidas por este sistema. Resultados basados en entrevistas y encuestas mostraron que los efectos de tele presencia tiene un refuerzo con displays estereoscópicos

2.1.1.1.3 Requerimientos de cámara y display

Los displays estereoscópicos proveen una representación visual inequívoca de la estructura espacial natural y de las imágenes generadas en computadora. Esto demuestra una ventaja sustancial sobre los displays 2D en varios campos de aplicación. Por otra parte, los usuarios de displays estereoscópicos se quejan a menudo de molestias visuales, como fatiga visual siendo básicamente el resultado del hecho de que la tecnología 3D actual puede aproximar, pero no copiar todas

las propiedades de los arreglos explotados por la visión binocular en un ambiente natural.

Un gran paralaje binocular en un display estereoscópico, por ejemplo, tiende a producir dolor de cabeza, mientras que bajo condiciones de visión natural, el sistema visual es capaz de cubrir disparidades de cualquier magnitud. Como consecuencia, el diseño de sistemas estereoscópicos de alta calidad debe tener en cuenta cuidar diferencias molestas entre la visión en displays y la visión natural bajo los principios subjetivamente tolerados.

2.1.1.1.3.1 Tomas de imágenes de televisión estereoscópica: Requerimientos de cámara.

Con despliegues estereoscópicos, los dos ojos de un observador han de recibir dos imágenes diferentes pero muy bien emparejadas, siendo las únicas diferencias entre dichas imágenes el desplazamiento entre los dos centros de perspectiva del sistema de imágenes usado para crear el par estereoscópico. En la visión natural, el sistema visual humano puede entonces ser capaz de evaluar estas diferencias de percepción de profundidad. Sin embargo debido a imperfecciones tecnológicas y/o ajustes incorrectos durante la producción, almacenamiento, transmisión y despliegue; pueden surgir diferencias adicionales entre las dos imágenes constituyentes de un par estereoscópico, tales errores pueden estorbar o incluso impedir la fusión binocular[3].

Es así como se debe encontrar límites admisibles para los errores de las imágenes, y de esta manera conocer los requerimientos específicos en el diseño de un equipo estereoscópico. Entre los errores más comunes encontrados en las cámaras se pueden mencionar los siguientes:

- ***Rotación o error de inclinación***, que ocurre si el eje vertical del censor de imagen de los ojos derecho e izquierdo no están alineados

[3] **Fusión Binocular** es el proceso de acomodación y enfoque que realiza nuestro cerebro para a partir de dos imágenes observadas por cada uno de los ojos, obtener una sola imagen nítida.

paralelamente. En equipo de cámara estereoscópica la rotación admisible de una o dos cámaras alrededor de su eje óptico es de 0.5 grados.

- ***Diferencia de longitud focal,*** que es la diferencia entre las dos lentes de un sistema de cámara estéreo, que provoca una amplificación desigual de las imágenes del ojo derecho e izquierdo. La diferencia de longitud focal entre dos cámaras no puede exceder un valor del 1%.

- ***Diferencia en contraste,*** es la afección del contraste lumínico de una imagen debido al cambio de la configuración en los niveles de blanco y negro. Se permite que entre las dos cámaras exista 1.5 dB de diferencia en el nivel de blanco y 0.1 dB de diferencia en el nivel de negro.

- ***Desviación de luminancia,*** en un par estereoscópico la diferencia de luminancia estacionaria afecta a la imagen entera debiendo no exceder de 3 a 6 dB. En particular áreas de superficie con diferentes niveles de intensidad deben ser iluminados para evitar molestos efectos de oposición binocular.

2.1.1.1.3.2 Presentación de imágenes de 3D-TV: Parámetros de display

Para reforzar la ilusión de presencia material ilustrada por las grabaciones de una cámara o por la generación de escenas en una computadora, es necesario desplegar la información de tal manera que se asegure una percepción sin distorsiones y una visión confortable.

Para asegurar una buena percepción se requiere un ajuste perfecto de la visión binocular por medio de una correcta alimentación monocular en la percepción del tamaño y distancia del espacio gráfico. La correspondencia insuficiente entre varios objetos aparenta distancias que no son verdaderas, con transferencia de paralaje estereoscópico y es probable que el tamaño angular percibido monocularmente haga que estos objetos se miren distorsionados en tamaño y/o

forma, la fuente de estos deterioros, incide e impacta en varias aplicaciones prácticas.

Para mantener una confortable y buena visión la magnitud de la disparidad retinal causada por cualquier par de objetos en el campo de la visión aguda no debe exceder un cierto límite. Es claro que el tamaño del cuadro influye en la impresión de realismo, por lo cual es razonable suponer que cuadros 3-D deben preservar un tamaño mínimo de tal manera que se evite el llamado efecto de teatro de marionetas[4].

El paralaje binocular esta enfocado a la medición de la distancia de objetos y dispara un proceso de ajuste de imágenes retinales de acuerdo a las leyes de la estereometría. Este mecanismo explica por qué en los ambientes reales el tamaño percibido de un objeto permanece constante independientemente de su distancia y dependiente del tamaño angular (fidelidad de tamaño). Si el mismo principio se mantiene para la percepción de cuadros 3D, los objetos pueden ser mirados agrandados o minimizados, siempre que la proporción entre su tamaño angular (tamaño en la pantalla) y la distancia estereoscópica (paralaje en la pantalla) difiera de la proporción correspondiente al mundo real. Una descripción de este proceso se presenta en la figura 2.2.

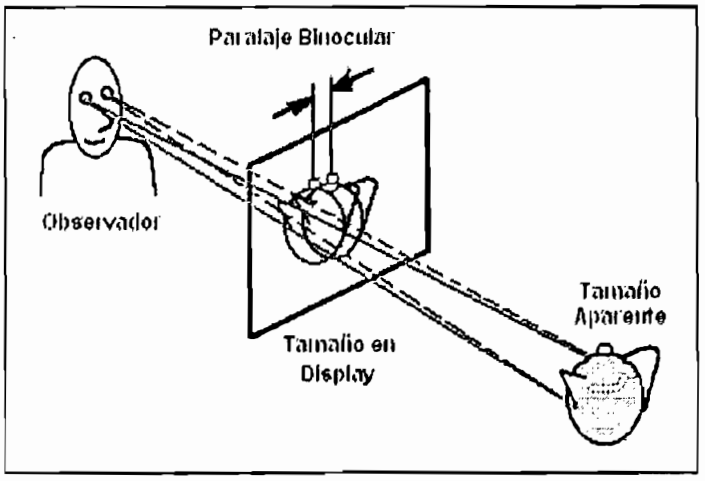


Figura 2.2 Percepción del tamaño en un display 3D.

[4] **UIT-R BT.1438 "Efecto teatro de marionetas:** describe un tipo de distorsión en imágenes 3D. A veces, los objetos estereoscópicos se perciben como anormalmente grandes o pequeños".

Como resultado de los experimentos realizados para establecer un tamaño de cuadro mínimo para mostrar imágenes estereoscópicas no se llegó a un resultado claro y contundente. Ya que la respuesta psicológica de mostrar imágenes estereoscópicas en diferentes tamaños, demostró ser bastante flexible.

Excesiva disparidad puede ocasionar numerosos fenómenos molestos (como presión en los ojos, tirones extraños en los músculos oculares y dolor de ojos), aunque la fusión aún es posible. Siendo especialmente molestos por el hecho de que los observadores no distinguen su origen y los afecta sin una advertencia previa. Usualmente se establece un límite de 70 minarc que fue encontrado de la apertura de los ojos de los humanos y de su profundidad de enfoque. Disparidades sobre los 35 minarc pero menores que 70 minarc no causan ninguna molestia, disparidades sobre los 70 minarc se deben evitar en despliegues que proporcionen una alta resolución espacial.

Con displays 3D es imposible separar completamente lo observado por el ojo izquierdo y el ojo derecho, debido a que un porcentaje de la imagen derecha es visible en el ojo izquierdo y viceversa. El crosstalk Interocular[5] está normalmente en el rango de 0.1 a 0.3 % con técnicas de polarización y en el rango de 4 a >10% con display 3D de tiempos multiplexados.

El crosstalk produce doble contorno (desdoblamiento de imagen) y es una causa potencial del dolor de cabeza en los espectadores. Los resultados muestran que la visibilidad de crosstalk aumenta con el incremento del contraste y el incremento de disparidad binocular (profundidad) de la imagen estereoscópica como se puede observar en la figura 2.3. Para producir un razonable rango de profundidad (sobre los 40 minarc) en un display de contraste alto (100:1), el crosstalk debería ser tan bajo como el 0.3%.

[6] **Crosstalk interocular:** se denomina así a lo que es observado por un ojo que no debería ser visto por este, por ejemplo la parte de la imagen derecha que es vista por el ojo izquierdo, esto produce una imagen de doble contorno.

En un display 3D de multi-vistas la perspectiva observada cambia con la posición de la cabeza del observador, proveyendo así un efecto de "mirar alrededor". Los saltos notables de imagen que se producen desde una perspectiva vista a la siguiente (image flipping[6]) perjudica grandemente la integridad del espacio estereoscópico. Un gran número de diferentes vistas es requerido para que el efecto flipping[7] sea imperceptible. Se espera que el crosstalk entre vistas adyacentes reduzca el efecto flipping creando una visión débil en las cercanías.

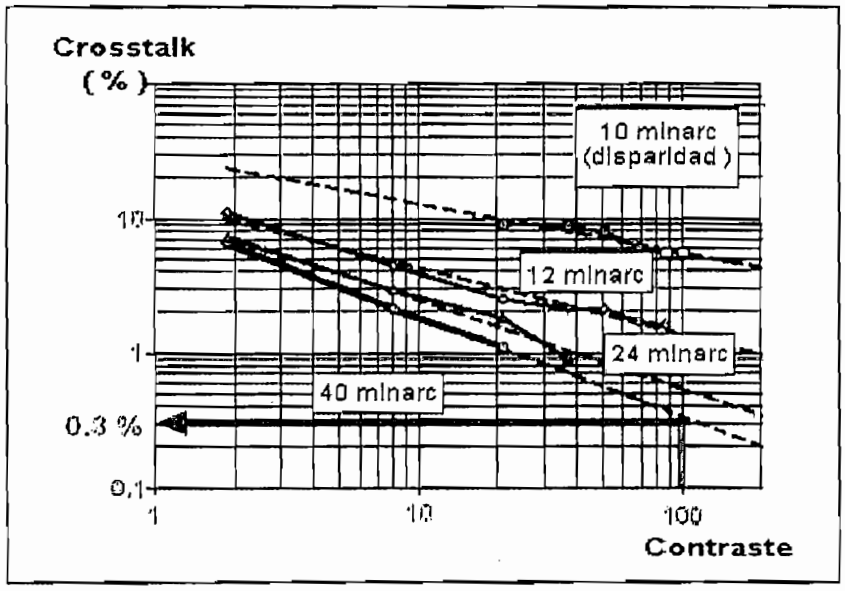


Figura 2.3 Umbrales de visibilidad para crosstalk como una función de contraste local y disparidad binocular.

2.1.1.2 Tecnología en el proyecto COST 230

Los logros tecnológicos dentro del proyecto COST 230 han desarrollado la creación de un sistema para TV estereoscópica y displays estereoscópicos. El sistema de televisión estereoscópica se encuentra estructurado en su forma general por: colector de imagen, grabación, mezcla y edición.

2.1.1.2.1 Componentes del sistema

Colector de imágenes.- Para la mayor parte de recolección de imágenes, se utiliza cámaras estereoscópicas. Estas usualmente tienen arreglos de dos o mas cámaras de TV con idénticas separaciones horizontales entre sus ejes ópticos. Las cámaras simples pueden tener una configuración del arreglo geométrico, mientras las mas sofisticadas tienen servo control del ángulo de convergencia de los ejes ópticos de dos cámaras, distancia de separación de cámara y tres funciones de lentes (acercamiento o zoom, iris y enfoque).

A menos que las imágenes sean procesadas posteriormente de alguna manera para quitar errores antes de presentar a los espectadores, las cámaras individuales deben ser alineadas cuidadosamente en los tres ejes espaciales. Como ejemplo de cámaras estereoscópicas tenemos a las desarrolladas dentro los proyectos MIRAGE y DISTIMA.

Grabación.- Grabaciones exitosas de dos canales (3D) han sido hechas en varios proyectos, siendo el método preferido el que usa cámaras sincronizadas que guardan sus datos en dos VCR´s (Video Camera Recorder) de calidad de estudio con sincronía temporal. Para las aplicaciones industriales y de otro tipo donde se quiere usar un medio de bajo costo, las vistas de la imagen derecha e izquierda deben ser multiplexadas en una cinta simple, encontrándose además resultados satisfactorios al usar formatos de calidad doméstica en algunas circunstancias.

Es probable que el desarrollo de nuevos medios de grabación digital baje el costo de grabado, aunque se debe tener cuidado con los artefactos de cuadro digital, los cuales pueden ser aceptables en imágenes 2-D pero no en 3-D.

Mezcla y edición.- Ha habido varias demostraciones en el proceso de 3-D TV. Las experiencias de producción del proyecto MIRAGE en el programa "Eye to Eye" han sentado precedentes. El proceso de post-producción utilizó un estándar

---

[7] **Efecto flipping:** se denomina así al efecto que causa sobre un observador el image flipping.

de alta calidad y proceso de edición en serie, el único fragmento adicional de equipo fue un monitor 3-D con calidad de estudio. Un técnico en estereoscopía fue quien ayudo al director del programa para la operación de edición en serie. Se puede concluir que teniendo un cuidado adecuado, la post producción de video estereo de alta calidad puede ser llevada a cabo usando modernos y convencionales equipos de edición con ligeras modificaciones.

2.1.2 PROYECTO RACE II – DISTIMA (R- 2045)

El proyecto DISTIMA (**DI**gital **ST**ereoscopic **IM**aging & **A**pplications) fué desarrollado como parte de la segunda fase del proyecto RACE (**R**esearch and **D**evelopment in **A**dvanced **C**ommunications technologies for **E**urope) de la Unión Europea de proyectos, la cual se dedica a la integración de nueva tecnología y la creación de prototipos para nuevos servicios y aplicaciones. Siendo actualmente conocido internacionalmente por su importancia no solo en Europa sino también en Estados Unidos y Japón.

La meta del proyecto fue probar la viabilidad de una cadena que va desde la grabación - codificación - transmisión - decodificación hasta el despliegue de secuencias de video estereoscópico de dos canales, utilizando para ello la red IBCN (**I**ntegrated **B**roadband **C**ommunicate **N**etwork), red de comunicación de banda ancha integrada.

Como resultado, el proyecto espera aumentar la competitividad de la industria europea en los nuevos servicios de la IBCN como telefonía estereoscópica y video conferencia, así como la distribución de video, y en las aplicaciones profesionales de imágenes estereoscópicas como manejo remoto en las aplicaciones industriales, educación, medicina con video ayuda, entre otras muchas.

En cada una de las aplicaciones mencionadas, se requiere una alta calidad de señal de video digital estereoscópico de dos canales, donde cada uno de los canales tiene por lo menos la resolución indicada por el estándar de televisión

digital normal, es decir una resolución espacial y temporal según recomendación UIT-R BT.601.

La investigación realizada por DISTIMA reveló que los algoritmos MPEG de codificación/decodificación pueden manejar la tasa de datos asociada con imágenes estereoscópicas (3D), es así, como se ha proyectado que las imágenes televisivas estereoscópicas pueden transmitirse a 1.5 veces la tasa de la HDTV.

2.1.2.1 Arquitectura fundamental para el sistema de difusión de imagen estereoscópica.

En Diciembre de 1994 se realiza la primera transmisión mundial de video estereoscópico en tiempo real sobre una red ATM, acompañada por sonido estereofónico sorround[8], ésta tuvo lugar en Europa como resultado del proyecto RACE DISTIMA. Los datos globales de la velocidad de conexión eran de 10 Mbps, uniendo los laboratorios de investigación de KPN en Leidschendam, en Holanda, y los laboratorios de Telekom en Berlin, Alemania, usando el equipo técnico desarrollado dentro del proyecto. La figura 2.4 muestra un esquema de dicha transmisión.

El sistema experimental de transmisión de video DISTIMA ATM, puede manejar 4 canales de video y puede ser empleado para transmitir video estereoscópico o video 3D de múltiples vistas.

Se tiene tres opciones de entrada básicas al sistema de transmisión: entrada de video en vivo a través de un sistema de cámara de multivisión, entrada de flujo de bits desde un reproductor de datos (dataplayer) y entrada de video digital. Pudiendo aceptar cualquier sistema, estereo o de múltiples vistas, de código MPEG1 o MPEG2.

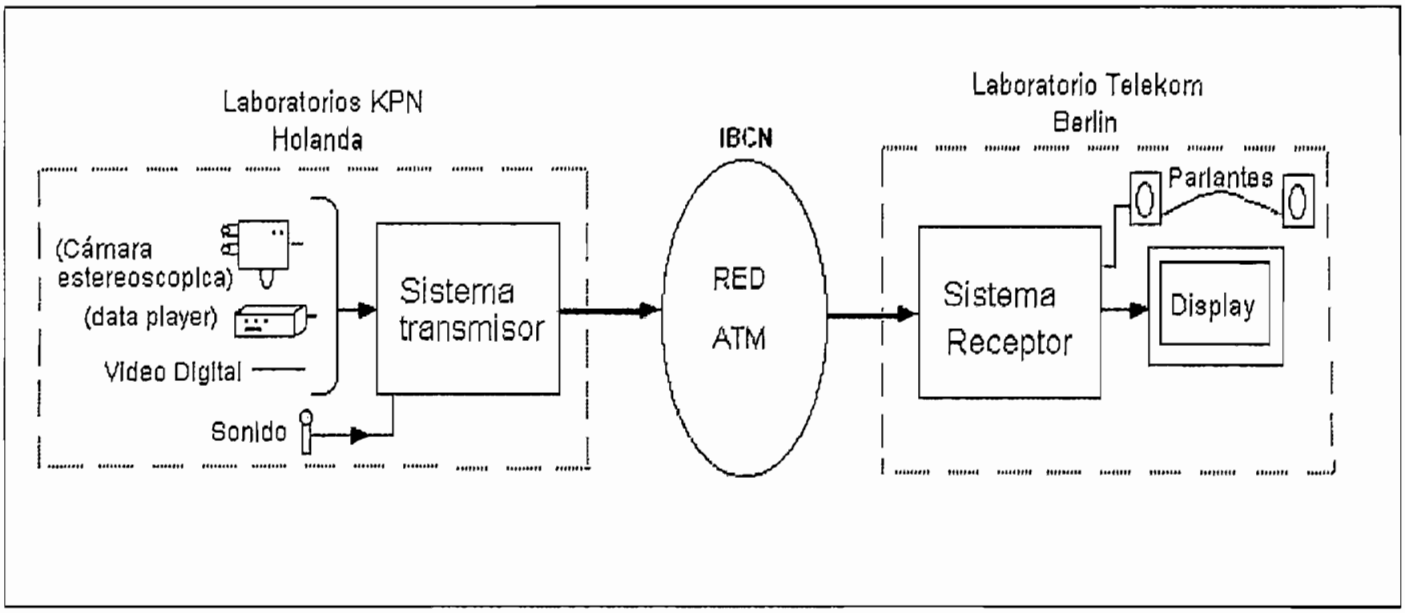


Figura 2.4 Esquema de transmisión del proyecto DISTIMA

El receptor DISTIMA podía manejar 4 canales de video para producir video tridimensional de múltiples vistas. Para reforzar el carácter especial de visión espacial el sistema DISTIMA estaba equipado con un sistema de sonido estereofónico Dolby surround.

El sistema de transmisión esta formado por un codificador estereoscópico, un multiplexor, un sistema de corrección de errores FEC, una capa de adaptación ATM (AAL), la capa ATM y la capa física.

El codificador estereoscópico lo conforma el sistema S-MPEG (Stereo-MPEG) desarrollado en DISTIMA, el cual codifica la señal de video del canal izquierdo mediante un tipo convencional de codificador híbrido DPCM/DCT que conforma esencialmente el estándar MPEG-2. Cualquier decodificador MPEG puede decodificar la señal codificada S-MPEG (solamente la vista del canal izquierdo). Para el canal de la vista derecha se utiliza un sistema de codificación basado en predicción, la predicción se obtiene no solo del cuadro anterior o siguiente (obteniéndose así los cuadros: predichos P y cuadro predichos bidireccionalmente

[8] **Sonido "Sorround"** que recrea el dramatismo, el ambiente y el realismo de los efectos especiales, mediante la codificación de un canal adicional de sonido posterior L/R, junto con la información de audio de un "canal central" para colocar las voces en su posición natural más cerca de la pantalla.

B), sino también del cuadro del canal izquierdo, conformando lo que se conoce como un codificador MPGE2 con estimación de disparidad y predicción compensada de disparidad de la señal de vista izquierda . Este modo alternante de predicción de cuadro incrementa la eficiencia del proceso de compresión para este canal.

El codificador S-MPEG propuesto por el proyecto, codifica los dos canales en un total de 9 Mbps, donde cada secuencia esta conforme a la UIT-R BT 601 (576 x 720 interlazado cada uno a 50 Hz). Un ancho de banda de 6 MHz es usado para los cuadros del canal izquierdo y de 3 MHz para el canal derecho.

2.1.2.2 Cámara avanzada de estudio - DISTIMA.

La cámara estereoscópica DISTIMA para estudio de televisión fue desarrollada a principios de los años 90. En la figura 2.5 se puede ver a la cámara en una toma. Esta versátil cámara de estudio de televisión es motorizada y maneja a control remoto las funciones de las lentes normales y los dos parámetros 3D, que son: separación de cámara (la distancia horizontal entre los ejes ópticos de las dos cámaras), y la distancia de la convergencia de la cámara (la distancia de la cámara a la que los ejes ópticos de las dos cámaras se interceptan).

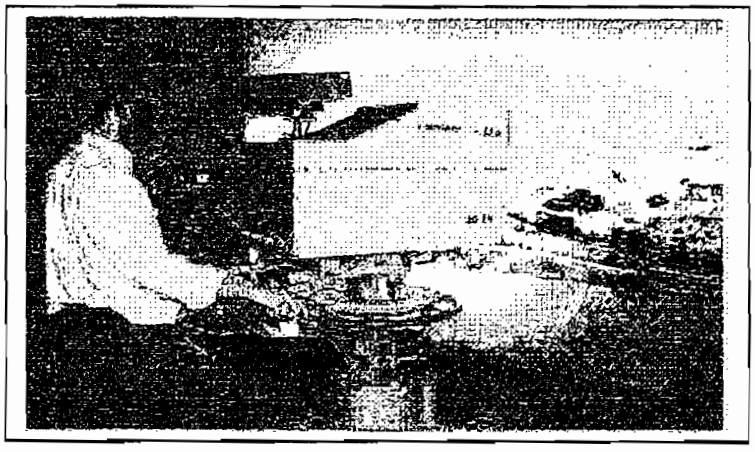


Figura 2.5. Cámara Avanzada de Estudio – DISTIMA

Las cabezas individuales dentro de la cámara 3-D fueron desarrolladas específicamente para la cámara 3-D por Multimedia Thomson. Durante el diseño de los sensores de la cámara se prestó particular atención a emparejar las dos cámaras para la geometría y colorimetría. Tal como los otros componentes de la cámara DISTIMA, los sensores de la cámara se diseñaron para controlarlos

remotamente por computador. Cada cámara diseñada usa una configuración de tres sensores CCD con un zoom de lentes de televisión de alta exactitud. El zoom de las lentes es modificado para dar una exactitud por servo control, así las dos lentes pueden ser operadas en forma sincrónica, permitiendo el zoom durante una filmación.

Para evitar el cansancio en el observador, deben emparejarse los dos cuadros para formar una imagen 3-D con precisión. También las pequeñas diferencias horizontales entre las dos imágenes (disparidad) le permiten al observador percibir profundidades, debiendo ser controladas con mucho cuidado.

La cámara es controlada por computadora con un interfaz de computador personal, siendo diseñada para comportarse y ser operada como cualquier cámara convencional de estudio de televisión. El interfaz de computador personal es usado para controlar los parámetros 3-D, calculando los valores usados en el contenido de profundidad de la escena y configurando los lentes.

2.1.2.3 Display estereoscópico

El objetivo de crear un display estereoscópico, era realizar un proyector estereo con una relación de aspecto de 4:3 y cuadros polarizados basados en tecnología LCD. Las principales características de tal sistema son: superposición geométrica de los cuadros izquierdo y derecho, emparejamiento fotométrico[9] entre los cuadros.

Varias clases de ajustes son realizadas, algunas debido a los problemas específicos de la estereoscopia, otras debido al sistema en sí, entre estas están:

- Registro de imagen: ajustes de cada compuerta (rotación, horizontal, vertical, enfoque)
- Contraste: ajuste de la polarización (rotación)

[9] **Emparejamiento fotométrico:** se refiere al emparejamiento de la intensidad de la luz entre las dos imágenes

- Dirección de la polarización: ajuste de ½ longitud de onda (rotación)
- Posición de cuadro: ajuste fino de la proyección de las lentes (horizontal y vertical)

Cada proyector se ajusta óptimamente con un filtro óptico para obtener un blanco D65 (aproximadamente). Un ajuste electrónico permite corregir el balance blanco. La lámpara es de metal – halide de 250 W con una eficiencia de aproximadamente 72 lúmenes / W. El sistema de iluminación proporciona 150 lúmenes de rendimiento a la salida del proyector. Las polarizaciones de salida son vertical y horizontal y una pantalla especial no polarizada es utilizada. La electrónica incluye tres partes, un rack " conversor 2:1" , un rack "procesador de video" y el procesador LCD, estas tarjetas son instaladas junto al proyector.

El conversor 2:1 realiza el cambio de campo interlazado con entradas 4:2:2 a un formato de salida de cuadro progresivo con norma europea (625/50/1). El rack del procesador de vídeo esta manejado por software a través de un interfaz de usuario para ajustar fácilmente ambos proyectores. En particular, el microprocesador maneja los modelos que permiten las medidas del color, esto también carga los coeficientes de la matriz de corrección de color.

El proceso LCD incluye una tarjeta análoga y un generador de base de tiempo para manejar cada compuerta. Es más, cada compuerta necesita un voltaje especifico de referencia para direccionarlo correctamente.

2.1.3 PROYECTO PANORAMA

El proyecto europeo PANORAMA (**PA**ckage for **N**ew **O**pe**R**ational **A**utostereoscopic **M**ultiview systems and **A**pplications), investiga en el desarrollo de hardware y software de un sistema auto estereoscópico de múltiples puntos de vista para ser usado en telecomunicaciones.

Este proyecto pretende construir un despliegue auto estereoscópico para realizar video conferencias estereoscópicas en tiempo real con adaptación del punto de vista. La meta es lograr una verdadera ilusión de telepresencia para los

compañeros remotos. Para este propósito, vistas intermedias en posiciones arbitrarias deben formarse a partir de vistas de un sistema estereoscópico de cámaras con una separación básica bastante grande, como se muestra en la figura 2.6. Esta separación es de 50 cm. para pantallas pequeñas y 80 cm. con pantallas grandes.

El punto de vista actual es adaptado de acuerdo a la posición de la cabeza del espectador, de tal manera que la impresión de paralaje en movimiento se produzca.

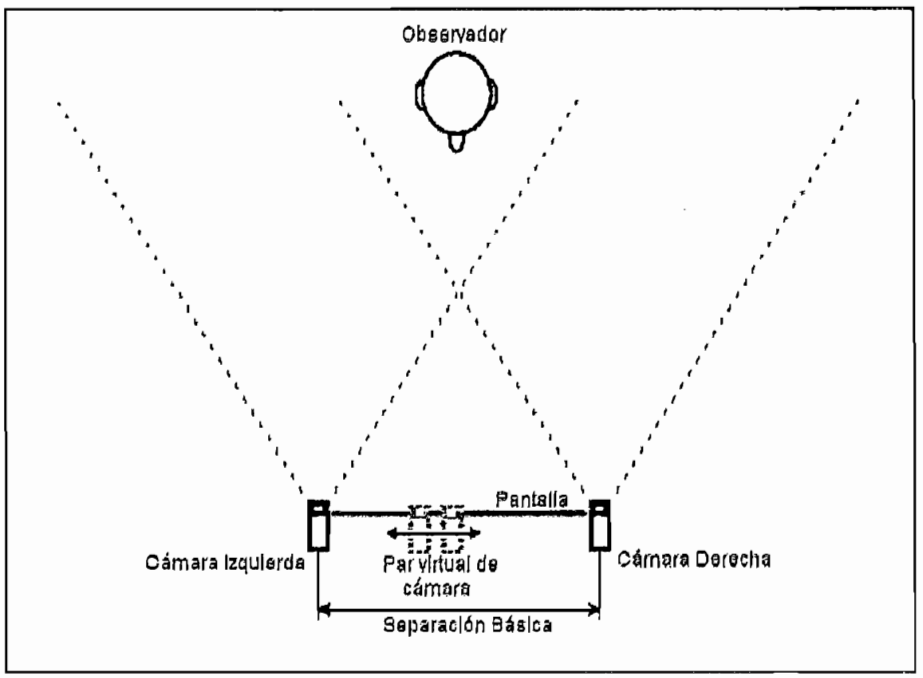


Figura 2.6 Arreglo de cámaras estereoscópicas sobre la pantalla y posición virtual variable de un par de cámara.

El sistema entero consiste de un estimador de disparidad[10], un codificador MPEG2 estereoscópico, codificador de disparidad y multiplexor en el lado del transmisor, y un demultiplexor, decodificador de disparidad, decodificador MPEG2 e interpolador con adaptación del punto vista en el lado del receptor. Como se muestra en la figura 2.7. Para la transmisión de la señal codificada una red ATM

[10] **Estimador de disparidad**, circuito que realiza el cálculo de disparidad entre las imágenes obtenidas por las cámaras derecha e izquierda.

es provista, siendo necesario un display auto estereoscópico para mostrar las imágenes.

Las señales de imagen de las vistas izquierda y derecha, además de la señal de audio , son codificadas por separado por codificadores MPEG-2 disponibles comercialmente.

Sin embargo es necesario proveer de un codificador por separado para el sub-muestreador de disparidad de campo que está fuera del estimador. El sistema multiplexor, que es compatible con el estandar MPEG-2, integra la disparidad codificada como un dato adicional en el flujo de datos, independiente de los datos de video. Además esto es necesario para sincronizar las imágenes independientes codificadas izquierda y derecha con los datos de disparidad.

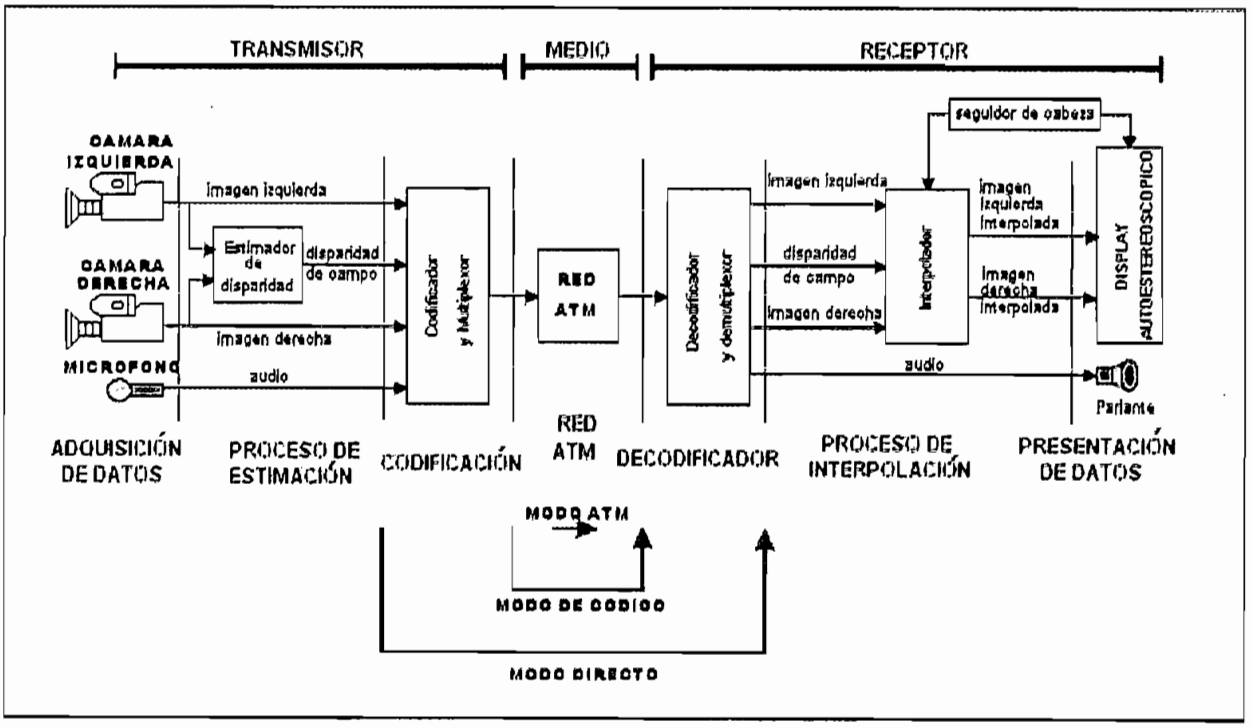


Figura 2.7 Diagrama de bloques de la cadena completa de sistema para proyecto PANORAMA

Para compensar el retardo del estimador de disparidad, una unidad de sincronización es insertada para asegurar la sincronización de los campos de disparidad y las secuencias de imagen grabadas a la entrada del codificador.

Al mismo tiempo, la información del seguidor de cabeza (headtracker), quien informa acerca de la posición de la cabeza del espectador, es usada para manejar el display auto estereoscópico, el cual es un sistema basado en la proyección hacia una pantalla lenticular y debe ser adaptado según el ángulo de observación, produciendo así que la impresión de paralaje en movimiento se produzca.

2.1.4 Proyecto MIRAGE (AC044)

El proyecto MIRAGE (**M**anipulation of **I**mages in **R**eal-time for the Creation of **A**rtificially **G**enerated **E**nvironments), desarrollado por ACTS (**A**dvanced **C**ommunications **T**echnologies and **S**ervices) fue comenzado a desarrollarse en Octubre de 1995, participando en su desarrollo los países de Estados Unidos, Alemania, Bélgica y España.

MIRAGE ayudó a la estrategia de la Comisión Europea para la implementación de la IBC (Integrated **B**roadband **C**ommunications) para proveer técnicas y tecnología para la producción de realidad virtual e imágenes estereoscópicas en ambientes virtuales. Además cubrió un número de proyectos ACTS como son:

- El escenario para la introducción de televisión digital y servicios interactivos multimedia.
- Análisis avanzado para imágenes e interpretación por telepresencia.
- Construcción de un modelo 3-D y síntesis para imágenes por telepresencia.
- Aspectos de comunicación de presencia virtual.
- Telepresencia y demostradores multimedia.

MIRAGE además esta dirigido a la creación y manipulación de sistemas y técnicas para televisión en realidad virtual para ser usado en difusión, multimedia, interactividad y tele presencia, teniendo como objetivos los siguientes:

- El desarrollo de una producción virtual a un precio económico.
- El desarrollo y demostración de sistemas de hardware y software para el uso de creadores de programas tradicionales.
- El desarrollo de técnicas de producción, definición de sistemas prácticos funcionales y dirección en problemas de estandarización.

El proyecto MIRAGE abarcó varios sistemas para la producción, creación de programas, post-producción y capturas de imágenes estereoscópicas en movimiento. El trabajo del proyecto abarca varios grupos de trabajo como son:

La producción virtual que explora y desarrolla nuevas plataformas y técnicas, en la producción de realidad virtual para el uso sencillo en la creación de nuevos programas con la ayuda de una práctica multicámara ligera. Este desarrollo se inicio con sistemas que controlan el movimiento de las cámaras y la unión de éstos, con dispositivos que introducen aleatoriamente imágenes grabadas o repetidas que se guardan en disco duro.

Edición virtual de series desarrolla un sistema de edición en tiempo real basado en gráficos de computadora y una estación de trabajo para el control, manipulación y alteración de ambientes virtuales. Esto se usará como una herramienta en la pre-producción para crear ambientes, como un sistema de edición en línea y en vivo para el uso durante la producción (es decir para el movimiento de la cámara y el control de iluminación) como un sistema de edición para la alteración de ambientes en post-producción.

Host virtual que realiza la creación y manipulación de caracteres virtuales a ser usados como organizadores del programa o presentadores. Este desarrolla en tiempo real sistemas de actuación animada para controlar caracteres de primer plano en un generador computarizado y adquiere ambientes que realmente imitan

movimientos humanos, respuestas y emociones. Los sistemas son basados en el gesto, movimiento y reconocimiento del dialogo.

Imágenes estereoscópicas enfoca aspectos de 3-D, es decir, producción de televisión estereoscópica para la integración en ambientes virtuales. Está dirigido a equipo, métodos de producción y problemas relacionados con factores humanos. El uso de cámaras estereoscópicas para la adquisición de la imagen en telepresencia fue investigado y una cámara liviana tridimensional ha sido desarrollada.

La Arena de los Juegos virtuales es donde todos los proyectos trabajan juntos para los ensayos y la creación para una muestra de Juegos virtuales en Europa a ser jugados a través de límites internacionales.

Dentro del proyecto MIRAGE existen dos tipos de ensayos:

El primer tipo consistió en ***comprobación continua de desarrollo de hardware y software en el estudio*** que concluyó con un segmento de programa permitiendo demostrar tanto fracasos como éxitos del sistema. Este material se usó para los informes, exhibiciones, demostraciones, entrenamiento y ensayos de la red, pasando al proyecto de demostración para evaluaciones subjetivas y psicológicas.

El segundo ensayo está dentro de la ***Arena de los Juegos virtuales*** que reúne las tecnologías desarrolladas y técnicas en la creación de una arena de juegos, donde compiten jugadores de diferentes países de Europa. Un programa grabado de ensayo se ha hecho y las negociaciones continúan para la transmisión de una serie del programa en el futuro.

2.1.4.1 Principales logros del proyecto

La tecnología de estudio de realidad virtual permite hacer los programas donde el paisaje no es más que un modelo estereoscópico cargado en memoria de computadora con fondos estereoscópicos.

Una de los propósitos de la tecnología existente es entregar imágenes estereoscópicas de alta calidad en los hogares, siendo éstos receptados en dispositivos económicos al alcance de todos.

El desarrollo del proyecto MIRAGE incluyó un programa de 25 minutos, llamado "eye to eye" para televisión virtual. Dos cámaras estereoscópicas fueron usadas: la cámara de estudio europea (construida por el proyecto EC RACE DISTIMA) que fue fabricada para la ITC por AEA Technology y una cámara de peso liviano para aplicaciones móviles.

El programa "eye to eye" da una revisión breve de la historia de la imagen estereoscópica desde las primeras fotografías hasta la filmación del cine en 3-D, y muestra alguna de las posibilidades para el futuro. Esto incluye: presentación, efectos especiales, juegos computarizados y realidad virtual. El programa fue mostrado usando tecnología de displays, estos incluyen dos monitores de imágenes observadas con gafas polarizadas especiales, donde el receptor de televisión trabaja a una frecuencia de 100 Hz., alternando imágenes entre el ojo derecho y izquierdo a una frecuencia de 50 Hz.

Figura 2.8 Cámara de estudio europea.

Con la experiencia ganada con eye to eye se desarrolló una segunda fase en receptar imágenes estereoscópicas, el diseño y construcción de una nueva cámara liviana de tele presencia.

Con lo cual se ha desarrollado un prototipo de cámara de peso liviano exacta y fiable usadas en aplicaciones de transmisión donde se usan cámaras pequeñas y rápidas.

El diseño permite lentes intercambiables, alineación geométrica de dos cámaras, mando manual de la separación de la cámara y convergencia. La cámara mostrada en la figura 2.9, fue diseñada para ser montada y trabajar confiablemente en un automóvil.

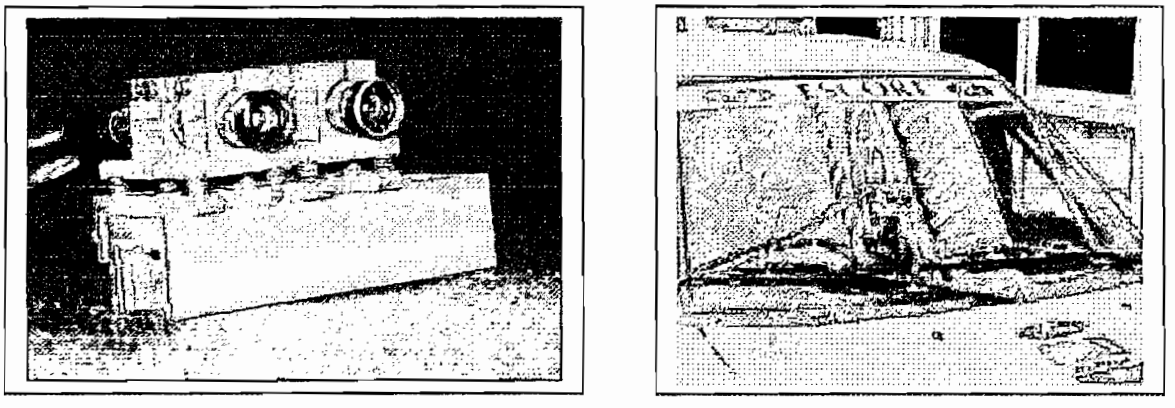


Figura 2.9 Cámara liviana de telepresencia 3-D

2.2 APLICACIONES

Desde hace mucho tiempo la estereoscopía ha despertado gran interés en los seres humanos por la ayuda que ha prestado en varios campos, y gracias al desarrollo tecnológico se facilita el uso de la televisión estereoscópica en ambientes donde es indispensable una buena apreciación de profundidad y volumen de las imágenes desplegadas.

Dentro de los campos de aplicación tenemos:

- Medicina
- Topografía y estudio del terreno.
- Estudio de la tierra y otros planetas

- CAD (Diseño Asistido por Computadora) y CAE (Ingeniería Asistida por Computadora)
- Ingeniería molecular
- Telepresencia
- Realidad Virtual

2.2.1 MEDICINA.

En este campo la generación de imágenes estereoscópicas proporciona una gran ayuda en la enseñanza, interpretación de imágenes para el diagnóstico y hoy en día presta una ayuda notable en intervenciones quirúrgicas.

En la enseñanza tiene evidentes aplicaciones en la visualización de imágenes o modelos en el interior del cuerpo humano, sean estos generados artificialmente o a partir de imágenes reales obtenidas por medio de TAC (Tomografía Axial Computarizada) o RMN (Resonancia Magnética Nuclear). En la figura 2.10 se muestra una exposición de anatomía virtual utilizando gafas para la visión estereoscópica realizado en la escuela de medicina de los Ángeles, Universidad de California.

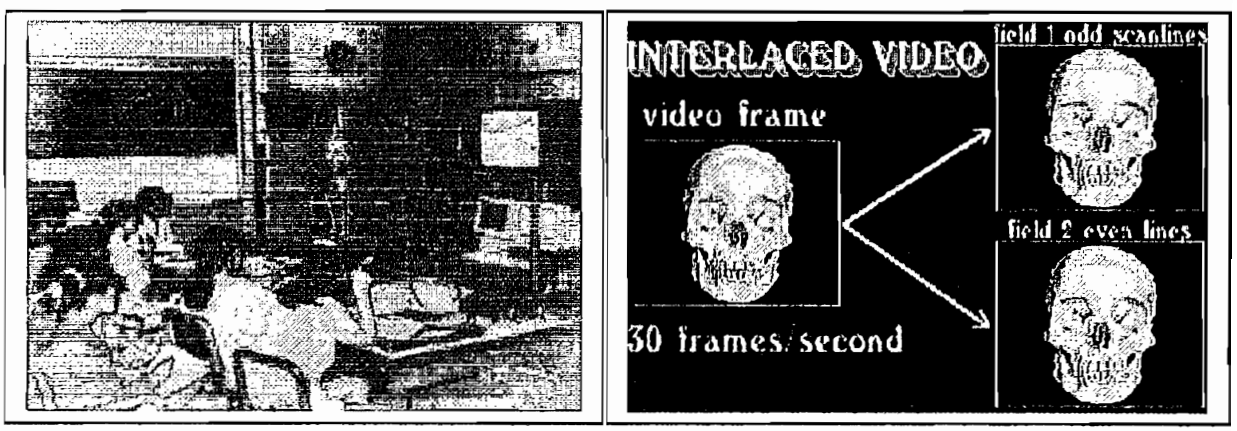


Figura 2.10 Exposición de Anatomía con gafas estereoscópicas.

Técnicas como la radiografía estereoscópica o imágenes de ultrasonido estereoscópico permiten situar cuerpos extraños y anomalías dentro del paciente, además se ha encontrado aplicaciones para el diagnóstico de enfermedades oculares e inclusive para distraer a los pacientes en odontología.

En el campo de la microcirugía ofrece grandes posibilidades, de tal forma que se cuenta ya con un sistema de microcirugía tridimensional conocido como el MediLive 3D, también existe un equipo denominado Vrex, el cual cuenta con un sistema de microcirugía orientado a la endodoncia. Ambos sistemas usan un multiplexor para enlazar las imágenes izquierda y derecha, mientras la visualización estereoscópica se consigue con gafas de cristal liquido. También la endoscopía cuenta con gran ayuda gracias a la estereoscopía, una ventaja de este sistema es que todo el equipo quirúrgico puede observar en una gran pantalla tridimensional una intervención si esta dotado de gafas para la visión estereoscópica, en la figura 2.11 se muestra una intervención quirúrgica con ayuda de un laparoscopio estereoscópico, además las imágenes tridimensionales pueden grabarse en video para su estudio posterior o emplearlas en docencia.

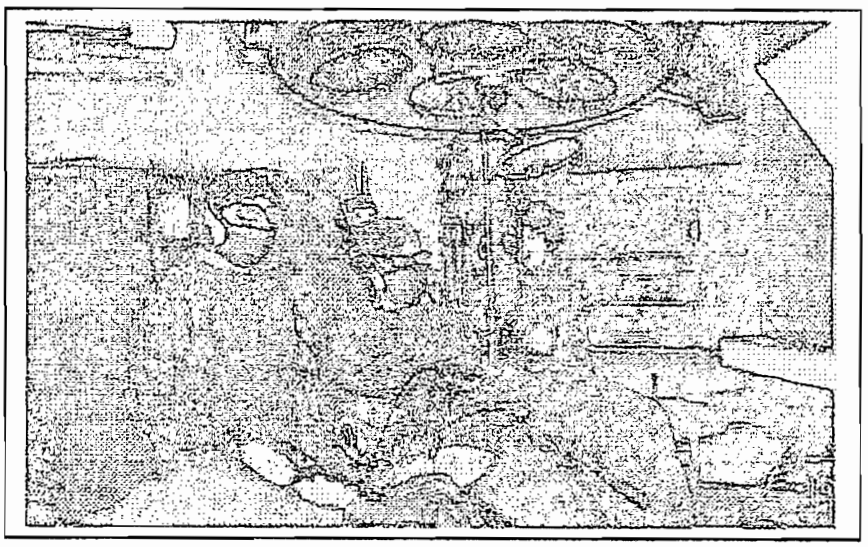


Figura 2.11 Operación mediante laparoscopía estereoscópica.

2.2.2. TOPOGRAFÍA Y ESTUDIO DEL TERRENO

Esta es una de las aplicaciones prácticas más antiguas en las que se han utilizado técnicas estereoscópicas para la visualización y medición del relieve terrestre, mediante fotografías aéreas, donde desde un avión se toman dos imágenes de una zona de terreno con una cierta distancia entre ellas para obtener un estereo–par que posteriormente se verá en relieve mediante un estereoscopio

especial, permitiendo calcular elevaciones del terreno mediante estereocomparadores.

Hoy en día se utilizan estos datos para generar imágenes 3-D simuladas mediante software. Una de las últimas técnicas de estudio de terreno se ha adaptado para ser utilizada bajo el agua mediante el uso de un sonar para obtener imágenes del relieve del fondo marino, donde datos sonoros son adaptados para su utilización en la generación de imágenes con perspectiva estereoscópica. En la figura 2.12 se muestra los equipos que componen al Mini-Rov HYDRATEC 3D, que es un sistema estereoscópico de TV3D, el cual nos permiten visualizar el relieve del fondo marino.

Un reciente ejemplo de trabajo topográfico es el realizado en febrero del 2000 desde el transbordador espacial Endeavour, dentro del proyecto SRTM (Shuttle Radar Topography Mission), que permite obtener mapas tridimensionales de una resolución extraordinaria.

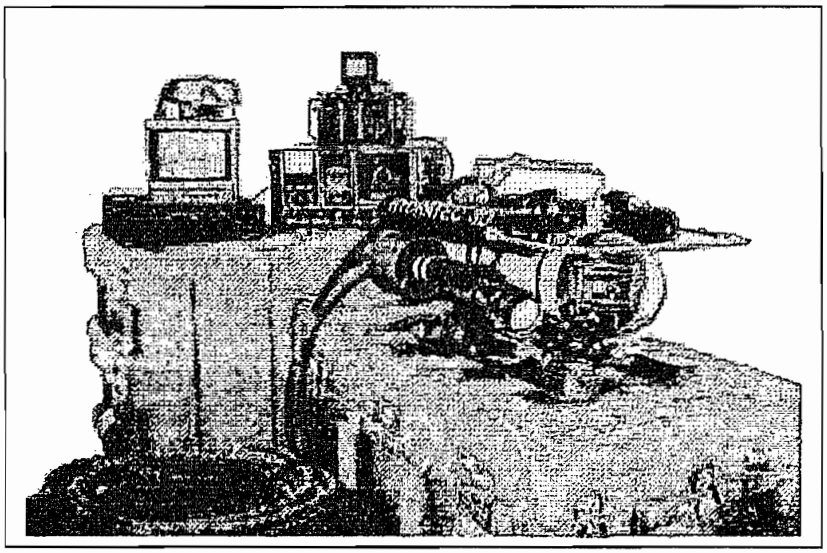
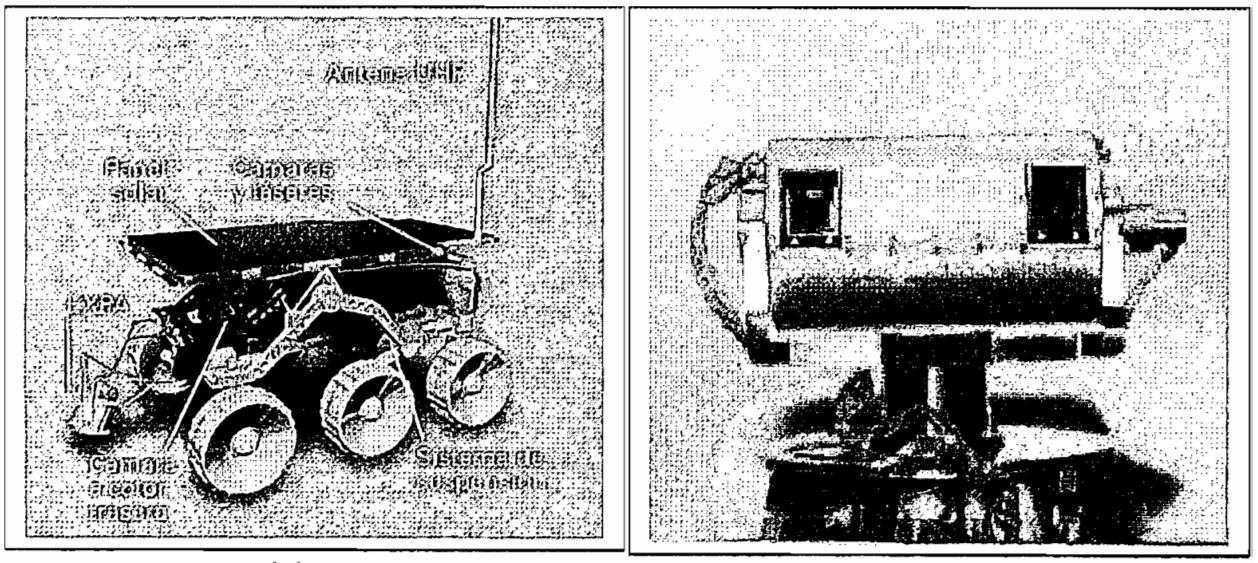


Figura 2.12 Mini-Rov HYDRATEC 3D, de la compañía Hydratec Tecnologia Submarina Ltda.

2.2.3. ESTUDIO DE LA TIERRA Y OTROS PLANETAS

Al igual que con la fotografía aérea, es posible obtener imágenes estereoscópicas de la Tierra, así como de otros planetas donde se pretende llegar

con algún tipo de robot para investigación y toma de muestras, haciéndose indispensable tener algún tipo de conocimiento del relieve que se quiere investigar, un ejemplo de esto es el esfuerzo realizado por la NASA, mediante la utilización de la sonda Pathfinder, para conocer mas acerca de la superficie de Marte. La toma de imágenes en estereo no solo sirvió para ver la superficie de Marte en 3D, sino para calcular distancias y tamaños de las rocas y conducir con más seguridad el vehículo, que de antemano se conoce debe ser operado con la ayuda de una cámara estereoscópica. La figura 2.13 muestra una imagen del vehículo utilizado en la exploración de Marte así como la cámara estereoscópica de filtros múltiples colocada en el pathfinder.



(a) (b)

Figura 2.13 (a) Sojourner, utilizado para explorar la superficie de Marte, (b) Cámara estereoscópica de filtros múltiples.

2.2.4. DISEÑO ASISTIDO POR COMPUTADOR (CAD) E INGENIERÍA ASISTIDA POR COMPUTADOR (CAE)

La utilización de la técnica estereoscópica es una poderosa herramienta para realizar cálculos y análisis de ingeniería, así como diseño y visualización de prototipos tanto en el campo de la Ingeniería Civil, industria automovilística, aeronáutica, etc. Consiguiendo con esto un importante ahorro en tiempo y dinero durante el desarrollo de dichos prototipos, los cuales no serian posible hacerlos

por métodos tradicionales; consiguiendo con la técnica estereoscópica optimizar piezas y conjuntos mecánicos, estructuras en obras civiles, etc.

2.2.5. INGENIERÍA MOLECULAR

En el campo de la Ingeniería Molecular, se ha hecho importante una visualización estereoscópica en las estaciones de diseño para poder apreciar sistemas moleculares complejos, es así como se han creado microscopios electrónicos con capacidad de entregar imágenes estereoscópicas. En la figura 2.14 se muestra un ejemplo de este equipo.

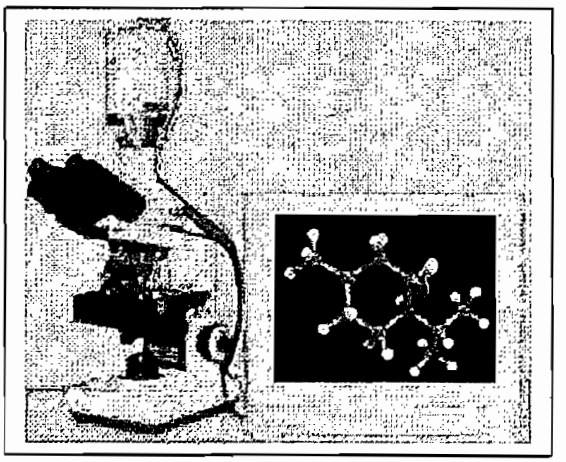


Figura 2.14 Microscopio estereoscópico electrónico, desplegando en pantalla el sistema molecular del menthol.

2.2.6. TELEPRESENCIA

En la telepresencia es de vital importancia la percepción de imágenes estereoscópicas ya que al tratarse de presencia a distancia y debido a que nuestro sistema visual de percepción es estereoscópico, se lo utiliza mucho para realizar trabajos en ambientes hostiles o de peligro, en donde se necesitan sistemas de video estero para una correcta teleoperación de los robots que generan la acción deseada, así como en sistemas de Telecomunicaciones. Un claro ejemplo de esto se encuentra en:

- ***la minería,*** donde debido al peligro que acarrea el excavar en las profundidades de la tierra, se utiliza control robótico remoto para todas las fases de operación como son exploración, colocación de cargas explosivas, descomposición de rocas y transporte del material buscado hacia la superficie, teniendo en cada fase una gran importancia la percepción por parte del operador del ambiente que explora el robot.

- ***manipulación radiactiva,*** este es uno de los principales campos en donde es indispensable el uso de la estereoscopia para una buena apreciación de la profundidad de los materiales y elementos radiactivos que son manejados en plantas nucleares, ya que éstos producen desechos que deben ser almacenados en contenedores especiales mientras decae su peligrosidad radiactiva. Es así como la estereo visión es considerada como esencial para proveer al teleoperador la habilidad de realizar operaciones en una manera muy diestra con un reducido riesgo de accidentes.

- ***Videoconferencia,*** es una de las aplicaciones de telecomunicaciones que permite que varias personas participen de una conferencia sin estar en un mismo lugar, donde la aplicación de la técnica visual estereoscópica permite apreciar de una manera mas real y vivida tanto a los conferencistas como al ambiente en que se desenvuelven.

Otro sistema de telecomunicaciones que está utilizando la técnica estereoscópica es la videotelefonía, que consiste en que las personas que intervienen en una conversación telefónica puedan visualizarse con sensación de profundidad.

2.2.7 REALIDAD VIRTUAL

Se entiende por realidad virtual a la interacción usuario-computador en la que imágenes estereoscópicas son generadas en tiempo real haciendo que el espectador se sienta inmerso en un escenario tridimensional artificial.

Esta aplicación surgió como un sistema de entretenimiento muy utilizado en video juegos, pero hoy se lo utiliza también en la arquitectura, la arqueología, medicina, industria automovilística y aéreo espacial. Siendo notorio la importancia del cálculo de los parámetros de visión estereoscópica ya que de ellos depende mucho el realismo del entorno virtual en todas estas aplicaciones.

Cabe anotar que las pantallas en las que se despliegan las imágenes virtuales pueden tener configuraciones de escritorio, paredes planas, paredes curvas y cuartos de 3 o 6 lados, la figura 2.15 muestra algunas de estas posibilidades.

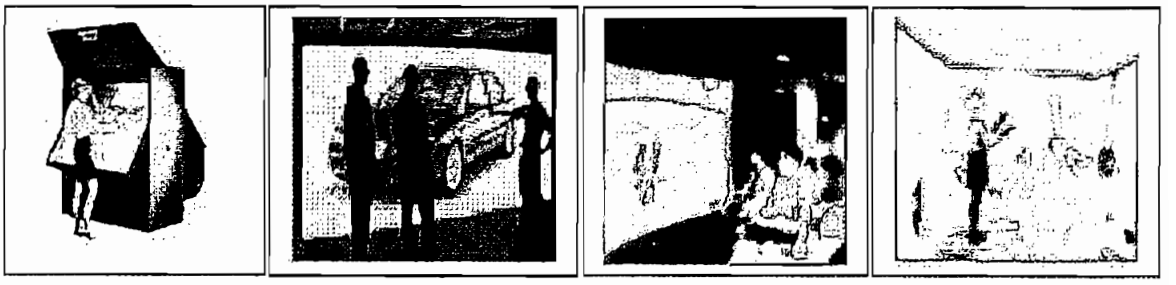


Figura 2.15 Configuraciones de pantallas auto estereoscópicas para realidad virtual.

CAPÌTULO III

3. PROCESOS UTILIZADOS EN LA GENERACIÓN DE SEÑAL DE TELEVISIÓN ESTEREOSCÓPICA

Para la generación de señales visuales estereoscópicas en general se deben seguir los siguientes pasos:

- Captación de la imagen.
- Compresión de las señales digitales estereoscópicas.
- Despliegue de las imágenes.

3.1 CAPTACIÓN DE LA IMAGEN

En la captación de imágenes estereoscópicas se han seguido dos tendencias tecnológicas distintas:

- Captación de la escena mediante 2 cámaras, lo que da origen a dos imágenes correspondientes a dos puntos distintos de visión.
- La captación de una escena con una única cámara, que *posee un* sistema óptico especial que permite tener dos imágenes simultaneas distintas: derecha e izquierda.

3.1.1 CAPTACIÓN DE LA ESCENA MEDIANTE DOS CÁMARAS

Dentro de la captación de la escena mediante dos cámaras, se puede tener dos tipos de configuración: Dos cámaras que están separadas una distancia interocular y cámaras separadas una distancia mayor que la distancia interocular.

En la primera configuración las dos cámaras necesitan guardar una alineación dentro de los límites permisibles en todos los tres ejes, controlando de esta manera el zoom, enfoque, la distancia interaxial y el punto de convergencia de los dos ejes ópticos, obteniéndose así un par estereoscópico que provee una información correcta de profundidad relativa.

En la segunda configuración las cámaras están separadas una distancia mucho mayor que la distancia interocular pero alineadas en los tres ejes (horizontal, vertical y profundidad), siendo necesario una síntesis de las dos imágenes obtenidas para de esta manera conseguir el par estereoscópico. Este sistema tiene la ventaja respecto al anterior que permite la simulación de movimiento de un par de cámaras virtuales en posiciones intermedias entre las dos cámaras reales, permitiéndole al observador experimentar la sensación de un paralaje en movimiento sin que las dos cámaras se hayan movido de su posición, este método es muy utilizado en video conferencias.

3.1.1.1 Geometría de la imagen estereoscópica

Las posiciones relativas y las orientaciones de los dos elementos de imagen en los dos planos sensores en un arreglo estereoscópico, constituyen la geometría de la imagen estereoscópica. Un arreglo de la imagen estereoscópica es mostrado en la figura 3.1. Un punto P en la escena 3D es proyectada en perspectiva hacia los puntos P_L y P_R en los sensores de la imagen izquierda y derecha, a través de los elementos de imagen izquierda y derecha L y R respectivamente (similar al tamaño de un agujero de alfiler[11] para lentes reales).

La disparidad del punto P (la distancia entre los puntos correspondientes P_L y P_R cuando las dos imágenes están alineadas una encima de la otra), es inversamente proporcional a la distancia de los centros de proyección. El problema de encontrar todos los pares dadas las vistas de las imágenes izquierda y derecha, es conocido como el problema de *correspondencia* o *estimación de disparidad.* La búsqueda de P_R para un P_L dado, es en general bidimensional. Sin embargo, cuando los ejes ópticos (los cuales son líneas perpendiculares a los planos de imagen pasando a través de los respectivos centros de proyección) son

[11] **Agujero de alfiler** se entiende como el modelo para un elemento de imagen que es infinitesimalmente pequeño, donde la imagen de un punto P en el mundo real, esta dada por la intersección del plano de imagen y la línea que une P y el agujero de alfiler.

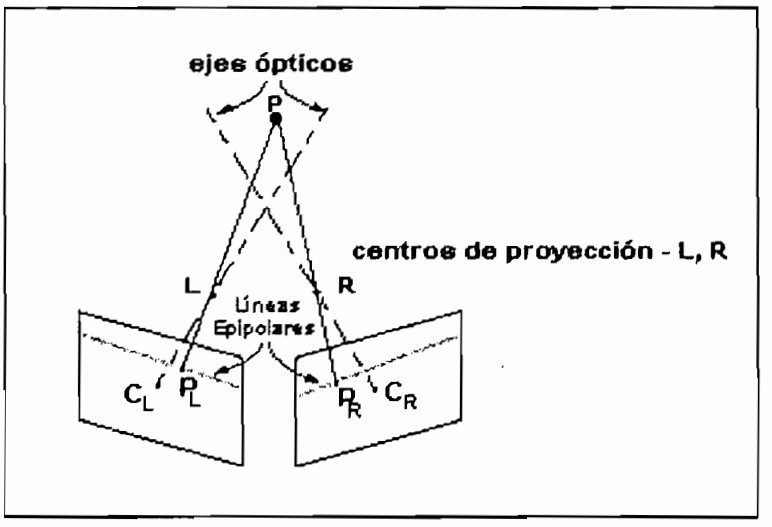


Figura 3.1 Geometría general de la imagen binocular

coplanares, los puntos correspondientes son forzados por la geometría para quedar delante de las líneas epipolares[12], definidas por las respectivas intersecciones de las dos imágenes planas con el plano definido por P, L y R. Así la búsqueda por el punto correspondiente P_L en la imagen izquierda para el punto P_R, es restringido a una dimensión. En el caso particular de que los ejes ópticos sean paralelos (Figura 3.2), las líneas epipolares llegan a ser líneas de exploración horizontal correspondientes. En este caso no hay necesidad de calcular la línea epipolar.

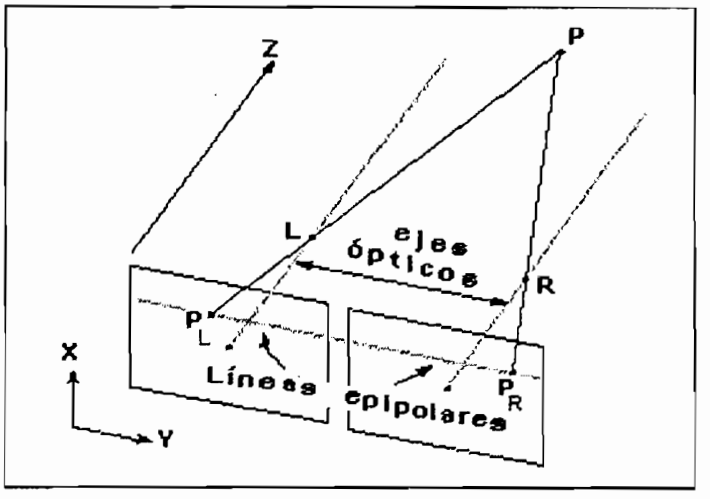


Figura 3.2 Geometría de imagen binocular con ejes paralelos

[12] **Línea epipolar** es la intersección del plano epipolar con los planos sensores de imagen, donde el **plano epipolar** es definido por los dos centros de proyección con el punto P.

Debido a la presencia de oclusiones (áreas que son visibles en una vista y no en la otra) no todos los puntos de la imagen tienen una correspondencia. La geometría apropiada de la imagen estereoscópica para observaciones estéreo esta estrechamente relacionada con la geometría del display estereoscópico, el cual involucra la posición de las pantallas del display izquierdo y derecho con respecto al observador y los ángulos de visión de la pantalla de display subtendido en los respectivos ojos.

Además de ser un arreglo favorable para el cálculo, la geometría de ejes paralelos es conocida por ser la *correcta geometría* para vistas estereoscópicas cuando las imágenes son mostradas en pantallas de display coplanar, esto se debe a que las dos vistas no tienen ninguna disparidad vertical entre los puntos correspondientes lo cual corrige la fatiga de ojos. Cuando la misma pantalla es usada para desplegar ambas vistas, la geometría pone restricciones adicionales en como posicionar los planos sensores de imagen con relación a los lentes.

3.1.1.2 Estereoscopia de múltiples vistas y síntesis de vistas intermedias.

Un par estereoscópico de imagen provee información de profundidad relativa de lo observado solamente desde un par de puntos de vista. De esta manera existe solamente una posición correcta de visión. Así, dos vistas son ajustadas por un observador en una sola ubicación.

Para permitir que múltiples observadores vean la perspectiva correcta y para proveer a un solo observador con señales de paralaje en movimiento durante el movimiento de su cabeza, se requieren mas de dos vistas. Puesto que esto puede ser prohibitivo en términos de adquisición, procesamiento y transmisión de vistas continuas, es preferible adquirir un mínimo juego de vistas para usar el conocimiento de la posición relativa de las cámaras y una estimación de un mapa de disparidad para sintetizar las vistas en medio de dos cámaras reales.
Así, la síntesis de vistas intermedias puede ser considerada como una forma de compresión. Sin embargo, la calidad de las vistas sintetizadas depende de la

exactitud de la estimación del mapa de disparidad y de la manera en que se manejan las oclusiones.

La estimación de disparidad se hace más confiable con un número creciente de vistas usando un arreglo básico de múltiples cámaras alineadas. Por lo general un juego de cámaras alineadas con distancias iguales entre ellas, como se muestra en la figura 3.3, se usan para adquirir múltiples vistas.

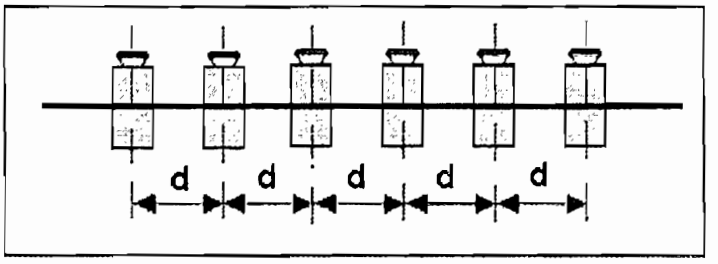


Figura 3.3 Configuración de múltiples cámaras equidistantes.

3.1.2 Captación de la escena mediante una cámara

Para captar secuencias de video estereoscópico con una cámara es necesario colocar en frente del lente de la cámara un adaptador óptico, el cual proporcionará un *campo secuencial de video 3D*. El adaptador estereoscópico consiste de una caja plástica hermética y resistente, un espejo reflector y un par de obturadores de cristal líquido (LCS). El haz de luz que pasa por la hendidura llega a las superficies polarizadas posicionadas ortogonalmente de los LCS's que abren y cierran las ventanas de luz para grabar tanto la imagen directa como la imagen reflejada en el espejo, en campos alternados de video. Como resultado la imagen izquierda es grabada durante el campo impar y la imagen derecha durante el campo par, o viceversa. Como se muestra en la figura 3.4(a) la sincronización de las ventanas de luz con los campos alternados de la cámara se consiguen mediante la conexión de un cable de video de salida de la cámara y el conector del adaptador.

En el gráfico 3.4(b), se puede apreciar como el adaptador produce un campo secuencial entrelazado de video estereoscópico mediante la grabación simultanea

de la vista del segundo ojo en la cámara. El campo secuencial resultante puede ser desplegado en monitores 2D (TV) o pantallas 2D con gafas estéreo especiales. El formato de campo secuencial entrelazado, sin embargo es un formato no conveniente para usarlo en varias aplicaciones de visión.

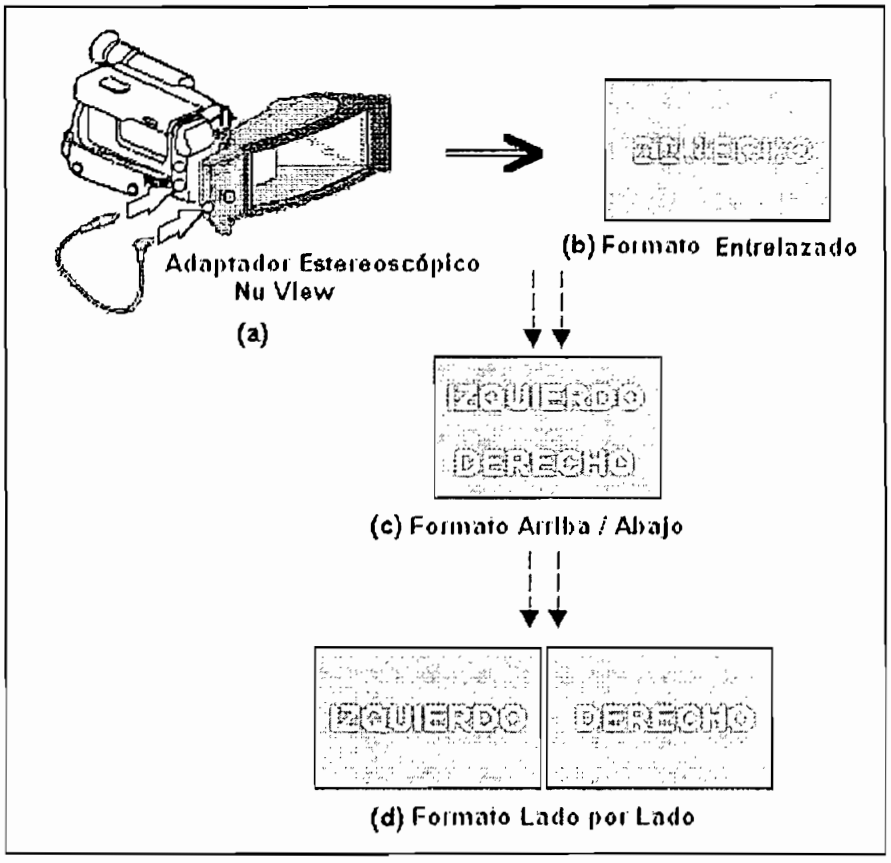


Figura 3.4 Captura de secuencias de video estéreo usando una cámara con adaptador estereoscópico.

Por ejemplo, aplicaciones de procesamiento, tales como filtración o transformación del campo secuencial de video que puede causar una pérdida en la calidad de imágenes estereoscópicas debido a los efectos del procesamiento de propagación de líneas interlazadas. Por la misma razón, el esquema de compresión de video no puede ser utilizado para grabar en espacio de disco duro o transmisión en canal de ancho de banda limitado. Por lo tanto, primero se separa el formato de campo secuencial entrelazado a un formato arriba/abajo (figura 3.4(c)), donde la parte izquierda es puesta en la parte superior de la imagen y la imagen derecha es puesta en la parte inferior, o viceversa.

Después de la separación del campo se transforma la imagen a un formato lado por lado (figura 3.4(d)). Necesitando ahora el desarrollo temporal o interpolación espacial de cada imagen para proveer una alta calidad de secuencias de imágenes de video 2D/3D. Este adaptador utiliza una frecuencia de 60 Hz para evitar los efectos de parpadeo (flicker). El video estereoscópico en 60 Hz no es tan uniforme comparado al video 2D en 60 Hz, porque el monitor 2D coloca 30 Hz para la imagen izquierda y 30 Hz para la imagen derecha.

Adicionalmente los displays (como los displays montados en la cabeza, pantallas polarizadas o displays autoestereoscópicos) requieren proyectar una imagen en tamaño original para proveer un confortable despliegue tridimensional. La interpolación espacial es también requerida en aplicaciones 2D solamente en la explotación de información de profundidad 3D. La interpolación espacial se logra por la copia de línea, duplicando el tamaño, o la interpolación lineal entre líneas, así tenemos que:

$$\begin{cases} F_L^{2i} = G_L^i \\ F_L^{2i+1} = \left(G_L^i + G_L^{i+1}\right)/2 \end{cases} \qquad \text{(Ec. 3.1)}$$

donde F_L y G_L denotan las imágenes de la figura 3.4 en lado a lado y arriba/abajo respectivamente. El exponente i representa el exponente de la fila en la imagen. La imagen derecha puede ser interpolada en una manera similar.

3.2 COMPRESIÓN DE LAS SEÑALES DIGITALES ESTEREOSCÓPICAS

3.2.1 NECESIDAD DE COMPRESIÓN DE VIDEO DIGITAL

Una señal nominal de video NTSC tiene 480 líneas activas[13] de barrido por cuadro donde cada línea de barrido contiene una componente de luminancia digitalizada por 720 muestras y cada componente de diferencia de color de 360 muestras, de acuerdo al muestreo 4:2:2 de las componentes Y , Cr y Cb se obtienen 1440 palabras por línea, con una composición de 8 bits por componente de píxel, con lo que una señal NTSC puede requerir cerca de 166 Mbps para una velocidad de 30 cuadros por segundo. Esto presenta un serio problema en lo que se refiere a transmisión y almacenamiento, ya que para transmisión el ancho de banda asignado a un canal de TV es de 6 MHz y para transmitir una señal digitalizada ahora sería necesario un ancho de banda de 83 MHz, asumiendo un esquema de modulación digital de 2 bits/Hz. De la misma forma para almacenar 60 segundos de video con formato NTSC sería necesario una capacidad aproximada de 1 Gigabyte. Estos cálculos demuestran que para transmitir y almacenar video es necesario comprimir las señales digitalizadas.

3.2.2 FACTORES QUE FACILITAN LA COMPRESIÓN.

La compresión de video digital se basa en principios de teoría de información y en modelos psicofísicos del sistema visual humano. Determinándose que se puede eliminar la información que presente una redundancia estadística o una redundancia perceptiva, para de esta manera obtener solo la información útil de la señal denominada **entropía**.

La redundancia estadística se presenta cuando existe una *redundancia de código* o *una redundancia de píxeles*. El código de una imagen representa el cuerpo de la información mediante un conjunto de símbolos. La eliminación del

[13] **Líneas activas:** las líneas que son vistas en la pantalla, del total de 525 líneas para NTSC solo 480 son activas.

por métodos tradicionales; consiguiendo con la técnica estereoscópica optimizar piezas y conjuntos mecánicos, estructuras en obras civiles, etc.

2.2.5. INGENIERÍA MOLECULAR

En el campo de la Ingeniería Molecular, se ha hecho importante una visualización estereoscópica en las estaciones de diseño para poder apreciar sistemas moleculares complejos, es así como se han creado microscopios electrónicos con capacidad de entregar imágenes estereoscópicas. En la figura 2.14 se muestra un ejemplo de este equipo.

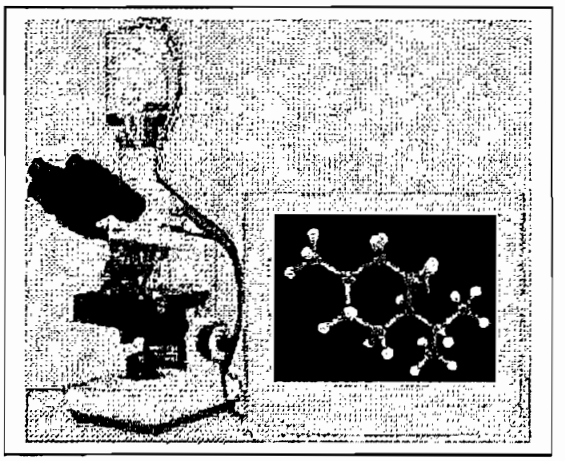


Figura 2.14 Microscopio estereoscópico electrónico, desplegando en pantalla el sistema molecular del menthol.

2.2.6. TELEPRESENCIA

En la telepresencia es de vital importancia la percepción de imágenes estereoscópicas ya que al tratarse de presencia a distancia y debido a que nuestro sistema visual de percepción es estereoscópico, se lo utiliza mucho para realizar trabajos en ambientes hostiles o de peligro, en donde se necesitan sistemas de video estero para una correcta teleoperación de los robots que generan la acción deseada, así como en sistemas de Telecomunicaciones. Un claro ejemplo de esto se encuentra en:

- ***la minería,*** donde debido al peligro que acarrea el excavar en las profundidades de la tierra, se utiliza control robótico remoto para todas las fases de operación como son exploración, colocación de cargas explosivas, descomposición de rocas y transporte del material buscado hacia la superficie, teniendo en cada fase una gran importancia la percepción por parte del operador del ambiente que explora el robot.

- ***manipulación radiactiva,*** este es uno de los principales campos en donde es indispensable el uso de la estereoscopia para una buena apreciación de la profundidad de los materiales y elementos radiactivos que son manejados en plantas nucleares, ya que éstos producen desechos que deben ser almacenados en contenedores especiales mientras decae su peligrosidad radiactiva. Es así como la estereo visión es considerada como esencial para proveer al teleoperador la habilidad de realizar operaciones en una manera muy diestra con un reducido riesgo de accidentes.

- ***Videoconferencia,*** es una de las aplicaciones de telecomunicaciones que permite que varias personas participen de una conferencia sin estar en un mismo lugar, donde la aplicación de la técnica visual estereoscópica permite apreciar de una manera mas real y vivida tanto a los conferencistas como al ambiente en que se desenvuelven.

Otro sistema de telecomunicaciones que está utilizando la técnica estereoscópica es la videotelefonía, que consiste en que las personas que intervienen en una conversación telefónica puedan visualizarse con sensación de profundidad.

2.2.7 REALIDAD VIRTUAL

Se entiende por realidad virtual a la interacción usuario-computador en la que imágenes estereoscópicas son generadas en tiempo real haciendo que el espectador se sienta inmerso en un escenario tridimensional artificial.

Esta aplicación surgió como un sistema de entretenimiento muy utilizado en video juegos, pero hoy se lo utiliza también en la arquitectura, la arqueología, medicina, industria automovilística y aéreo espacial. Siendo notorio la importancia del cálculo de los parámetros de visión estereoscópica ya que de ellos depende mucho el realismo del entorno virtual en todas estas aplicaciones.

Cabe anotar que las pantallas en las que se despliegan las imágenes virtuales pueden tener configuraciones de escritorio, paredes planas, paredes curvas y cuartos de 3 o 6 lados, la figura 2.15 muestra algunas de estas posibilidades.

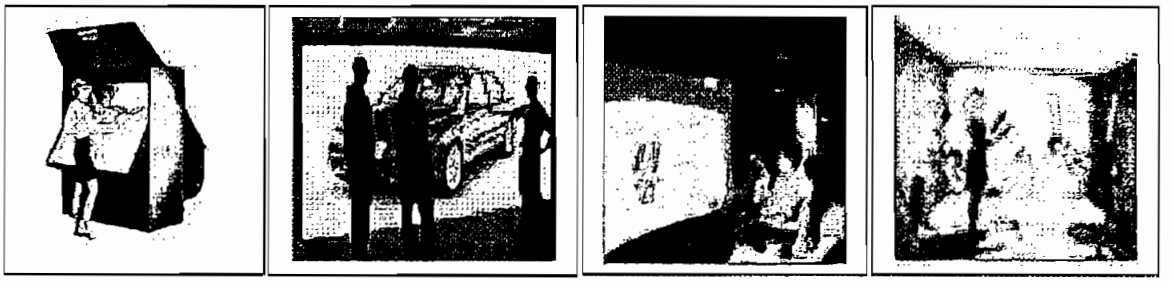


Figura 2.15 Configuraciones de pantallas auto estereoscópicas para realidad virtual.

CAPÌTULO III

3. PROCESOS UTILIZADOS EN LA GENERACIÓN DE SEÑAL DE TELEVISIÓN ESTEREOSCÓPICA

Para la generación de señales visuales estereoscópicas en general se deben seguir los siguientes pasos:

- Captación de la imagen.
- Compresión de las señales digitales estereoscópicas.
- Despliegue de las imágenes.

3.1 CAPTACIÓN DE LA IMAGEN

En la captación de imágenes estereoscópicas se han seguido dos tendencias tecnológicas distintas:

- Captación de la escena mediante 2 cámaras, lo que da origen a dos imágenes correspondientes a dos puntos distintos de visión.
- La captación de una escena con una única cámara, que posee un sistema óptico especial que permite tener dos imágenes simultaneas distintas: derecha e izquierda.

3.1.1 CAPTACIÓN DE LA ESCENA MEDIANTE DOS CÁMARAS

Dentro de la captación de la escena mediante dos cámaras, se puede tener dos tipos de configuración: Dos cámaras que están separadas una distancia interocular y cámaras separadas una distancia mayor que la distancia interocular.

En la primera configuración las dos cámaras necesitan guardar una alineación dentro de los límites permisibles en todos los tres ejes, controlando de esta manera el zoom, enfoque, la distancia interaxial y el punto de convergencia de los dos ejes ópticos, obteniéndose así un par estereoscópico que provee una información correcta de profundidad relativa.

En la segunda configuración las cámaras están separadas una distancia mucho mayor que la distancia interocular pero alineadas en los tres ejes (horizontal, vertical y profundidad), siendo necesario una síntesis de las dos imágenes obtenidas para de esta manera conseguir el par estereoscópico. Este sistema tiene la ventaja respecto al anterior que permite la simulación de movimiento de un par de cámaras virtuales en posiciones intermedias entre las dos cámaras reales, permitiéndole al observador experimentar la sensación de un paralaje en movimiento sin que las dos cámaras se hayan movido de su posición, este método es muy utilizado en video conferencias.

3.1.1.1 Geometría de la imagen estereoscópica

Las posiciones relativas y las orientaciones de los dos elementos de imagen en los dos planos sensores en un arreglo estereoscópico, constituyen la geometría de la imagen estereoscópica. Un arreglo de la imagen estereoscópica es mostrado en la figura 3.1. Un punto P en la escena 3D es proyectada en perspectiva hacia los puntos P_L y P_R en los sensores de la imagen izquierda y derecha, a través de los elementos de imagen izquierda y derecha L y R respectivamente (similar al tamaño de un agujero de alfiler[11] para lentes reales).

La disparidad del punto P (la distancia entre los puntos correspondientes P_L y P_R cuando las dos imágenes están alineadas una encima de la otra), es inversamente proporcional a la distancia de los centros de proyección. El problema de encontrar todos los pares dadas las vistas de las imágenes izquierda y derecha, es conocido como el problema de *correspondencia* o *estimación de disparidad*. La búsqueda de P_R para un P_L dado, es en general bidimensional. Sin embargo, cuando los ejes ópticos (los cuales son líneas perpendiculares a los planos de imagen pasando a través de los respectivos centros de proyección) son

[11] **Agujero de alfiler** se entiende como el modelo para un elemento de imagen que es infinitesimalmente pequeño, donde la imagen de un punto P en el mundo real, esta dada por la intersección del plano de imagen y la línea que une P y el agujero de alfiler.

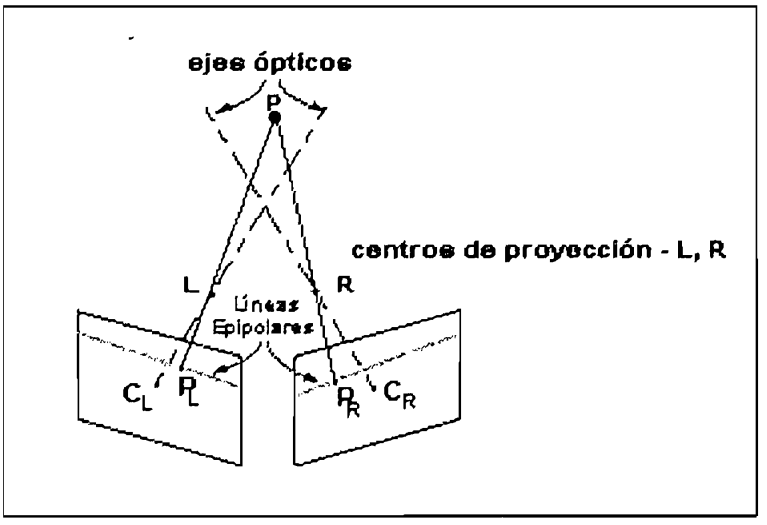


Figura 3.1 Geometría general de la imagen binocular

coplanares, los puntos correspondientes son forzados por la geometría para quedar delante de las líneas epipolares[12], definidas por las respectivas intersecciones de las dos imágenes planas con el plano definido por P, L y R. Así la búsqueda por el punto correspondiente P_L en la imagen izquierda para el punto P_R, es restringido a una dimensión. En el caso particular de que los ejes ópticos sean paralelos (Figura 3.2), las líneas epipolares llegan a ser líneas de exploración horizontal correspondientes. En este caso no hay necesidad de calcular la línea epipolar.

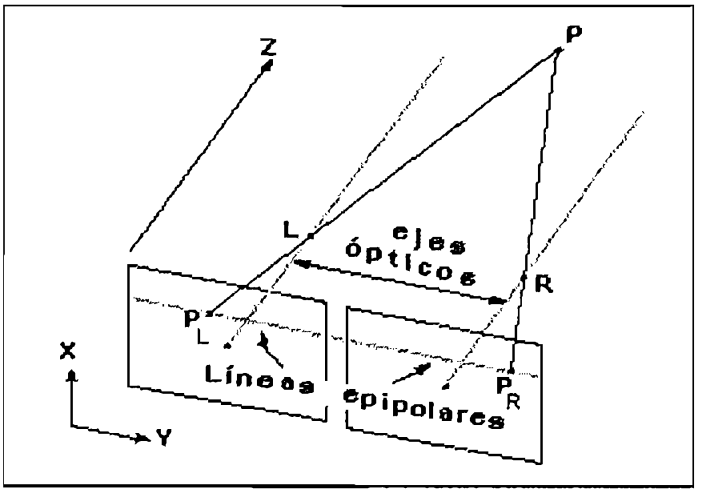


Figura 3.2 Geometría de imagen binocular con ejes paralelos

[12] **Línea epipolar** es la intersección del plano epipolar con los planos sensores de imagen, donde el **plano epipolar** es definido por los dos centros de proyección con el punto P.

Debido a la presencia de oclusiones (áreas que son visibles en una vista y no en la otra) no todos los puntos de la imagen tienen una correspondencia. La geometría apropiada de la imagen estereoscópica para observaciones estéreo esta estrechamente relacionada con la geometría del display estereoscópico, el cual involucra la posición de las pantallas del display izquierdo y derecho con respecto al observador y los ángulos de visión de la pantalla de display subtendido en los respectivos ojos.

Además de ser un arreglo favorable para el cálculo, la geometría de ejes paralelos es conocida por ser la *correcta geometría* para vistas estereoscópicas cuando las imágenes son mostradas en pantallas de display coplanar, esto se debe a que las dos vistas no tienen ninguna disparidad vertical entre los puntos correspondientes lo cual corrige la fatiga de ojos. Cuando la misma pantalla es usada para desplegar ambas vistas, la geometría pone restricciones adicionales en como posicionar los planos sensores de imagen con relación a los lentes.

3.1.1.2 Estereoscopia de múltiples vistas y síntesis de vistas intermedias.

Un par estereoscópico de imagen provee información de profundidad relativa de lo observado solamente desde un par de puntos de vista. De esta manera existe solamente una posición correcta de visión. Así, dos vistas son ajustadas por un observador en una sola ubicación.

Para permitir que múltiples observadores vean la perspectiva correcta y para proveer a un solo observador con señales de paralaje en movimiento durante el movimiento de su cabeza, se requieren mas de dos vistas. Puesto que esto puede ser prohibitivo en términos de adquisición, procesamiento y transmisión de vistas continuas, es preferible adquirir un mínimo juego de vistas para usar el conocimiento de la posición relativa de las cámaras y una estimación de un mapa de disparidad para sintetizar las vistas en medio de dos cámaras reales.
Así, la síntesis de vistas intermedias puede ser considerada como una forma de compresión. Sin embargo, la calidad de las vistas sintetizadas depende de la

exactitud de la estimación del mapa de disparidad y de la manera en que se manejan las oclusiones.

La estimación de disparidad se hace más confiable con un número creciente de vistas usando un arreglo básico de múltiples cámaras alineadas. Por lo general un juego de cámaras alineadas con distancias iguales entre ellas, como se muestra en la figura 3.3, se usan para adquirir múltiples vistas.

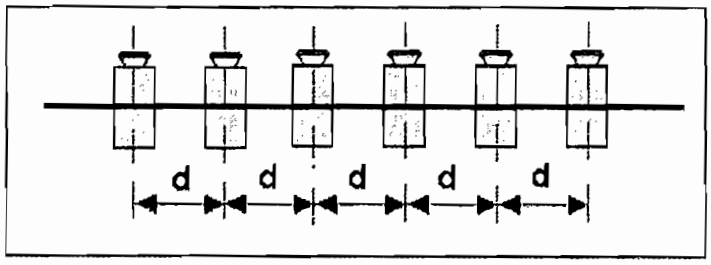


Figura 3.3 Configuración de múltiples cámaras equidistantes.

3.1.2 Captación de la escena mediante una cámara

Para captar secuencias de video estereoscópico con una cámara es necesario colocar en frente del lente de la cámara un adaptador óptico, el cual proporcionará un *campo secuencial de video 3D*. El adaptador estereoscópico consiste de una caja plástica hermética y resistente, un espejo reflector y un par de obturadores de cristal líquido (LCS). El haz de luz que pasa por la hendidura llega a las superficies polarizadas posicionadas ortogonalmente de los LCS's que abren y cierran las ventanas de luz para grabar tanto la imagen directa como la imagen reflejada en el espejo, en campos alternados de video. Como resultado la imagen izquierda es grabada durante el campo impar y la imagen derecha durante el campo par, o viceversa. Como se muestra en la figura 3.4(a) la sincronización de las ventanas de luz con los campos alternados de la cámara se consiguen mediante la conexión de un cable de video de salida de la cámara y el conector del adaptador.

En el gráfico 3.4(b), se puede apreciar como el adaptador produce un campo secuencial entrelazado de video estereoscópico mediante la grabación simultanea

de la vista del segundo ojo en la cámara. El campo secuencial resultante puede ser desplegado en monitores 2D (TV) o pantallas 2D con gafas estéreo especiales. El formato de campo secuencial entrelazado, sin embargo es un formato no conveniente para usarlo en varias aplicaciones de visión.

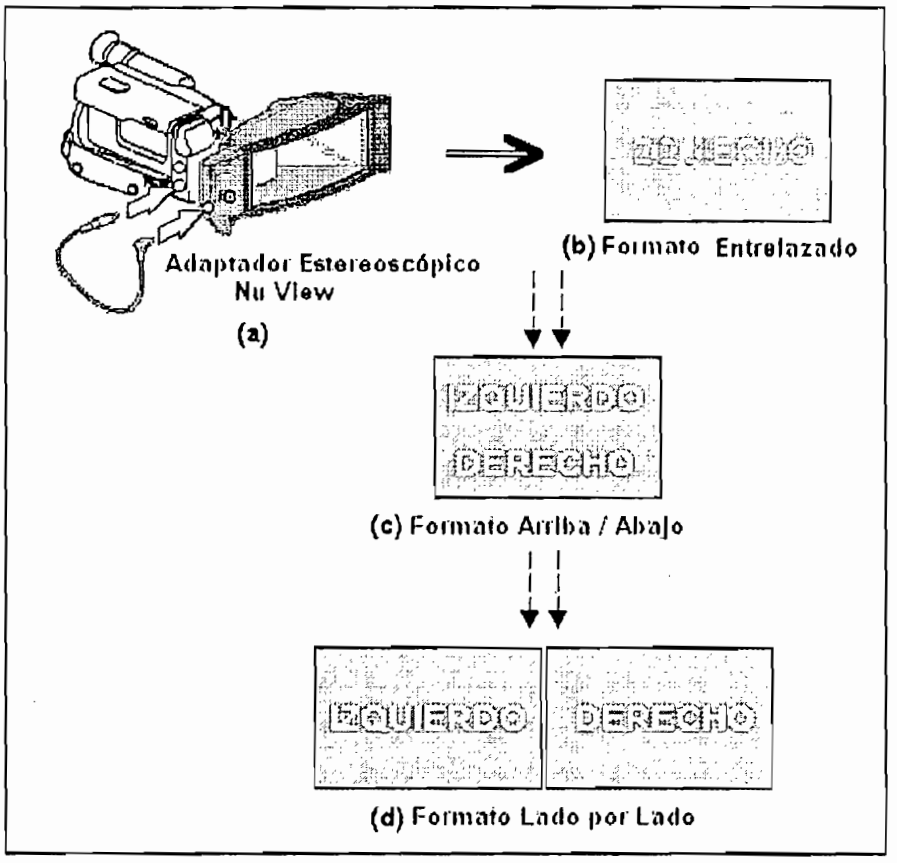


Figura 3.4 Captura de secuencias de video estéreo usando una cámara con adaptador estereoscópico.

Por ejemplo, aplicaciones de procesamiento, tales como filtración o transformación del campo secuencial de video que puede causar una pérdida en la calidad de imágenes estereoscópicas debido a los efectos del procesamiento de propagación de líneas interlazadas. Por la misma razón, el esquema de compresión de video no puede ser utilizado para grabar en espacio de disco duro o transmisión en canal de ancho de banda limitado. Por lo tanto, primero se separa el formato de campo secuencial entrelazado a un formato arriba/abajo (figura 3.4(c)), donde la parte izquierda es puesta en la parte superior de la imagen y la imagen derecha es puesta en la parte inferior, o viceversa.

Después de la separación del campo se transforma la imagen a un formato lado por lado (figura 3.4(d)). Necesitando ahora el desarrollo temporal o interpolación espacial de cada imagen para proveer una alta calidad de secuencias de imágenes de video 2D/3D. Este adaptador utiliza una frecuencia de 60 Hz para evitar los efectos de parpadeo (flicker). El video estereoscópico en 60 Hz no es tan uniforme comparado al video 2D en 60 Hz, porque el monitor 2D coloca 30 Hz para la imagen izquierda y 30 Hz para la imagen derecha.

Adicionalmente los displays (como los displays montados en la cabeza, pantallas polarizadas o displays autoestereoscópicos) requieren proyectar una imagen en tamaño original para proveer un confortable despliegue tridimensional. La interpolación espacial es también requerida en aplicaciones 2D solamente en la explotación de información de profundidad 3D. La interpolación espacial se logra por la copia de línea, duplicando el tamaño, o la interpolación lineal entre líneas, así tenemos que:

$$\begin{cases} F_L^{2i} = G_L^i \\ F_L^{2i+1} = \left(G_L^i + G_L^{i+1}\right)/2 \end{cases} \qquad \text{(Ec. 3.1)}$$

donde F_L y G_L denotan las imágenes de la figura 3.4 en lado a lado y arriba/abajo respectivamente. El exponente i representa el exponente de la fila en la imagen. La imagen derecha puede ser interpolada en una manera similar.

3.2 COMPRESIÓN DE LAS SEÑALES DIGITALES ESTEREOSCÓPICAS

3.2.1 NECESIDAD DE COMPRESIÓN DE VIDEO DIGITAL

Una señal nominal de video NTSC tiene 480 líneas activas[13] de barrido por cuadro donde cada línea de barrido contiene una componente de luminancia digitalizada por 720 muestras y cada componente de diferencia de color de 360 muestras, de acuerdo al muestreo 4:2:2 de las componentes Y , Cr y Cb se obtienen 1440 palabras por línea, con una composición de 8 bits por componente de píxel, con lo que una señal NTSC puede requerir cerca de 166 Mbps para una velocidad de 30 cuadros por segundo. Esto presenta un serio problema en lo que se refiere a transmisión y almacenamiento, ya que para transmisión el ancho de banda asignado a un canal de TV es de 6 MHz y para transmitir una señal digitalizada ahora sería necesario un ancho de banda de 83 MHz, asumiendo un esquema de modulación digital de 2 bits/Hz. De la misma forma para almacenar 60 segundos de video con formato NTSC sería necesario una capacidad aproximada de 1 Gigabyte. Estos cálculos demuestran que para transmitir y almacenar video es necesario comprimir las señales digitalizadas.

3.2.2 FACTORES QUE FACILITAN LA COMPRESIÓN.

La compresión de video digital se basa en principios de teoría de información y en modelos psicofísicos del sistema visual humano. Determinándose que se puede eliminar la información que presente una redundancia estadística o una redundancia perceptiva, para de esta manera obtener solo la información útil de la señal denominada **entropía**.

La redundancia estadística se presenta cuando existe una *redundancia de código* o *una redundancia de píxeles*. El código de una imagen representa el cuerpo de la información mediante un conjunto de símbolos. La eliminación del

[13] **Líneas activas:** las líneas que son vistas en la pantalla, del total de 525 líneas para NTSC solo 480 son activas.

código redundante consiste en utilizar el menor número de símbolos para representar la información. La redundancia de píxeles se presenta debido a que la mayoría de las imágenes presentan semejanzas o correlaciones entre sus píxeles. Estas correlaciones se deben a la existencia de estructuras similares en las imágenes, puesto que no son completamente aleatorias. De esta manera, el valor de un píxel puede emplearse para predecir el de sus vecinos.

Las técnicas de compresión que eliminan estas redundancias de código y de píxeles utilizan cálculos estadísticos para lograr eliminar este tipo de redundancia y reducir la ocupación original de los datos en espacio y tiempo, dando lugar a la compresión espacial y temporal.

De ahí que métodos de codificación que explotan solamente la redundancia espacial son llamados métodos de codificación ***intraframe*** (o simplemente intracoding), en donde se codifica basándose completamente en la redundancia propia de la imagen. Y los métodos que explotan solamente la redundancia temporal se denominan métodos de codificación ***interframe*** (o predictivo) donde la codificación se basa en la información repetitiva en tramas de video consecutivas. La eliminación de la redundancia estadística conduce a los métodos de compresión **lossless**, sin pérdida de la información y que alcanza factores de compresión[14] muy bajos, alrededor de 2:1.

La redundancia perceptiva también conocida como *redundancia visual*, es creada por el mecanismo de percepción del sistema visual humano (entre el ojo y el cerebro), consistiendo en la remoción de las irrelevancias perceptuales, ya que el ojo humano responde con diferente sensibilidad a la información visual que recibe, la información a la que es menos sensible se puede descartar sin afectar a la percepción de la imagen.

[14] **Factor de compresión:** también conocido como relación de compresión, es la relación entre el número de bits usados para representar una imagen o secuencia antes de la compresión y el numero de bits necesarios para representarla después de la compresión.

Cuando se elimina la redundancia perceptiva se obtiene la denominada compresión Lossy, con pérdida de información y que logra alcanzar unos factores de compresión más elevados (10:1, 50:1 o mayores), a costa de sufrir una pérdida de información sobre la imagen original.

3.2.3 MÉTODOS DE CODIFICACIÓN BASADOS EN LA FORMA DE ONDA

Estos métodos están basados principalmente en propiedades estadísticas de las intensidades de la imagen y no utilizan ninguna información derivada de objetos físicos que están presentes en la escena. Estos métodos son principalmente 2D (espacial) y 3D (espacial-temporal), siendo extensiones de métodos de codificación de forma de onda de señales 1D. Algunos métodos de codificación de forma de onda usados ampliamente son: *modulación diferencial por impulsos codificados* (DPCM), *codificación mediante transformadas*, *codificación en subbandas*, *cuantificación vectorial* (VQ) y *compresión mediante fractales*.

La técnica **DPCM** se basa en la eliminación de las redundancias entre píxeles muy próximos, extrayendo y codificando únicamente la nueva información que aporta cada píxel. Se define la nueva información de un píxel como la diferencia entre el valor real y el valor estimado de ese píxel.

Las Figuras 3.5 y 3.6 muestran los componentes básicos de un sistema de codificación predictiva sin pérdidas (lossless). El sistema consta de un codificador y un decodificador, conteniendo ambos un predictor idéntico.

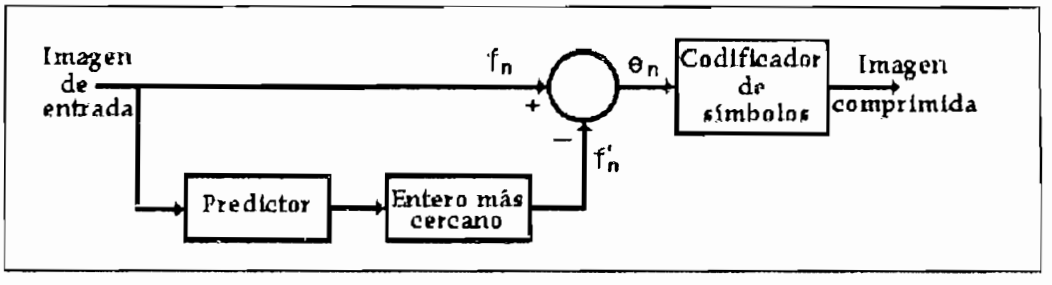


Figura 3.5 Codificador DPCM con técnica lossless

A medida que se va introduciendo sucesivamente cada píxel de la imagen de entrada, representado por f_n, en el codificador (Figura 3.5), el predictor genera el valor anticipado de dicho píxel en función de algún número de entradas anteriores. La salida del predictor se redondea después al entero más cercano, representado por f'_n, y se utiliza para construir la diferencia, o error de predicción como se muestra en la ecuación 3.2 :

$$e_n = f_n - f'_n \qquad \text{(Ec. 3.2)}$$

que se codifica utilizando un código de longitud variable (por medio de un codificador de símbolos) para generar el siguiente elemento del flujo de datos comprimidos. El decodificador de la Figura 3.6 reconstruye e_n a partir de las palabras código de longitud variable y realiza la operación inversa:

$$f_n = e_n + f'_n \qquad \text{(Ec. 3.3)}$$

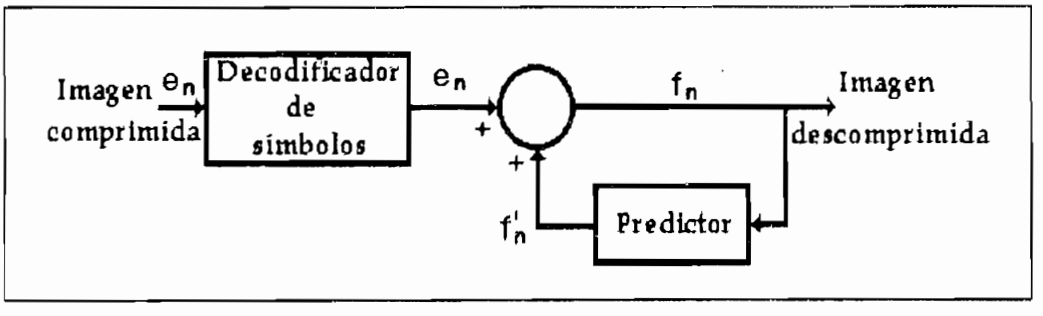


Figura 3.6 Decodificador DPCM con técnica lossless.

En la codificación predictiva de dos dimensiones, la predicción es una función de los píxeles anteriores de una exploración de izquierda a derecha y de arriba abajo de una imagen.

La estructura fundamental para la codificación predictiva de una imagen, es la modulación de pulsos codificados diferenciales (DPCM). Esto implica, que la cantidad que se codifica es la diferencia entre píxeles de brillo.

El esquema de compresión DPCM opera en la imagen completa, píxel por píxel. El primer píxel, en la esquina superior izquierda de la imagen, permanece inalterado; este es exactamente codificado con su brillo original. El proceso se mueve ahora al segundo píxel en la línea, donde el siguiente valor de brillo del píxel se sustrae de los actuales píxeles de brillo. El resultado de la sustracción es el nuevo valor codificado para el segundo píxel en la imagen. Este proceso se repite por toda la línea. Al inicio de la próxima línea, el proceso comienza de nuevo, y este continúa hasta que la imagen entera es codificada. Las operaciones de compresión y descompresión de la codificación predictiva sin pérdidas se muestran en la Figura 3.7.

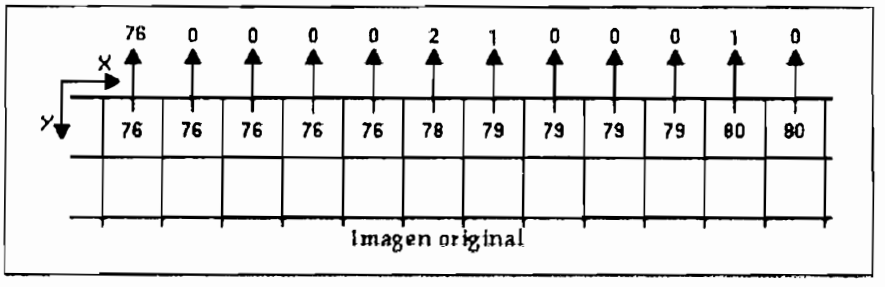


Figura 3.7 Operación de una codificación predictiva lossless.

Como ejemplo, se toman los cinco primeros píxeles de la línea de una imagen que contiene los siguientes valores de brillo: 23, 48, 76, 56, y 83. Se asumirá que la imagen fue originada con valores de brillo de 8 bits. Los valores DPCM codificados son mostrados en la tabla 3.1.

Imagen original	Valores de 8 bits	Código DPCM de 6 bits
Píxel # 1	23	23
Píxel # 2	48	48-23=25
Píxel # 3	76	76-48=28
Píxel # 4	56	56-76=-20
Píxel # 5	83	83-56=27
Total de bits	8x5=40 bits	6x5=30 bits

Tabla 3.1 Ejemplo de codificación DPCM con 6 bits.

Los cinco primeros píxeles de brillo son comprimidos desde valores de 5 x 8 bits = 40 bits a valores de la diferencia de 5 x 6 bits = 30 bits.

El esquema de compresión DPCM trabaja con la suposición de que los píxeles vecinos serán similares o altamente correlacionados. Como resultado, sus diferencias normalmente serán valores muy pequeños. Mirando los valores en el ejemplo anterior, ninguno es mayor de 31 o menor de -32, éstas diferencias de valor se pueden codificar usando números de 6 bits en lugar de números de 8 bits, permitiendo un factor de compresión de 8/6 = 1.333:1. Si todo los valores de las diferencias estuvieran debajo de 16, sólo serían necesarios números de 4 bits, permitiendo un factor de compresión de 2:1.

La operación de compresión en DPCM trabaja mejor en imágenes que no tienen un número desmesuradamente grande de brillo que oscila entre píxeles adyacentes. Cuando se aplica a imágenes normales, la codificación DPCM puede proporcionar factores de compresión alrededor de 2:1. Para las imágenes con series largas de valores de píxeles constantes, los factores de compresión se pueden incrementar significativamente.

En la ***codificación mediante transformadas***, se utiliza una transformada lineal reversible, para hacer corresponder una imagen con un conjunto de componentes fundamentales o coeficientes, en el dominio de la frecuencia. La imagen en el dominio de la frecuencia se puede transformar inversamente al dominio espacial, reproduciendo la imagen tal y como estaba originalmente. Este principio es el fundamento para las técnicas de compresión por transformadas. Los sistemas más prácticos de codificación por transformación se basan en la *Transformada de Coseno Discreta* (**D**iscrete **C**osine **T**ransform, **DCT**), que tiene un compromiso entre la capacidad para concentrar la información y la complejidad de cálculo.

La transformada de coseno discreto (DCT) presenta una mayor eficiencia para imágenes naturales típicas, este método presenta una mejor reducción de redundancias que los métodos DPCM. Un codificador DCT típico es mostrado en la figura 3.8.

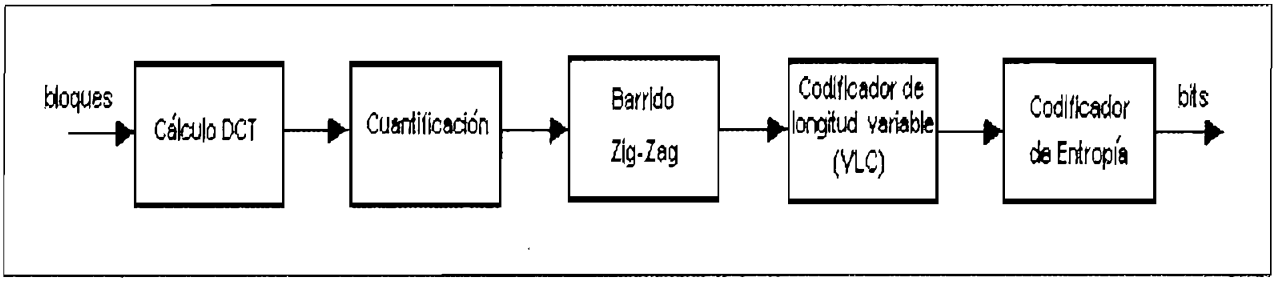


Figura 3.8 Codificador DCT típico

Para el cálculo de la DCT se divide la imagen en bloques de píxeles de tamaño 8x8 como se muestra en la Figura 3.9, que se procesan de izquierda a derecha y de arriba abajo. Según se va encontrando cada bloque o subimagen de 8x8, se cambian los niveles de sus 64 píxeles, sustrayendo de los mismos la cantidad 2^{n-1}, siendo 2^n, el máximo número de niveles de gris. Esto es, para las imágenes de 8 bits se resta 128 de cada píxel. Después se calcula la Transformada Discreta del Coseno bidimensional del bloque, con las siguientes fórmulas:

$$DCT \quad Svu = \frac{1}{4} CuCv \sum_{X=0}^{7} \sum_{Y=0}^{7} S_{yx} COS\left[\frac{(2x+1)u\Pi}{16}\right] COS\left[\frac{(2y+1)v\Pi}{16}\right] \qquad \text{(Ec. 3.4)}$$

$$DCT \quad Inversa \quad Syx = \frac{1}{4} CuCv \sum_{U=0}^{7} \sum_{V=0}^{7} SvuCOS\left[\frac{(2x+1)u\Pi}{16}\right] COS\left[\frac{(2y+1)v\Pi}{16}\right] \qquad \text{(Ec. 3.5)}$$

donde $Cu \quad y \quad Cv = \frac{1}{\sqrt{2}} \quad cuando u, v = 0,0$ componente DC

$Cu \quad y \quad Cv = 1$ en los demás casos

Svu = Celda designada para el coeficiente DCT

Syx = Celda designada para el píxel reconstruido

produciendo un conjunto de 64 valores conocidos como coeficientes de la DCT, como se ve en la Figura 3.10

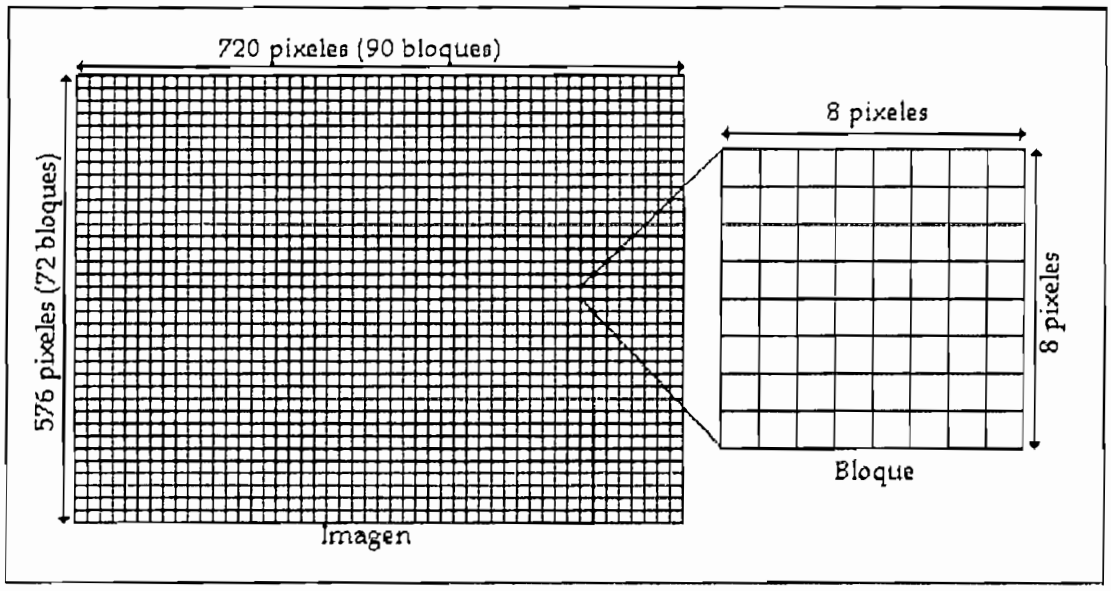


Figura 3.9 División en bloques o subimágenes de 8x8 píxeles

En el cuantificador los 64 coeficientes son entonces cuantificados, produciendo en algunos de ellos su reducción a cero. Los coeficientes son codificados en umbral, usando una matriz de cuantificación y son preparados para la codificación de entropía convirtiéndolos en una cadena unidimensional de 64 coeficientes en orden cuasi ascendente de los componentes de frecuencia. Para convertir los coeficientes en esta cadena unidimensional se reordenan usando una exploración o barrido en zig-zag.

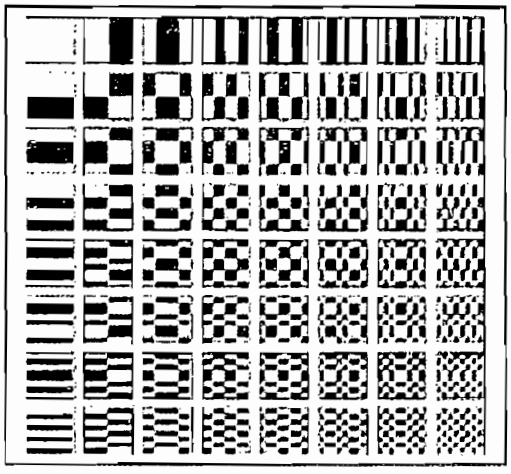


Figura 3.10 Coeficientes de un bloque de 8x8

El primer coeficiente del barrido en zig-zag es conocido como el coeficiente DC mientras que el resto son los coeficientes AC, esto se ilustra en la figura 3.11. A la matriz de cuantificación se le pueden aplicar factores de escala para obtener

diversos niveles de compresión. Las entradas de la matriz de cuantificación son usualmente determinadas según consideraciones psicovisuales.

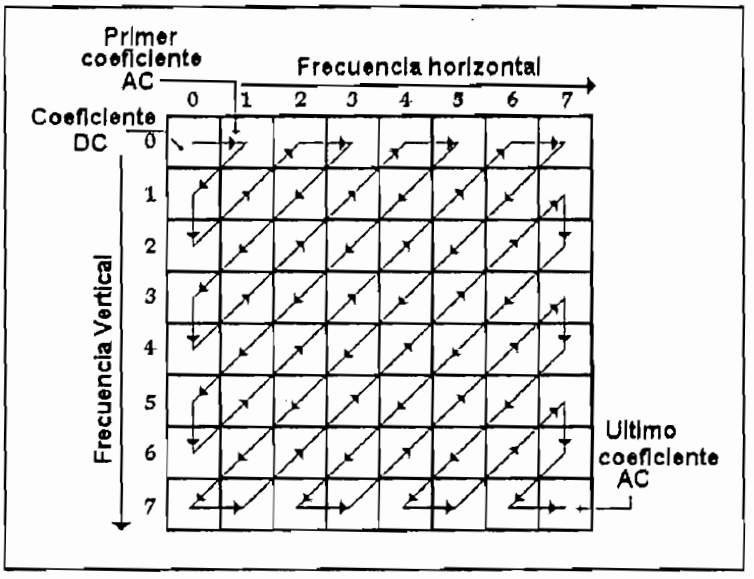


Figura 3.11 Barrido o exploración en zig-zag

En la asignación del Código de Longitud Variable (VLC) el coeficiente DC de cada bloque es codificado usando DPCM. Es decir, se codifica la diferencia entre coeficiente DC del presente bloque y el del bloque previamente codificado. Puesto que la cadena unidimensional reordenada según el barrido en zig-zag de la Figura 3.11 está distribuida cualitativamente según una frecuencia espacial creciente, los coeficientes AC no nulos se codifican utilizando un código de longitud variable que define el valor del coeficiente y el número de ceros precedentes.

Los métodos de ***codificación en subbandas*** son otra clase de métodos de codificación de formas de onda, que aprovechan la distribución no uniforme de la energía a través de diferentes bandas de frecuencia. Estos métodos dividen la imagen en dos bandas diferentes, cada una es codificada independientemente según un esquema óptimo de asignación de bit. De tal manera que la imagen entera se filtra y se submuestrea para obtener las subbandas, estos métodos no experimentan discontinuidades artificiales visibles a través de límites de bloques, esto es común en métodos de codificación de transformadas basados en bloques.

La codificación de subbandas ha mostrado ser equivalente a la codificación usando una extensión de bloque (no sobrepuesto) llamado transformada ortogonal solapada o sobrepuesta (**L**apped **O**rthogonal **T**ransform, **LOT**).

La codificación basada en ***vectores de cuantificación*** es una extensión de los principios de cuantificación escalar (**S**calar **Q**uantizer, **SQ**) para vencer la barrera de 1 bit por píxel (bpp) asociada con SQ. La idea principal de la cuantificación vectorial es particionar el espacio vectorial en sectores (figura 3.12), cada uno de los cuales será representado por un solo vector que puede ser el centroide.

El conjunto de centroides viene a ser el libro de códigos (codebook) que conforman los niveles de cuantificación y a cada uno se le asigna una dirección y etiqueta. Para efectuar la cuantificación de un vector de entrada lo que se realiza es asignarle la dirección del vector del libro de códigos más cercano evaluado mediante una medida de similitud. Un aspecto muy importante de cualquier sistema de cuantificación vectorial es la obtención del libro de códigos, el espacio vectorial debe ser dividido en sectores los cuales se hallan partiendo de vectores de entrenamiento. Dichos vectores deben representar fielmente el espacio de interés. El libro de códigos se obtiene empleando un algoritmo conocido como LBG (cuyo nombre se deriva de los creadores **L**inde, **B**uzo y **G**ray)[15].

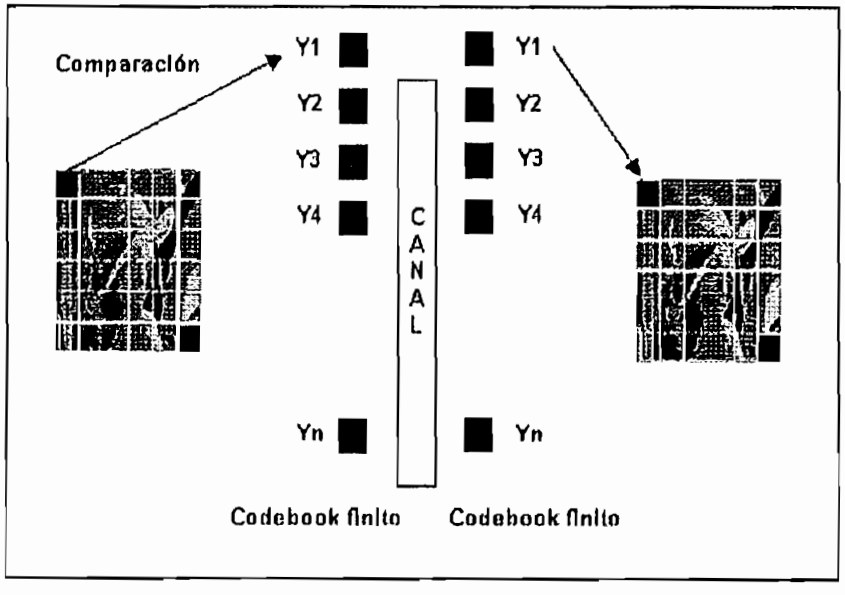


Figura 3.12 Cuantificación vectorial (VQ)

[15] **Referencia internet:** http://alek.pucp.edu.pe/~dflores/cuantif.html

Varios métodos computacionalmente eficientes (como es el árbol estructurado VQ) que reduce la complejidad de la búsqueda al encontrar el vector del código de mejor aproximación y varios sistemas con variantes han sido propuestos. El VQ puede ser usado por codificación de imagen directa, codificación residual, o codificación de subbandas.

La ***codificación de imagen fractal*** se basa en el método de M. F. Barnsley, que a partir de una imagen natural, obtiene una familia de contracciones que generan un fractal que se aproxima a la imagen natural tanto como queramos. Así, en vez de comprimir la información de cada punto de la imagen, nos basta con guardar la familia de contracciones que generan el fractal.

Lo primero que se realiza es tomar una partición de la imagen en subconjuntos llamados "regiones dominio". Cada una de estas regiones se sustituirá por la transformada afín que la genera. Cuanto mayor sea el tamaño de los subconjuntos de la partición, mayor será la compresión de la imagen y peor su calidad.

Luego se debe conseguir las "regiones rango", que son subconjuntos de la imagen, de tamaño mayor que las regiones dominio (dos o tres veces mayores), que no tienen que cubrir todo el conjunto, y que pueden superponerse. La idea del algoritmo de compresión es buscar transformaciones contractivas que transformen a las regiones rango en regiones dominio.

Para cada dominio buscamos entre todas las regiones rango la que mediante una transformación contractiva más se parezca al dominio y se almacena dicha transformación. Este proceso es muy lento, al tener que trabajar con un gran número de conjuntos. Se debe tener en cuenta que la codificación de imagen fractal es similar a la cuantificación vectorial con un codebook que contiene todas las posibles combinaciones de las transformaciones aplicadas a los bloques de dominio.

3.2.4. MÉTODOS DE CODIFICACIÓN DE SEGUNDA GENERACIÓN

Los métodos de codificación de segunda generación son adaptaciones de los métodos de codificación de forma de onda, que dividen las imágenes en regiones homogéneas de diferentes formas y tamaños, dependiendo de algunas propiedades como textura, color o movimiento. Estos métodos también son conocidos como *métodos de codificación basados en región o segmentación.* Con ayuda de estos métodos se ha logrado mejorar la eficiencia en codificación utilizando técnicas de adaptación del tamaño de la imagen a píxeles o bloques, además estos métodos mejoran la calidad percibida reduciendo los artefactos[16] que sobreponen dos áreas a la vez no homogéneas (tal como obscurecimiento de los bordes separando las dos imágenes, algo frecuente en la codificación basada en bloques). Por cada segmento, la forma, la situación y los parámetros que regulan la intensidad y distribución del color dentro de ese segmento necesita ser codificado.

Los *métodos de región creciente*, son métodos de segmentación, que emplean una combinación de técnicas de discriminación de textura y borde para tener áreas texturizadas homogéneamente.

Recientemente la morfología matemática ha sido usada para segmentar imágenes. La *codificación de contornos* es la codificación de formas arbitrarias sobre una cuadrícula discreta. La *codificación de cadena* es la manera más simple de codificar exactamente un contorno y no es eficiente a nivel de bit. Los contornos pueden ser codificados aproximadamente escogiendo un juego de vértices de control y por la definición de un polígono o ajustándose a una curva a través de estos vértices.

La otra opción para evitar regiones formadas arbitrariamente es comúnmente conocida como bloque de tamaño variable basado en segmentación. Un ejemplo de esta segmentación es el llamado quadtree. Este código está basado en el principio de descomposición recursiva del espacio. Inicialmente la imagen es

---

[16] **Artefactos o artefactos de bloques:** son las distorsiones de mosaico resultantes al realizar la compresión.

descompuesta en 4 cuadrantes de igual tamaño. Si uno de los cuadrantes no tiene región uniforme, él es subdividido en 4 cuadrantes. Esta descomposición iterativa se detiene, si todo el cuadrante contiene una región uniforme, o si el mismo contiene solamente un píxel. Los quadtrees son construidos en cualquiera de las siguientes maneras top-down o bottom-up, o como una combinación de ambas. La construcción top-down requiere de la descomposición recursiva de un bloque (conocido como un nodo del quadtree), en cuatro subbloques dependiendo del criterio de descomposición. La construcción bottom-up requiere el particionamiento de la imagen en pequeños subbloques y entonces recursivamente se fusionan 4 bloques basados en un criterio de fusión. Las técnicas de descomposición y fusión construyen un quadtree top-down y entonces unen subbloques cercanos para obtener una colección de subbloques que se aproximan a la forma de la región original. La estructura del árbol puede ser codificada eficientemente con un bit por descomposición/fusión. Sin embargo como las formas delineadas son arbitrarias y los bloques son rectangulares, el número final de subbloques es generalmente mucho más alto que con los métodos de región creciente. Algunas extensiones en base a una segmentación de tipo árbol pueden reducir el número de subbloques permitiendo particiones diagonales, en adición a las particiones horizontales y verticales.

El criterio normalmente usado para la homogeneidad es la variación de la intensidad. La intensidad dentro de cada segmento es normalmente modelada como un planar o superficie cuadrática, y los parámetros de estas superficies son calculados por la resolución del sistema de ecuaciones obtenidos por la aplicación del modelo de cada píxel en la región. Entonces los residuos después de ajustarse al modelo son codificados usando métodos convencionales.

3.2.5 CODIFICACIÓN INTERFRAME

Las secuencias de imágenes tienen una considerable redundancia temporal con objetos en la escena ya que la cámara típicamente es sometida a desplazamientos pequeños entre cuadros sucesivos. Los métodos de codificación

que explican esta redundancia que existe entre cuadros adyacentes son conocidos como métodos de codificación interframe.

La *predicción compensada en movimiento* (**M**otion **C**ompensated **P**rediction (MCP)) es el método de codificación interframe más usado. Incluso las extensiones espacio - temporales de métodos de codificación de transformadas y subbandas incluyen una fase de compensación de movimiento. En una secuencia de imagen típica, el movimiento de cuadro a cuadro es una composición de los movimientos del objeto y el movimiento de la cámara en el espacio 3D, proyectado en el plano de la imagen. El movimiento de la cámara da lugar a un movimiento global mientras que los movimientos del objeto causa variaciones locales. La MCP confía en el hecho de que los movimientos local y global pueden ser estimados, entonces un cuadro a ser codificado puede predecirse de un cuadro de referencia cercano temporalmente. El error de imagen después de la predicción, llamado la diferencia de cuadro desplazado (**D**isplaced **F**rame **D**ifference (DFD)) puede ser codificado usando intracoding o usando métodos de codificación basados en segmentación.

Típicamente la composición del movimiento local se estima usando una aproximación a la base actual del modelo de movimiento 3D. La región usada para la estimación del movimiento es normalmente considerada un pedazo planar que esta siendo sometido al movimiento y a una conveniente transformación proyectiva usada para modelar la proyección sobre el plano de la imagen. La traslación solamente del movimiento paralelo al plano de la imagen es la aproximación de mayor uso. Este modelo simple requiere solo de dos parámetros denominados la componentes horizontal y vertical de traslación. La transformación 2D afín (modelo de 6 parámetros) es usualmente una buena aproximación al movimiento real para objetos a distancia razonable, esto puede ser considerado para la traslación, rotación, ajuste y corte de la interframe. La transformación de la perspectiva 2 D (modelo de 8 parámetros) es la mas apropiada para modelar el movimiento de un pedazo planar bajo la proyección de la perspectiva; es así como se puede considerar para las distorsiones de perspectivas inducidas (más notable en objetos cercanos).

La estimación del movimiento (ME) es normalmente realizada para un grupo de píxeles que es probable que tengan los mismos parámetros de movimiento. La estimación de movimiento con un bloque rectangular de píxeles y con el modelo de solo traslación es comúnmente conocido como el *emparejamiento de bloques*, el cual corresponde a encontrar un bloque en el cuadro de referencia que mejor encaje (dando una sensación de distorsión mínima) con el bloque a ser predecido. La función de distorsión es evaluada sobre un rango de búsqueda centrado alrededor de la localización de traslación cero. Sin embargo mínimos y máximos errores cuadrados medios a través de las correlaciones han sido usados como criterio para el mejor emparejamiento, por simplicidad de cálculo, el criterio de diferencia absoluta mínima (**M**inimun **A**bsolute **D**ifference, MAD) definido a continuación es el más usado.

$$MAD = \underset{i}{Min}\sum_{j}\sum \left| I(k,l) - I_{ref}(k+i,l+j) \right|, \quad (i,j) \in S \qquad \text{(Ec. 3.6)}$$

siendo S el vecindario de búsqueda.

Si la función de distorsión es evaluada en todos los posibles desplazamientos de píxel dentro del vecindario de búsqueda, entonces la búsqueda del mejor emparejamiento es llamado *búsqueda exhaustiva*. Ya que la búsqueda en la vecindad puede ser bastante larga en situaciones reales, la complejidad de la búsqueda exhaustiva puede ser lo bastante alta para ser práctica. Varias estrategias de reducción de búsqueda se han sugerido, basándose en asumir que la función de distorsión es monótona en el rango de búsqueda. Las más notables de estas son la *búsqueda logarítmica*, la *búsqueda de 3 pasos* y la *búsqueda de dirección conjugada*. El *emparejamiento de bloque jerárquico* (**H**ierarchical **B**lock **M**atching, HBM) también es logarítmicamente eficiente pero no hace la suposición de monotonía. Cuando se obtiene el mejor de los emparejamientos en desplazamientos de todos los píxeles, la estimación puede interpolarse con exactitudes de subpíxel. La interpolación bilineal usada comúnmente utiliza una combinación lineal de los cuatro píxeles más cercanos para producir el valor del subpíxel. Las traslaciones de las componentes horizontal y vertical de un bloque se conoce como vector de movimiento. Los vectores de movimiento se basan

usualmente en la técnica DPCM para aprovechar la uniformidad del campo en movimiento sobre la imagen.

Considerando pedazos triangulares y estimaciones del vector de movimiento de cada uno de los vértices, los seis parámetros de los modelos afines pueden obtenerse. De igual manera los ocho parámetros del modelo de transformación de perspectiva 2D pueden estimarse de los vectores de movimiento de los vértices de un cuadrilátero.

La MCP en estos casos procede de la siguiente manera: una imagen es particionada de manera estática o adaptiva en particiones triangulares o cuadriláteras; la estimación de los vértices de los vectores de movimiento usan una área pequeña alrededor de cada píxel y los parámetros afines son estimados. La predicción para un pedazo es obtenida deformando el correspondiente triángulo en el cuadro de referencia de acuerdo al modelo afín a ese pedazo. Ya que la estimación del movimiento de vértices puede ser no confiable, una aproximación alterna es refinar iterativamente la estimación del modelo en movimiento, usando la gradiente en descenso o métodos de búsqueda de Gauss-Newton, sobre el conjunto de píxeles dentro de un pedazo.

3.2.6 CODIFICACIÓN BASADA EN MODELOS

Estos métodos de codificación han surgido recientemente y son el resultado de la sinergia entre los tres campos denominados: codificación de imagen, entendimiento de imagen (análisis de la escena) y gráficos computarizados (síntesis de la imagen). Estos métodos van más allá de la información 2D y modelan los objetos físicos diferentes en una escena basada en atributos 3D obteniendo la información disponible a priori sobre la escena. Puesto que las imágenes son codificadas en base a su contenido, estos métodos también ajustan el posicionamiento de la imagen y realizan operaciones de recuperación desde las bases de datos de video.

Mientras las técnicas de codificación convencionales tienen un buen desempeño en altas y medianas velocidades de bits, su desempeño es inadecuado para bajas velocidades. Esto surge parcialmente del hecho de que los métodos convencionales son métodos de codificación de propósito general y no aprovechan los tipos de escena específicos. Por ejemplo en una videoconferencia, el movimiento de la cámara es despreciable y la naturaleza de la escena es usualmente del tipo "cabeza y hombros". El ojo y los movimientos de labios son considerados más importantes. Sin embargo los métodos convencionales no se aprovechan de la naturaleza de la escena y generalmente asignan bits a todas las áreas de la escena con igual importancia.

De aquí que en bajas velocidades de bits, la calidad percibida se degrada severamente. También la magnitud de la compensación de movimiento lograda se considera muy importante a velocidades de bits bajas, ya que muy pocos bits están disponibles para codificar regiones subcompensadas. Ahora los modelos de compensación de movimiento simple serán reemplazados por modelos más complejos. Si los objetos reales en la escena 3D y su movimiento 3D pueden modelarse, entonces la secuencia puede ser sintetizada desde los parámetros de los modelos para transmitir solo el objeto codificado y parámetros del modelo de movimiento. Esta es la finalidad perseguida por los métodos de codificación basado en modelos, generalmente los métodos de codificación basados en modelos rastrean los objetos sobre el tiempo, como oposición a la predicción de codificación de cuadro desde un cuadro de referencia.

Los módulos tipicos de tales esquemas de codificación son modelados de acuerdo a: los modelos de análisis de la imagen, codificación de parámetros modelos, manipulación de falla de modelo y síntesis de la imagen de los modelos. La fase del análisis consiste generalmente de una fase de segmentación para obtener las diferentes regiones homogéneas en la escena. Si la naturaleza de los objetos es conocida con anterioridad, como en una videoconferencia, pueden usarse superficies 3D convenientes o modelos volumétricos. El modelo de fallo

(**M**odel **F**ailure (MF)) corresponde a regiones que no pueden ser modeladas correctamente (tal como fondos descubiertos). Estas regiones son generalmente manejadas por métodos de codificación de formas de onda. Asumiendo que los objetos son flexibles y usando modelos de movimiento para objetos flexibles, las regiones MF son considerablemente reducidas. Las regiones MF también son reducidas de tal forma que pueden permitir tener distorsiones geométricas (errores pequeños en el tamaño y posición de objetos) que son más tolerables perceptualmente que la distorsión introducida debido a la cuantización inadecuada de las áreas MF (común para velocidades de transmisión muy bajas).

Para secuencias en las que el movimiento de la cámara es dominante sobre los movimientos de los objetos y probablemente cubre ubicaciones espaciales adyacentes sobre un periodo largo de tiempo, una nueva clase de métodos conocidos como *métodos de codificación basados en mosaicos* se han desarrollado. Estos métodos registran los cuadros en el tiempo usado técnicas apropiadas de distorsión para calcular el movimiento de la cámara y obtener una composición panorámica de imagen en mosaico. De tal forma que las redundancias temporales se eliminan. El mosaico es codificado usando métodos intracoding estándar. El mosaico codificado y los parámetros de registro son suficientes para reconstruir la secuencia. Se manejan regiones con movimiento local a través de la operación "cortar y pegar".

Debe notarse que debido al conocimiento usado anteriormente, un codificador basado en modelo para un tipo particular de escena no es óptimo para codificar otro tipo de escena. Además la eficiencia de codificación con estos codificadores muestra que solo puede ser usada para escenas no tan complejas. Algunos investigadores han propuesto un codificador de switcheo híbrido, que use codificación basada en modelos para objetos que cumplan con el modelo y una codificación basada en formas de onda para regiones de fallo, con el objetivo de poder codificar escenas más complejas.

3.2.7 ESTRUCTURA MULTIRESOLUCIÓN PARA CODIFICACIÓN DE VIDEO

Una estructura multiresolución es una estructura eficiente de datos para codificación de imagen que ofrece varias características deseables, tales como escalabilidad espacial de algoritmos complejos, transmisión progresiva y una base psicofísica para análisis y representación de imágenes. A continuación se provee una apreciación de la representación de estructura de múltiple resolución y se delinea brevemente las características anteriormente citadas.

3.2.7.1 Descomposición multiresolución

La descomposición multiresolución, también conocido como descomposición piramidal de una imagen, es la descomposición de una imagen en subimágenes con una progresiva disminución de las resoluciones espaciales. Tal descomposición posibilita el refinamiento jerárquico de varios métodos de análisis de imagen desde un simple nivel descriptivo, hasta los niveles de realce y refinamiento en la resolución espacial. La descomposición también ofrece medios compactos de codificación de imagen como se describirá a continuación.

El refinamiento de la resolución espacial es eficiente en los procesos de cálculo y permite escalabilidad espacial. También pueden hacerse refinamientos prematuros a un nivel global, sin ser alterados por detalles espaciales locales. Los experimentos en fisiología visual humana y psicofísica han mostrado que el sistema visual humano es selectivo en espacio-frecuencia y que el ancho de banda de estos filtros espaciales es como de una octava. En otras palabras, las bandas de las diferentes frecuencias tienen aproximadamente el mismo ancho en una escala logarítmica; esto sugiere la posibilidad de que el propio sistema visual humano emplee una representación multiresolución.

Una descomposición que emplea filtros de ancho de banda de octavas (llevando a cabo un submuestreo por un factor de 2) para obtener las subimágenes de múltiples resoluciones se conoce como una descomposición Dyadic. Puesto que un filtro Gaussiano tiene buen desempeño en ambos dominios, espacial y

frecuencial, la primera descomposición propuesta de múltiples resoluciones usa dicho filtro. Sin embargo este filtro no tiene unidad de ganancia en el pasa banda entero con lo cual resulta un excesivo alisamiento de la señal. La colección de subimágenes con una progresiva disminución de resolución es llamada una *pirámide gaussiana* y será usada para un refinamiento progresivo. Por sobre muestreo la imagen en el nivel (l +1) se multiplica por un factor de 2 y se interpola usando el mismo filtro pasa bajos, así una imagen de baja resolución con la misma magnitud espacial como la imagen en el nivel l puede ser obtenida.

La diferencia entre estas dos imágenes que tienen la misma magnitud espacial, proporciona los detalles de alta frecuencia espacial presentes en el nivel l. La colección de los detalles de las imágenes en los diferentes niveles de resolución se llama una *pirámide Laplaciana*, la diferencia de las imágenes filtradas con el método gaussiano está en aplicar directamente un operador Laplaciano. La figura 3.13 ilustra la construcción de las pirámides guasiana y Laplaciana.

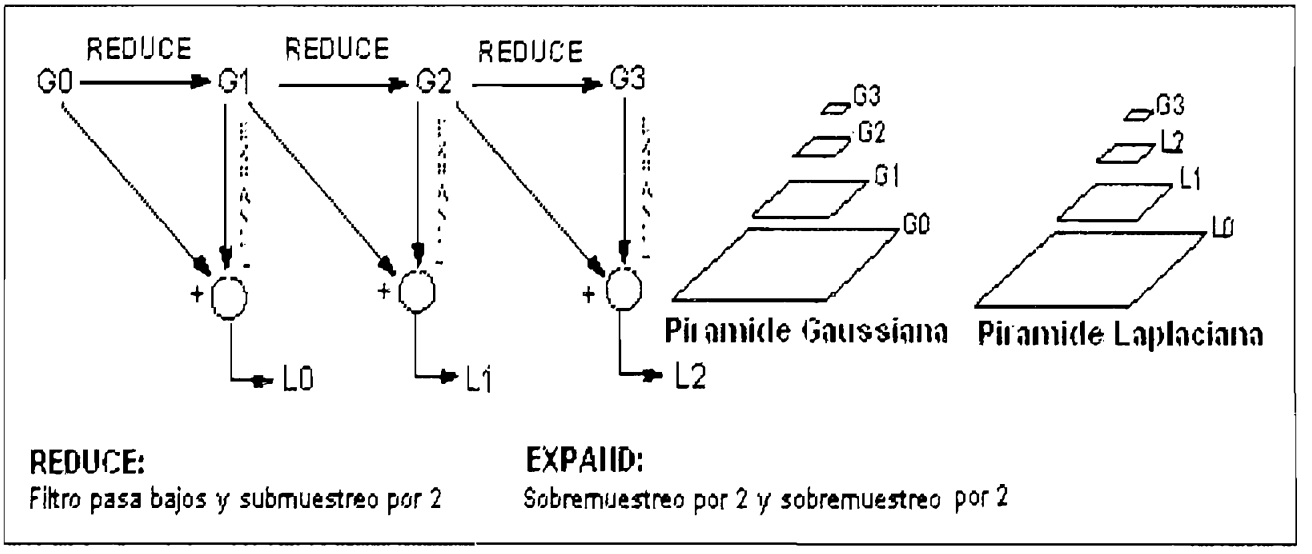


Figura 3.13 Pirámide Gaussiana y Laplaciana.

El nivel de menor resolución de subimagen de la pirámide Laplaciana es el mismo que el nivel de menor resolución de la pirámide Gausiana. Puesto que los detalles de la imagen son típicamente escasos, estos pueden comprimirse eficazmente. La imagen de pasa bajo contiene la mayoría de la energía y puede

ser codificada eficazmente debido a su magnitud espacial reducida. Así la pirámide Laplaciana constituye una eficiente representación codificada de la imagen original.

3.2.7.2 Teoría de bancos de filtros multifrecuenciales

Aún cuando la descomposición multiresolución y la motivación para operadores Gausianos y Laplacianos surge de investigaciones de visión, los principios en que se basan vienen de la teoría de un banco de filtros multifrecuenciales en procesamiento de señales. Esta teoría presenta la estructura para el diseño de filtros convenientes requeridos en sistemas que manejan diferentes velocidades de muestreo. El diseño de filtros apropiados ayuda en el logro de características importantes tales como: cancelación del aliasing, reconstrucción perfecta y reducción de la distorsión de amplitud y fase.

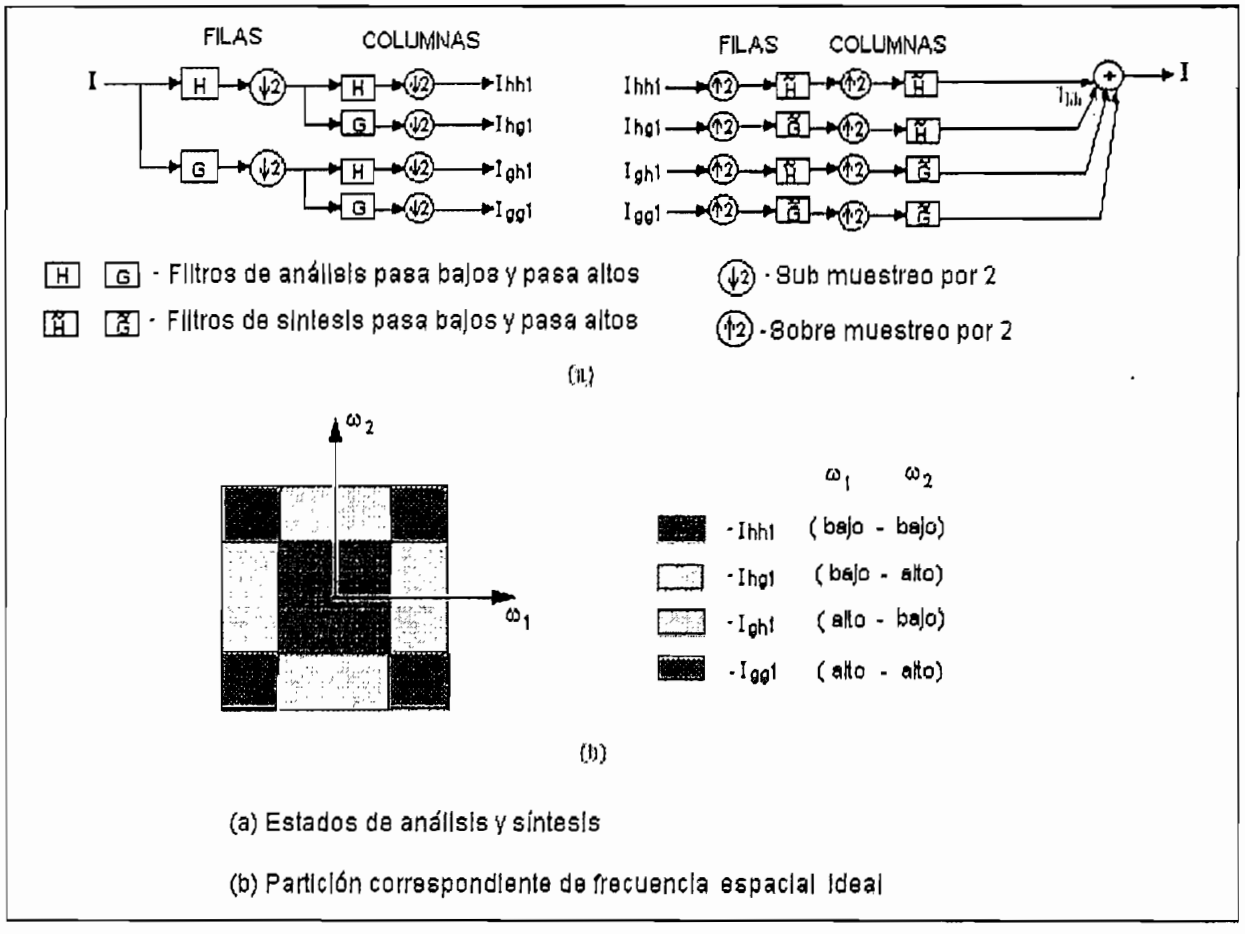


Figura 3.14 Descomposición de subbandas Dyadic de una imagen I.

Así esta teoría forma la base para la descomposición de subbandas, en las que una imagen es descompuesta en varias imágenes no sobrepuestas (o mínimamente traslapadas) en subbandas de frecuencia espacial durante la fase de análisis.

Cada una de estas bandas puede ser procesada de manera diferente. Por ejemplo el sistema visual humano es conocido por ser mas sensible a las orientaciones espaciales horizontales y verticales que a otras orientaciones arbitrarias. Esto puede ser explotado por una cuantización de menor resolución de subbandas con orientación diagonal. Durante la síntesis, todas las subbandas procesadas son sobre muestreadas e interpoladas usando correctamente los filtros de reconstrucción diseñados y sumándolas conjuntamente.

Para el caso de descomposición Dyadic, los dos filtros de análisis son imágenes espejo una de otra con respecto a la frecuencia de cuadratura 2π/4; así los filtros son referidos tal como filtros de cuadratura de espejo. La figura 3.14 ilustra los pasos de análisis y síntesis para una fase simple de descomposición Dyadic[17] y el resultado ideal de la partición de frecuencia.

Las correspondientes ecuaciones de análisis y síntesis son las siguientes:

$$Ihhl(m,n) = \sum_k h(k) \sum_l h(l) I(2m-k, 2n-l) \qquad \text{(Ec. 3.7)}$$

$$I_{hh}(m,n) = \sum_k \tilde{h}(2k+i) \sum_l \tilde{h}(2l+j).Ihhl\left(\left[\frac{m}{2}\right]-k, \left[\frac{m}{2}\right]-l\right) \qquad \text{(Ec. 3.8)}$$

donde: i y j son 0 y 1 dependiendo si m y n respectivamente son par o impar. Una descomposición de múltiples resoluciones es lograda por la descomposición recursiva de solamente subimágenes de pasa bajo, como se muestra en la figura 3.15.

[17] Un filtro separable 2D como se muestra en la figura 3.14, produce 4 subbandas y es equivalente a las subbandas obtenidas después de dos niveles de descomposición usando un filtro 2D no separable, sin embargo desde el punto de vista de compresión con base psicofísica, un filtro no separable es considerado mejor.

La pirámide de la resolución así obtenida es similar a la pirámide gaussiana.

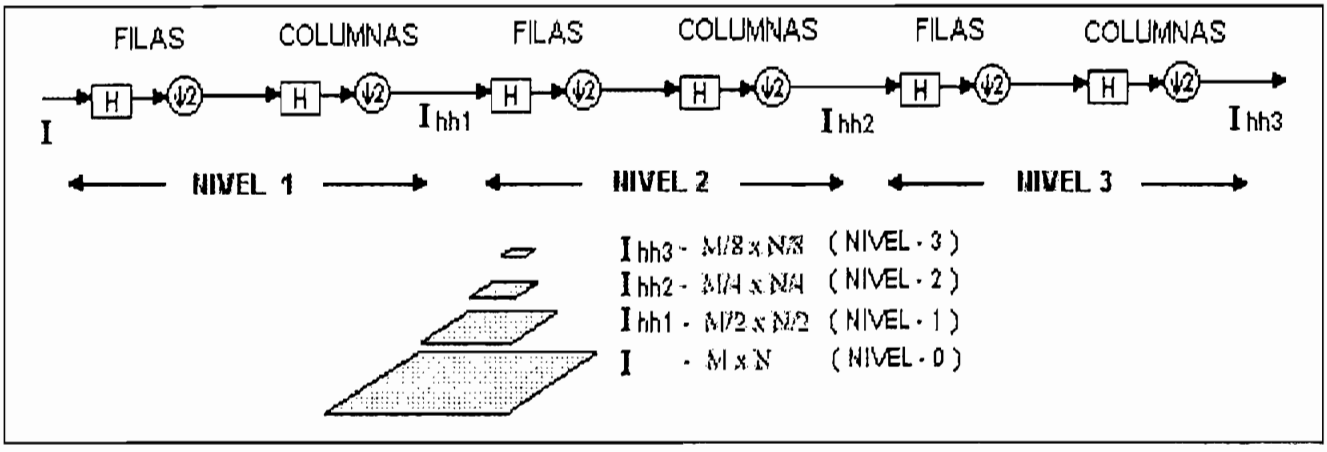


Figura 3.15 3 - niveles de descomposición multiresolución y la pirámide de resolución.

3.2.7.3 Teoría de descomposición Multiresolución y Wavelet.

La descomposición wavelet (de ondículas) es una poderosa alternativa a la tradicional técnica de análisis de fourier para análisis de señales. Las técnicas de análisis de Fourier usan funciones bases con un soporte espacial (o temporal) fijo para analizar todas las frecuencias. Así una buena localización en ambos dominios, espacial y temporal no es posible. La descomposición wavelet emplea un conjunto de funciones bases que son copias trasladadas y dilatadas (en una escala espacial / temporal) de una sola función conocida como la *función escalar* (scaling). Así el conjunto de funciones base consiste de funciones con soporte variable donde una buena localización en ambos dominios es posible. La relación íntima entre la teoría del banco de filtros, análisis wavelet y la descomposición multiresolución fue hecha popular por la teoría de descomposición de múltiples resoluciones de Mallat. La estrecha relación entre el banco de filtros de múltiples resoluciones y la teoría wavelet provee una rica variedad de familias de filtros de donde escoger, dependiendo de los requerimientos específicos. La clase de filtros normalmente usada basados en wavelet, son los wavelets de soporte sólido ortonormal de Daubechies. Como el nombre lo sugiere, estos filtros tienen un soporte compacto (deseable para eficacia computacional) aún manteniendo una razonable característica de filtro de media banda (necesario para minimizar el aliasing). Los correspondientes coeficientes del filtro son derivados aplicando la

ortonormalidad bajo traslaciones uniformes y regularmente forzadas (las cuales imponen ceros adicionales en el muestreo de frecuencia para atenuar la respuesta a altas frecuencias del filtro).

Los filtros pasa bajos y pasa altos son filtros de espejo en cuadratura (Quadrature Mirror Filter, QMF) y los filtros de síntesis son simplemente versiones inversas de filtros de análisis. Sin embargo, los filtros ortogonales tienen un número igual de coeficientes y son asimétricos; así estos tendrán una respuesta de fase no lineal. Esta distorsión de fase da lugar a desplazamientos espaciales variantes sobre la imagen, lo cual no puede ser aceptable en ciertas aplicaciones que requieren una extracción precisa de la posición. Una clase de filtros simétricos con un número impar de coeficientes, conocido como filtros *biortogonales*, han sido diseñados para superar este inconveniente. En este caso, los filtros pasa bajos y pasa altos tienen diferentes longitudes.

3.2.7.4 Pirámide Laplaciana vs descomposición de subbanda para codificación

Aunque la descomposición piramidal y de subbanda son en principio similares, éstas ofrecen dos diferentes representaciones de la imagen original. La representación de pirámide Laplaciana requiere cuatro tercios del número de píxeles en el nivel de resolución mas alto. Este aumento en el número de píxeles se debe a la presencia de redundancia en la representación. Por otro lado, la representación de una imagen en términos de sus subbandas no resulta en el incremento del número de píxeles. Esto se debe al bajo muestreo por un factor de 2 en cada dirección. La introducción del aliasing debido a los filtros de media banda no ideales puede ser cancelada por un adecuado diseño de los filtros de análisis y síntesis. Así la descomposición de subbanda es usualmente preferida sobre la descomposición piramidal para propósitos de codificación. Sin embargo, la codificación de pirámide Laplaciana tiene la ventaja que los errores de cuantización en los niveles mas altos de la pirámide pueden ser incluidos en las imágenes con detalle de nivel mas bajo, evitando así la acumulación de errores. Solo los errores de cuantización en la codificación del nivel 0 de detalle permanecen en la imagen. Tal realimentación de error de cuantización no es

posible en codificación de subbanda y los errores de cuantización también pueden llevar al aliasing durante la reconstrucción. Por otro lado , la codificación de subbanda puede explotar la sensitividad de orientación del sistema visual humano.

La correlación a través de las subbandas puede ser explotada por la cuantización de los vectores formados por los coeficientes correspondientes en las diferentes subbandas. Ambas representaciones ofrecen capacidad de transmisión progresiva en la que las subimágenes de menor resolución se transmiten primero y las imágenes de detalle se agregan progresivamente. Esto encuentra aplicación en buscadores de bases de datos de imágenes como son usuarios que pueden descargar primero las detalles de menor resolución de las imágenes y si es necesario, puede descargar luego las imágenes en detalle, ahorrando así un considerable ancho de banda.

También, en canales de transmisión propensos a error, las subimágenes de menor resolución que son mas críticas pueden ser protegidas con códigos de corrección de error. Al respecto, la descomposición multiresolución también permite priorizar la información.

3.2.7.5 Emparejamiento de bloque jerárquico en la resolución piramidal

Como se mencionó en la descomposición multiresolución, la múltiple resolución piramidal permite el refinamiento jerárquico de las estimaciones de movimiento. El emparejamiento de bloque jerárquico fue introducido en la codificación interframe como una técnica computacional eficiente de emparejamiento de bloque. Generalmente, la mayoría del cálculo de un codificador de video es la estimación del movimiento. Una búsqueda exhaustiva sobre un rango de ±S píxeles horizontales y verticales requiere de $(2S+1)^2$ búsquedas. La complejidad de cada búsqueda es proporcional al número de píxeles N usados en el cálculo del MAD. Algunas estrategias de reducción de la búsqueda, las que asumen un mínimo único dentro del área de búsqueda fueron presentados en la codificación interframe. Sin embargo, debido al ruido en las áreas sin rasgos distintivos y

posibilidades de patrones periódicos, la función MAD sobre el rango de búsqueda tiene múltiples mínimos.

Así estos métodos de reducción de búsqueda probablemente llevan a estimaciones erróneas del vector de movimiento. Por otro lado en el emparejamiento de bloque jerárquico la estimación empieza a un nivel de resolución menor, donde los detalles locales se han promediado y solo los detalles de menor resolución permanecen en la imagen. Así más rasgos globales son emparejados en los niveles de menor resolución y estas estimaciones fiables son refinadas de acuerdo a los detalles mas finos en los subsiguientes niveles de resolución.

Si se emplean n niveles de descomposición, el rango de búsqueda en el nivel n es $\pm S/2^n$ y solamente $(S/2^{n-1}+1)^2$ búsquedas son requeridas en el nivel de menor resolución. Puesto que el número de píxeles en el nivel l es $N/4^l$, la complejidad para la búsqueda también es baja. En los subsecuentes niveles, las estimaciones del nivel de resolución previo pueden ser refinadas sobre un rango de $\pm k$ píxeles centrado alrededor de la estimación. Así la complejidad de la búsqueda global para un bloque de N-píxeles sobre un rango de búsqueda de $\pm S$ píxeles es dado por:

$$\alpha.N\left\{\frac{1}{4^n}\cdot\left(\frac{S}{2^{n-1}}+1\right)^2+\frac{4}{3}\cdot(2k+1)^2\right\} \qquad \text{(Ec. 3.9)}$$

donde α es la constante de proporcionalidad asociado con la complejidad de búsqueda y los (4/3)N es la suma de píxel asintótico sobre la pirámide. La velocidad de la búsqueda compleja para la búsqueda exhaustiva, y emparejamiento de bloque jerárquico puede ser dado aproximadamente por,

$$\frac{1}{2^{-4n}+\frac{4}{3}\left(\frac{2k+1}{2S+1}\right)^2} \qquad \text{(Ec. 3.10)}$$

Ambos términos del denominador son significativamente menores que la unidad para n moderados, S grande y k pequeño. Así la unión de bloque jerárquico resulta en una significante reducción en la complejidad del cálculo. Para un

ejemplo típico con S=64, n=3 y k=2, el número de cálculos se reduce por un factor de 445.

En el refinamiento descrito anteriormente, el número de píxeles usados para el emparejamiento de bloque disminuye con la resolución. Esto puede producir un emparejamiento inestable en los niveles de menor resolución puesto que hay menos rasgos para emparejar dentro de un bloque.

Una alternativa es mantener constante el tamaño de bloque en todas las resoluciones. Así un bloque en nivel l corresponderá a cuatro bloques en el nivel (l + 1), la figura 3.16 ilustra esta juntura de bloque jerárquico. En este caso, el número de cálculos por bloque es el mismo que el descrito por la ecuación 3.9.

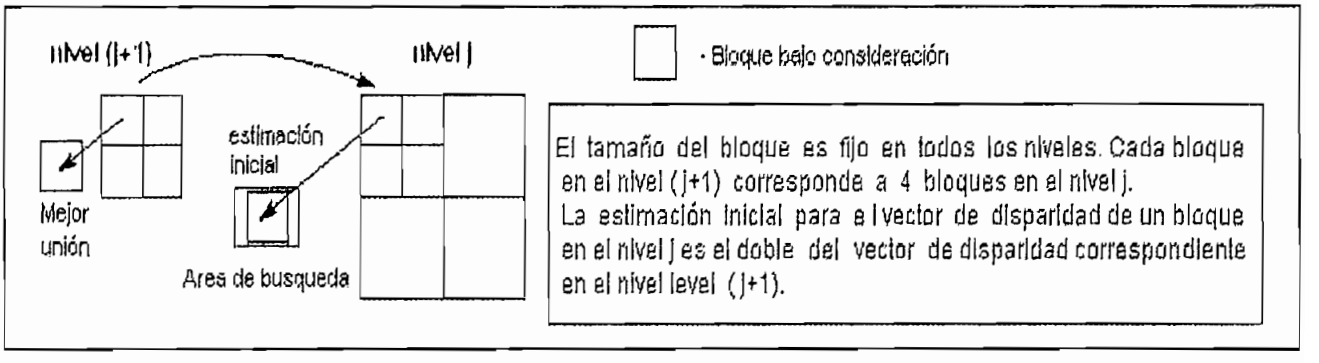


Figura 3.16 Movimiento jerárquico o estimación de la disparidad en una pirámide multiresolución Dyadic.

3.2.7.6 Otras aplicaciones de filtros multifrecuenciales en codificación de video.

La interoperabilidad de codificadores y decodificadores de video requiere el manejo de una amplia variedad de formatos de despliegue. Los diferentes estándares de televisión tales como NTSC, PAL y SECAM que se usan en diferentes partes del mundo tienen diferentes tamaños de displays. El propuesto por HDTV tiene una relación de aspecto de 16:9 y las películas modernas tiene una relación de aspecto de 3:2, como oposición a la convencional relación de aspecto de 4:3. Así para poder hacer uso del máximo de la resolución disponible en un display, un reajuste eficiente del esquema es necesario. Mientras la descomposición dyadic provee un escalamiento solamente por múltiplos de 2, las

proporciones entre estos diferentes sistemas no son enteros. La teoría de bloque de filtros multifrecuenciales proporciona una eficiente manera de manejar el submuestreo y sobremuestreo por diferentes factores. Esto proporciona un incentivo adicional para usar una multiresolución basada en aproximaciones, así que el mismo recurso de hardware puede ser compartido para decodificar y desplegar escalamiento sobre una variedad de formatos de displays. La diferente velocidad de tramas entre diferentes fuentes de video (60 Hz y 50 Hz de velocidad de repetición de campos en TV y 24 cuadros por segundo en películas) pueden también ser manejadas si el concepto de múltiples resoluciones es extendido en la dimensión temporal.

3.2.8 COMPRESIÓN DE IMÁGENES ESTEREOSCÓPICAS

En esta sección se explicará la Predicción Compensada en Disparidad (DCP), que permite predecir una vista de un par de imagen estéreo dada la otra vista, además se enfoca el desarrollo de una Segmentación Basada en Disparidad (DBS), un modelo de codificación de estructura de árbol y segmentación de disparidad. Finalmente el método DBS es comparado con el método de bloque de tamaño fijo (FBS) que se basa en una prueba fija de pares de imágenes estereoscópicos. La segmentación, predicción y residuos luego de la predicción son mostrados por un muestreo de imágenes pares.

3.2.8.1 Predicción Compensada en Disparidad (DCP)

Anteriormente se mencionó el concepto de estimación de disparidad, además de conocer que un par de imágenes estereoscópicas es formado por dos vistas de la misma escena desde dos perspectivas ligeramente diferentes. Ahora en el barrido de los píxeles que son ocluidos por objetos de la escena o por límites del cuadro, existe una correspondencia uno a uno entre los píxeles en las dos vistas[18]. Este

[18] La correspondencia es en general aproximada, y es exacta solamente en el caso limitado de píxeles infinitesimalmente pequeños.

hecho puede explotarse para predecir el par de imagen de una vista dada la otra, así como lo muestra la figura 3.17. Sin embargo resolver la correspondencia o problema de estimación de disparidad es muy importante.

Esto se da debido a lo que es conocido en la teoría de la visión como el problema de la abertura. Las correspondencias pueden ser no confiables si una región muy pequeña es considerada durante la exploración, mientras que incluyendo áreas muy grandes durante la exploración pueden llevar a estimaciones erróneas como que dos objetos de diferente profundidad sean considerados juntos y un valor común de disparidad se asigna a esa región.

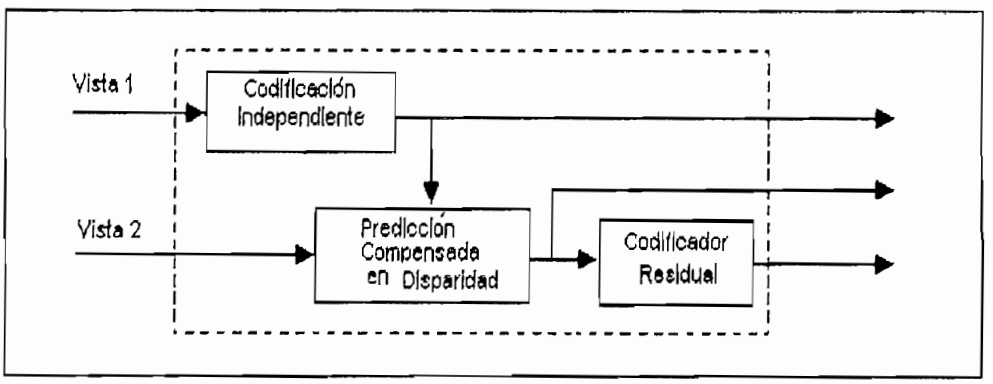


Figura 3.17 Codificación basada en predicción compensada en disparidad de un par de imagen estereoscópico.

Así para diferentes regiones de la imagen, se necesitan diferentes tamaños de bloques que dependen del detalle de disparidad local. Puesto que los detalles de la estimación local no están inicialmente disponibles, una estimación iterativa de la disparidad es requerida. El problema se presenta cuando las correspondencias tienen que ser decodificadas. La estimación del píxel-útil requeriría codificación de disparidad por cada píxel. Esto no produce una buena compresión.

Así los métodos de estimación de disparidad usados para codificar un par estereoscópico (en contraposición con los métodos usados para obtener profundidad estéreo) típicamente asumen una disparidad constante sobre un bloque de píxeles[19]. En este caso el problema es similar a los métodos de

[19] Físicamente, esto implica un pedazo planar que queda paralelo a los sensores de la imagen a una profundidad fija.

codificación interframe mencionados anteriormente. Sin embargo, la mayor diferencia en este caso es que, debido a las restricciones epipolares mencionadas en la geometría de la imagen estereoscópica, la búsqueda por el píxel correspondiente (o bloque) se restringe a una sola dimensión. En contraste, la estimación de movimiento requiere una exploración 2D. Para la geometría de imagen estereoscópica con los ejes paralelos, la exploración para el mejor emparejamiento de un bloque se restringe para estar dentro de las correspondientes líneas analizadas en la otra vista. Además de simplificar la exploración, esto también mejora la codificación de disparidades, ya que las disparidades en este caso son escalares.

3.2.8.2 Predicción compensada en disparidad (DCP) basada en tamaño de bloque fijo (FBS)

Varios investigadores han desarrollado esquemas de codificación de imagen estereoscópica basados en DCP. Aquí se describe algunos de estos métodos y se señala sus limitaciones, una secuencia de imagen estereoscópica es modelada como procesos fijo y estocástico discreto que emiten dos enteros desde un conjunto finito de enteros que representan el conjunto de todas las posibles imágenes (para un tamaño de cuadro dado y un número de niveles de intensidad). Basado en este modelo, se muestra que la estructura del codificador de la figura 3.17 proporciona una representación de codificación óptima si las imágenes son codificadas por el método lossless. También se muestra que esta estructura es casi óptima si las imágenes son codificadas con respecto a un criterio de fidelidad. Sin embargo, la cercanía a lo óptimo pueden lograrse si la dependencia de una vista en la otra puede ser totalmente explotada. El modelo estocástico simple descrito anteriormente no lo provee ningún método. Desde un punto de vista práctico, se presenta un algoritmo de emparejamiento de bloques basado en bloques de tamaño fijo (**F**ixed-**B**lock-**S**ize based **B**lock **M**atching **A**lgorithm (FBS-BMA)) para la estimación de la disparidad. Los estándares de codificación de video internacional adoptan FBS-BMA para estimación del movimiento debido a su simplicidad de aplicación, sin embargo estos métodos tienen ciertas limitaciones inherentes de las que se hablará a continuación.

Los pares típicos de imagen estereoscópica tienen áreas grandes de disparidad binocular cercana y constante. La compensación de disparidad basada en tamaño de bloque fijo falla al no aprovechar tales regiones y da como resultado una disparidad significativamente más alta codificando la imagen más de lo necesario. Si el mapa de disparidad estimado es uniforme, la imagen puede ser codificada eficazmente por codificación predictiva. Sin embargo usando emparejamiento de bloque con áreas pequeñas sin rasgos distintivos, conducen a emparejamientos falsos que conlleva a una codificación predictiva de disparidad de bloque inefectiva. Cuando los bloques de tamaño fijo fallan a través de objetos en dos profundidades diferentes, estimaciones incorrectas son producidas. Así los errores después de la compensación de disparidad son más significativos en los bordes de los objetos, requiriendo una codificación residual elevada. Además las vistas intermedias, basadas en sintetización en un mapa de disparidad con falsos e incorrectos emparejamientos son inexactos.

3.2.8.3 Segunda generación y métodos de estimación de disparidad basada en modelos.

Varios métodos basados en bordes para resolver el problema de correspondencia han sido propuestos, y algunos de estos métodos han sido extendidos para su uso en aplicaciones de codificación. Estos métodos típicamente detectan la intensidad de los bordes mediante la utilización de la operación Laplaciana de Gausiana y extrayendo los cruces por cero. Los bordes extraídos son aproximados a segmentos de línea recta y etiquetados. La correspondencia es establecida para un borde en una vista por la búsqueda de un borde con similar orientación y longitud en la otra vista, usando un método de optimización conveniente. Métodos de programación dinámica han sido propuestos para establecer tales correspondencias, las cuales en los bordes necesitan ser propagadas a otros píxeles. En general, el contorno o los esquemas de estimación de disparidad basadas en los bordes son computacionalmente intensivos y no son eficaces desde el punto de vista de la codificación.

Recientemente los métodos de codificación de imagen basados en modelos se han aplicado para hacer la compensación de disparidad adaptable a los objetos actuales presentes en la escena. Estos métodos de codificación son satisfactorios solo para aplicaciones restringidas. En general el rendimiento de estos métodos no encajan bien con el número de objetos en la escena y con la complejidad de la cámara y el movimiento de los objetos. También el estado de análisis de los objetos en estos métodos son de cálculo complejo, así estos métodos no pueden ser aplicados para escenas arbitrarias. El mejoramiento en el rendimiento de la codificación sobre métodos convencionales para imágenes en general aún no ha sido establecido.

3.2.8.4 Motivos para una nueva aproximación.

De esta manera el cálculo simple de los métodos de predicción compensada de disparidad basada en bloques de tamaño fijo no proporcionan una representación de codificación óptima. Estos métodos avanzados manejan este problema, pero no trabajan bien para imágenes arbitrarias. Lo que se necesita es un nuevo acercamiento en bits de codificación de disparidad al detalle de la disparidad local presente en la imagen, mientras se mantiene una baja elevación en la codificación de estos segmentos a una moderada complejidad computacional. Concluyendo que una representación óptima para la codificación de disparidad puede ser obtenida por segmentación del par de imagen estereoscópico basado en la disparidad.

Se asume que un modelo conveniente puede ser formulado para mapear un juego de píxeles en una vista del par estéreo a un conjunto correspondiente de píxeles en la otra vista. Se considera N regiones arbitrarias tal que la correspondencia para píxeles dentro de cada región es especificada por los parámetros para esa región. R_k^{modelo} es el número de bits necesarios para codificar los parámetros modelos para le región K-esima[20]. Se considera además que R_k^{forma} es el

[20] Se asume un diferente R_k^{modelo} para diferentes regiones con la finalidad de que los bits por codificar en base a los parámetros modelos pueden ser reducidos por codificación predictiva y entropía.

número de bits necesarios para codificar la región K-esima, en un modelo lossy o lossless. Se realizan aproximaciones en el modelo y en la forma, habrá errores después del modelo de predicción. R_k^{error} es el número de bits necesarios para codificar estos errores sujetos a un criterio de fidelidad. Además habrá regiones que debido a la oclusión, no tienen una región correspondiente en la otra vista. R^{occ} es el número de bits necesarios para codificar estas regiones, por intracodificación o encontrando una región similar en la otra vista y codificando los residuos. El número total de bits necesarios para codificar una vista dada la otra, sujeta a un criterio de fidelidad es:

$$R^{total} = \sum_{k=1}^{N} (R_k^{modelo} + R_k^{forma} + R_k^{error}) + R^{occ} \qquad \text{(Ec. 3.11)}$$

Esta expresión muestra los diferentes problemas que afectan el desempeño de la codificación. Para los métodos basados en FBS, el término $R_k^{forma} = 0$ debido a que las regiones son escogidas independientemente de las imágenes y el número de bloque N^{fbs} es mucho más grande que N. Además el incremento de bits debido a que el N^{fbs} es más grande, produce que el término R_k^{error} también se incremente para bloques que contienen objetos de diferente profundidad. Los métodos de codificación basados en modelos así como en contornos, no se basan en la segmentación de disparidad y generalmente tienen un mayor número de regiones que N. Además el término R_k^{forma} es de un valor muy alto. Se desarrolla luego una nueva aproximación que se basa en segmentación de disparidad la cual minimiza R_k^{forma} utilizando multiresolución basada en métodos de descomposición en estructura de árbol.

3.2.9 Segmentación basada en disparidad

A continuación se habla sobre un nuevo acercamiento para la codificación compensada en disparidad de pares de imágenes estereoscópicas. Este acercamiento se refiere a la segmentación basada en la disparidad, combinando intensidad e información de disparidad para segmentar una vista de un par de imagen estereoscópica dada la otra y lograr una representación de codificación que corresponde con el detalle de disparidad local. Una descomposición quadtree es empleada como oposición a la segmentación basada en contornos, debido a las buenas escalas de una elevada codificación de estructura segmentada para escenas complejas. Una solución computacional eficiente no iterativa, que reduce la elevada segmentación, es obtenida por el uso de una estructura de multiresolución. Las ubicaciones particionadas por la generalización QTD (Quadtree Decomposition) son calculadas usando el esquema de detección de borde descrito anteriormente.

3.2.9.1 Estructura Multiresolución para segmentación basada en disparidad (DBS).

Para segmentación basada en disparidad binocular, se necesita un mapa de disparidad exacto. Pero un mapa exacto de disparidad solo puede ser obtenido con una buena segmentación de la imagen tomando en cuenta disparidad, discontinuidad y fidelidad. Esto implica una solución iterativa, que no se la puede realizar por métodos computacionales. Sin embargo, una estructura multiresolución permite refinar progresivamente ambas particiones y sus disparidades de una resolución baja ó alta, reduciendo así significativamente la carga computacional asociada. Esta estructura también proporciona varias características deseables:

(1) Una resolución mixta de un esquema de codificación de imagen estereoscópica puede realizarse con facilidad dentro de la estructura;

(2) Como se describirá en la sección 3.2.9.2, la estimación multiresolución permite aplicar diferentes estrategias de subdivisión para reducir la información de codificación debido a la elevada segmentación;

(3) La estimación de multiresolución también reduce emparejamientos falsos evitando mínimos locales durante el emparejamiento de bloque;

(4) Además el esquema de codificación entero puede llegar a ser escalable en resolución. La exactitud de estimación no depende mucho de la selección de los filtros de análisis.

3.2.9.2 Descomposición quadtree general.

La descomposición quadtree de una imagen es una partición recursiva estructurada que divide una imagen en bloques rectangulares basados en un criterio de subdivisión. La figura 3.18 muestra un quadtree típico. En cada nivel del árbol, los bloques están formados por los nodos que pueden subdividirse y por los nodos sin división. Generalmente, un bloque es dividido solamente en los puntos medios de sus lados. En dicha descomposición regular, la estructura del árbol, el tamaño y localización de cada nodo pueden representarse usando solamente 1 bit/nodo. Aquí la elevada necesidad para representar la estructura del árbol, denominada como segmentación elevada, es muy pequeña. Sin embargo como la localización de las particiones son obtenidas independientemente de las características dentro de la imagen, generalmente la descomposición regular ha resultado en un número más grande de bloques.

La homogeneidad espacial de un bloque y movimiento de bloque han sido usados como criterio de subdivisión. A continuación se expone un novedoso criterio de particionamiento. Puesto que las escenas típicas tienen regiones grandes que están aproximadamente a una distancia constante de la cámara, una segmentación orientada al objeto puede ser obtenida usando la disparidad o profundidad de un bloque como el criterio de subdivisión. Así el número de bloques de disparidad a ser codificado después de la DCP es considerablemente reducido para escenas típicas. El uso de descomposición regular podría disminuir la elevada segmentación, pero podría incrementarse el número final de bloques

después de la descomposición y de esta manera se podría aumentar el número de bits necesarios para codificar estas disparidades.

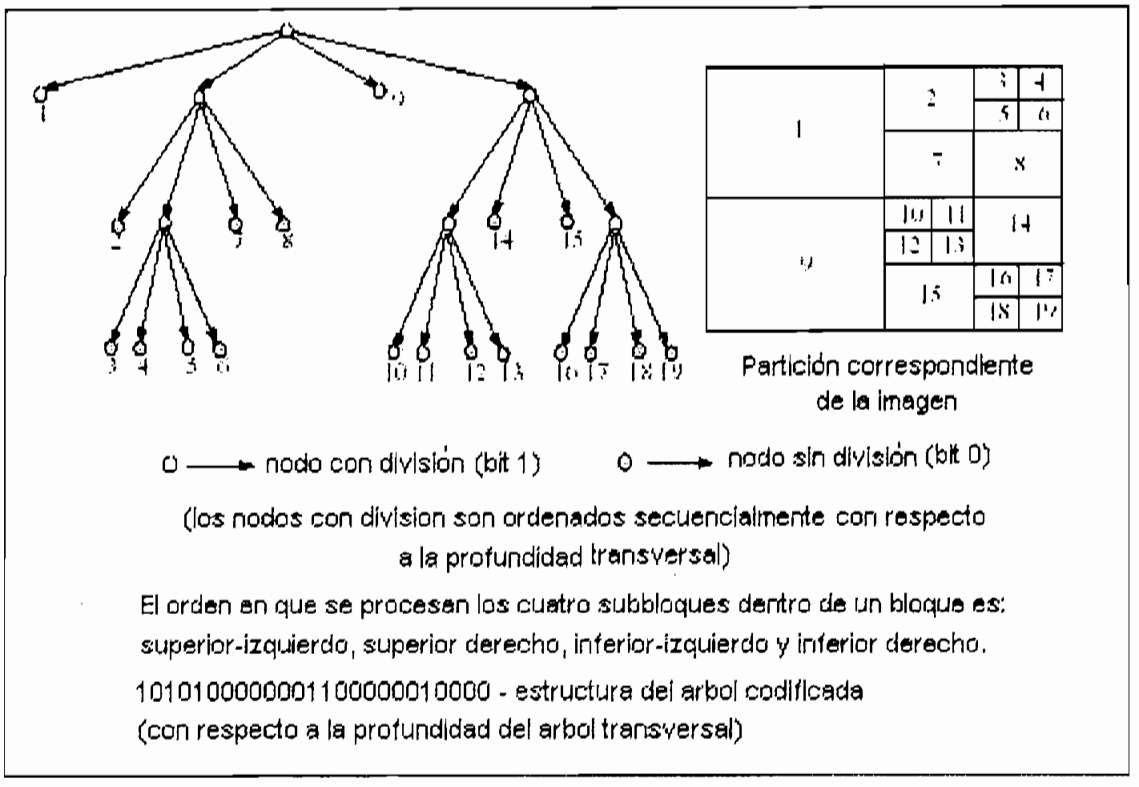


Fig 3.18 Descomposición de un quadtree general.

El objetivo entonces, es minimizar el número total de bits requeridos para codificar la estructura quadtree y las disparidades de bloque. El número de nodos divisibles puede ser minimizado alineando la localización de particiones con disparidades discontinuas. Sin embargo, codificar las localizaciones de las particiones arbitrarias horizontal y vertical dentro de un bloque requiere de $log_2(tamaño\ del\ bloque)$ [bits/nodo]. En lugar de siempre crear cuatro subbloques, el número de nodos sin división puede ser reducido considerando divisiones horizontales (H) y verticales (V) independientemente. Esto habría exigido sin embargo que se requiera de 2 bits/nodo para codificar los cuatro casos, denominados:

- únicamente H
- únicamente V
- H y V
- ni H ni V.

Un bloque puede ser dividido horizontalmente y verticalmente en 2^k-1 divisiones que están uniformemente separadas. Donde k es el número de bits asignados por dirección por nodo para detallar las divisiones particionadas. La división tiene lugar en la localización permitida que queda cerca de una forma de discontinuidad de disparidad. Puesto que las discontinuidades de disparidad no están disponibles antes de la segmentación, la intensidad de los bordes que normalmente constituye un excelente juego de discontinuidades de disparidades son usados. En la figura 3.19 se ilustra el procedimiento de descomposición del quadtree generalizado.

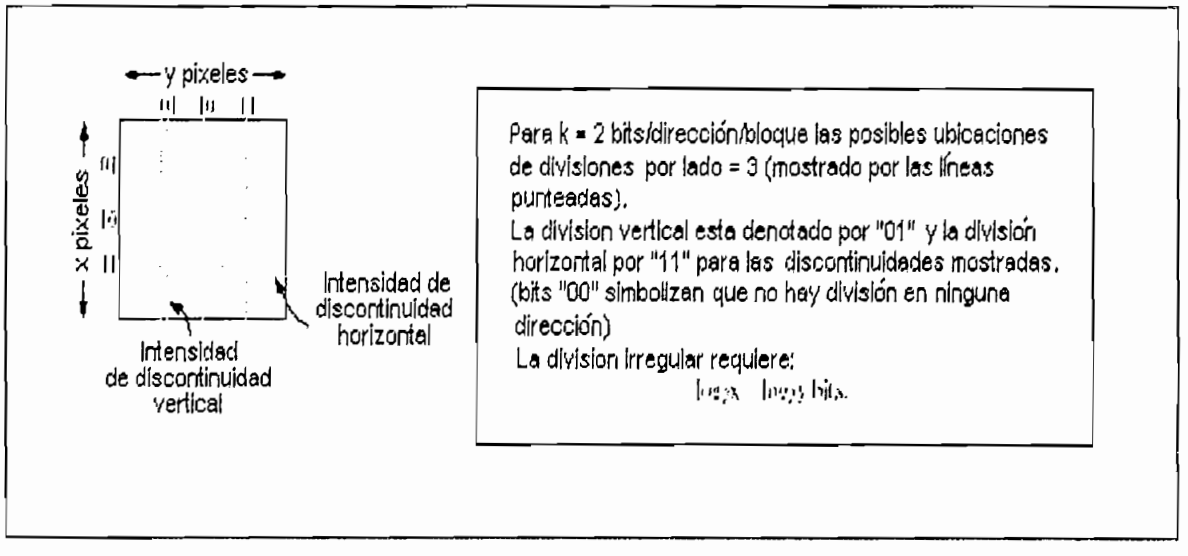


Figura 3.19 Descomposición generalizada quadtree – ubicaciones particionadas para k = 2.

La descomposición regular corresponde a (k=0) y descomposición irregular corresponde a (k= longitud/ancho de los bloques[21]).

Una multiresolución basada en descomposición quadtree procede de la resolución más baja a los niveles de resolución más finos. Los nodos sin división en una resolución llega a ser los nodos raíz en el próximo nivel de resolución. Este única estructura multiresolución para descomposición quadtree simplifica la complejidad de la descomposición y también ayuda a minimizar la elevada codificación. Por

[21] En los niveles de resolución respectivos las unidades se expresan en píxeles.

ejemplo en la parte superior de el árbol, si la compensación de disparidad se realiza a la resolución original, la búsqueda tiene que ser dirigida para tamaños de bloques que están cerca del tamaño de la propia imagen, mientras que con la estructura de multiresolución, la estimación se realiza a un nivel de resolución más bajo. Empleando diferentes valores de k en los diferentes niveles de resolución, la elevada segmentación y la codificación total de disparidad de bits necesaria puede ser minimizadas conjuntamente. Por ejemplo, pueden escogerse valores de k más grandes en las resoluciones más bajas ya que ahí serán pocos bloques inicialmente. La partición regular puede usarse para resoluciones más finas, ya que el número de bloques es alto y el error debido a las particiones fijas es pequeño a estas resoluciones debido a los tamaños de los bloques más pequeños. Puesto que las disparidades pueden ser codificadas de manera diferente en el árbol, el número requerido de codificación de disparidades de bits también se reduce.

3.2.9.3 Cálculo de las ubicaciones particionadas.

El objetivo primario de una descomposición irregular es alinear el límite del bloque con el límite de la característica que es usada en el criterio de subdivisión. En este caso el límite es la discontinuidad de disparidad. La discontinuidad de disparidad surge de un límite típico de objeto falso en una discontinuidad de intensidad de imagen (borde). En ausencia de un mapa de disparidad (que es lo que se intenta estimar), los bordes en una imagen proporcionan una localización para las particiones. La detección de bordes convencionales requiere convolución del bloque con dos operadores de gradiente (tal como un operador de Sobel) en direcciones ortogonales. La gradiente de intensidad de cada píxel es entonces umbralizada para obtener un mapa de borde. La convolución 2D con los operadores es computacionalmente costosa. También se necesita sólo los bordes horizontales y verticales dominantes dentro de un bloque. Así se usa un algoritmo de localizaciones de borde dominante vertical y horizontal. Para un bloque de tamaño w X h que empieza en la ubicación (x,y) en imagen I, la fila y columna promedio se calculan como:

$$m_{fila}(i) = \sum_{j=x}^{x+w-1} I_{(i,j)} \qquad \text{(Ec.3.12)}$$

$$m_{columna}(j) = \sum_{i=y}^{y+h-1} I_{(i,j)} \qquad \text{(Ec. 3.13)}$$

Estos promedios nos proveen con dos señales 1-D. El efecto de los detalles locales y el ruido son promediados fuera y los bordes dominantes a lo largo de las direcciones horizontal y vertical llegan a ser enfatizadas en las columnas y filas promediadas. Un filtro pasa altos de diferencia simétrica se aplica a la fila y columna promediados. Encontrando los valores picos sobre los valores absolutos de las salidas de los filtros, las localizaciones divididas horizontal y vertical se calculan como sigue:

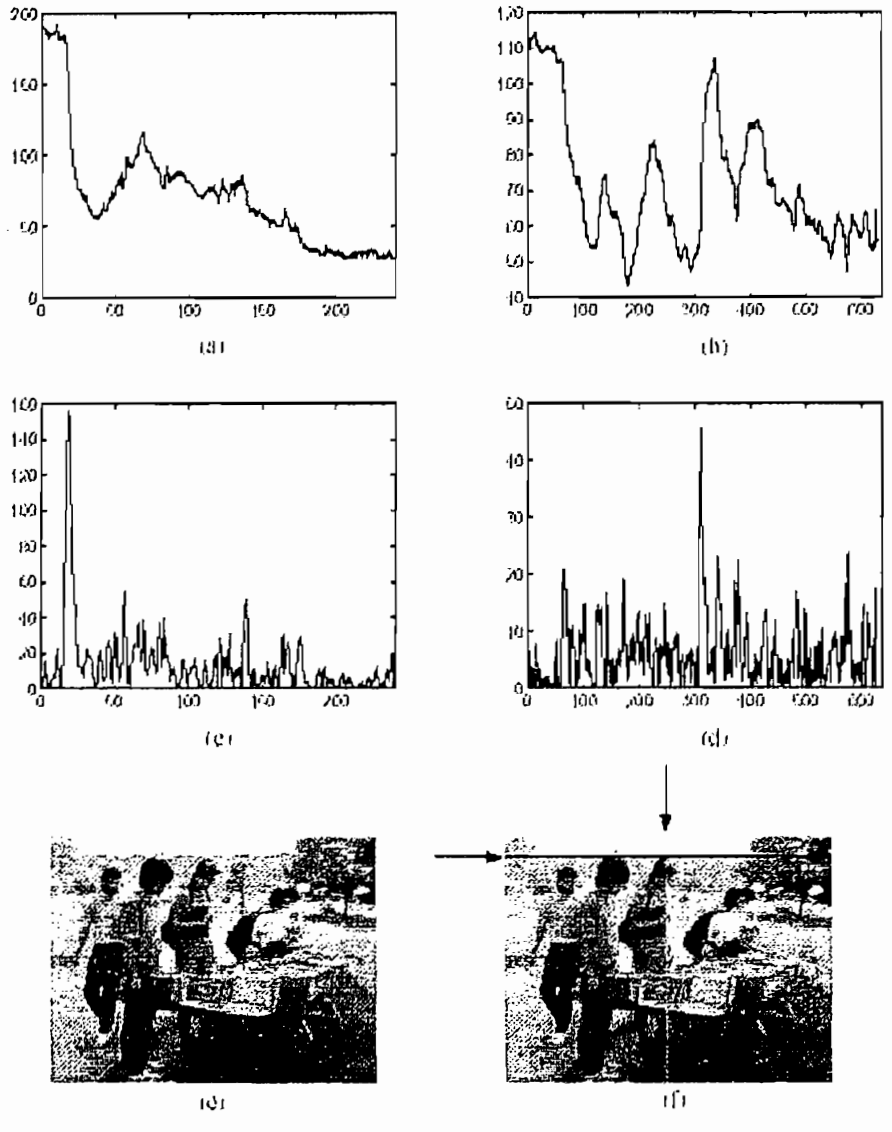
Para x:

$$Horizontal: l_h = \underset{t}{Max}\left(g_h(t) = \left|\left(m_{fila} \otimes f\right)\right| - n; (0 \le t < h+2n)\right) \qquad \text{(Ec. 3.14)}$$

Para y:

$$Vertical: l_v = \underset{j}{Max}\left(g_v(j) = \left|\left(m_{columna} \otimes f\right)\right| - n; (0 \le j < w+2n)\right) \qquad \text{(Ec. 3.15)}$$

Donde el operador ⊗ representa la convolución discreta y f un filtro de diferencia simétrico de longitud $(2n+1)$. Los filtros que generalmente se usan son de orden n = 1 y n = 2 (específicamente (-1, 0, 1) y (-1, -2, 0, 2, 1)). Un número grande de n proporciona una localización fiable del borde, por alisamiento de la salida de las variaciones locales pero reduce el número de las posibles localizaciones de particionamiento debido a efectos del borde. Este procedimiento se ilustra para un test de imagen en la figura 3.20. Puede verse que se obtienen buenas posibles ubicaciones que son alineadas con discontinuidades de intensidad y obtenidas usando un procedimiento simple de cálculo.



(a) Promedios de la fila, (b) Promedios de la columna, (c) Valores absolutos del filtro pasa alto de los promedios de la fila, (d) Valores absolutos del filtro pasa alto de los promedios de la columna, (e) Imagen usada, (f) Imagen con particiones horizontal y vertical que corresponden al máximo en (c) y (d). (Las fuertes discontinuidades de intensidad horizontales y verticales sobre el bloque). La diferencia simétrica que usó el filtro pasa alto es (-1 -2 0 2 1).

Figura 3.20 Ilustración del cálculo de las ubicaciones particionadas.

3.2.9.4 Codificación de segmentación superior.

La segmentación superior es considerable para descomposiciones irregulares, por la imposición de límites máximos y mínimos en las dimensiones de un bloque y operando dentro de la estructura de multiresolución, la segmentación puede ser reducida considerablemente.

Si el ancho y la altura de un bloque son *w* y *h* respectivamente, entonces la codificación de localizaciones de una partición arbitraria vertical y horizontal requiere $log_2(w \cdot h)$ bits. Puesto que las dimensiones de los bloques decrecen progresivamente como los procedimientos de descomposición del quadtree, el número de bits requeridos para codificar las localizaciones decrece logarítmicamente.

Si *n* niveles de descomposición multiresolución dyadic son empleados, entonces el tamaño de la subimagen en la resolución más baja es 4^{-n} veces el tamaño de la imagen real. De esta manera la codificación de bits de las particiones localizadas en un bloque que tiene dimensiones *w* x *h* de una resolución completa, sólo requiere $(log_2(w \cdot h) - 2n)$ bits de un nivel de resolución *n*. Usando un valor de *k* grande en resoluciones bajas (donde el hecho anterior puede ser explotado) y valores de *k* pequeños en resoluciones finas es descrito en la descomposición quadtree generalizada, la codificación superior puede ser reducida considerablemente.

Además si las dimensiones de bloques máximas y mínimas aceptables son descritas por S^{n}_{max} y S^{n}_{min} respectivamente, en el nivel de resolución *n*, entonces desde el nivel *(n-1)* hacia delante al límite superior para codificación superior por partición podría ser $log_2(2S^{k+1}_{max} - 2S^{k}_{min})$ en el nivel de resolución *k*. Los valores de S_{max} ayudan en el límite superior de la complejidad que necesita ser manejada por un elemento del proceso en una implementación paralela al proceso. Los valores S_{min} preveen la formación de bloques sumamente pequeños y también consideran la inexactitud de los bordes de un bloque mientras se usa los filtros de diferencia simétrica. Ya que el mismo número de bits podría necesitarse para

codificar la localización particionada arbitraria independientemente de si un bloque es dividido o no, la estructura del árbol es codificada en dos niveles separados. En el primer nivel, 2 bits superiores por nodo son usados para especificar si un nodo no fue dividido, o fue dividido, horizontalmente, verticalmente, o ambas. Como los tamaños de los bloques pueden ser calculados en el decodificador, esta superioridad puede hacerse en un rango de 0 a 2 bits dependiendo de que el ancho o la altura de un bloque sea tan pequeño o tan grande que S_{min}. Las localizaciones particionadas son codificadas en un segundo nivel, solamente en la dirección en la que la división ocurre.

Las ventajas de la descomposición irregular sobre la descomposición regular se ilustran en la figura 3.21 por un prueba de imagen sintética. La segmentación superior es casi la misma para ambas descomposiciones. Sin embargo la disparidad de cada nodo sin división tiene que ser codificada, entonces la partición irregular podría quedar fuera de la descomposición regular. Las ecuaciones generales que describen el modelo para codificación de segmentación superior son desarrolladas posteriormente.

3.2.9.5 Algoritmo de segmentación basado en disparidad.

Dentro de la estructura multiresolución diferentes criterios de subdivisión pueden ser utilizados en diferentes niveles de resolución. Para obtener una razonable segmentación inicial y para evitar realizar un emparejamiento de bloque con una gran cantidad de bloques en el comienzo, una homogeneidad espacial basada en descomposición se emplea en el nivel de resolución más bajo. La Homogeneidad espacial de un bloque es medida en términos de la variación de la intensidad dentro del bloque. En los subsecuentes niveles de resolución, la diferencia de disparidad entre subbloques hace el criterio de subdivisión. Los pasos del algoritmo son descritos a continuación:

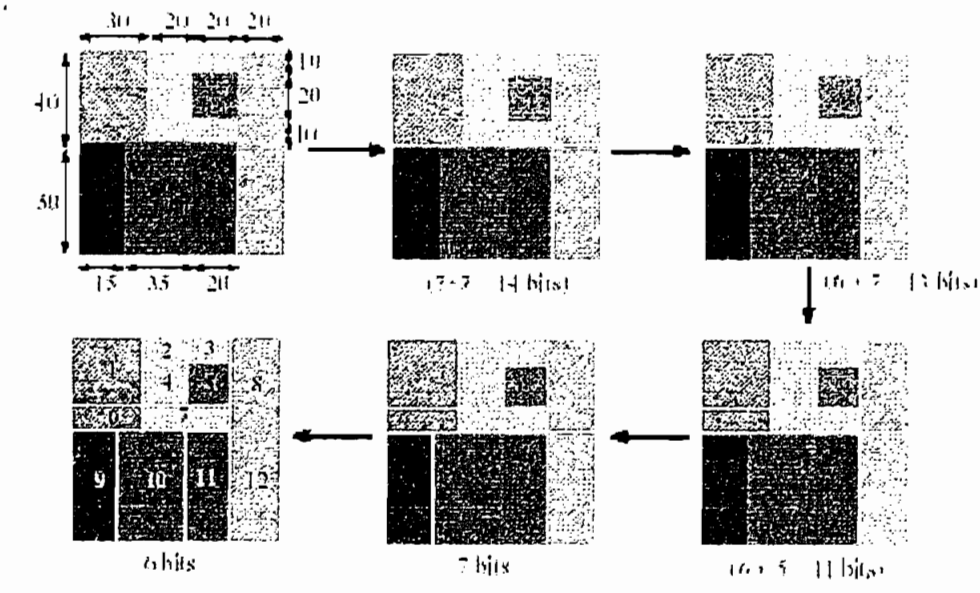
1. Construir las pirámides multiresolución izquierda y derecha mediante filtros pasa bajos recursivos y entonces submuestrear empleando el método de la figura 3.15.

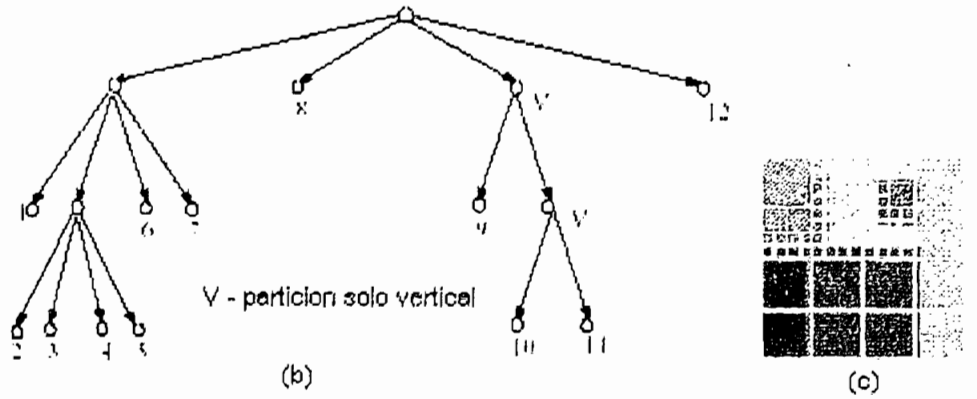
2. Empezar en el nivel de resolución mas bajo con la subimagen entera como un bloque. Fijar un umbral en la máxima varianza (T_{max}) permitida dentro del bloque. Fijar las dimensiones máximo y mínimo del bloque permitido (S_{max} y S_{min}) en la actual resolución.

3. Recursivamente, para cada bloque de altura h y ancho w:
   Si (($h < S_{min}$) y ($w < S_{min}$)), entonces se declara al bloque como un nodo sin división.
   Caso contrario,
   a. Calcular la varianza (var) del bloque.
   b. Si ($var < T_{max}$) y ($h < S_{max}$) y ($w < S_{max}$), se declara al bloque como un nodo sin división.
      Caso contrario, calcular la localización de los bordes dominantes horizontal y vertical (l_h y l_v píxeles respectivamente de la esquina superior izquierda en el bloque) como se discutió en la sección 3.4.3.
      Si (($h - l_h > S_{min}$) y ($l_h > S_{min}$)), dividir el bloque horizontalmente.
      Si (($w - l_v > S_{min}$) y ($l_v > S_{min}$)), dividir el bloque verticalmente.

4. Para los nodos sin división en la resolución, calcular la disparidad de bloque por emparejamiento de bloque con la correspondiente subimagen en la resolución de la otra vista. Si n niveles de descomposición dyadic son empleados entonces el rango de búsqueda en la resolución menor será 2^{-n} veces del rango de búsqueda deseado en la resolución más alta, en las direcciones horizontal y vertical.

5. Proceder al siguiente nivel de resolución más alto. Duplicar cada una de las dimensiones del bloque de los nodos sin división y las disparidades de bloque correspondientes. Fijar un umbral para la máxima diferencia absoluta permisible en las disparidades de bloque (D_{max} – generalmente pequeña) entre sub-bloques. Fijar las dimensiones del bloque máxima y mínima permisibles (S_{max} y S_{min}) en la resolución actual.





(a) Pasos de la descomposición – Número de bits necesarios para codificar las localizaciones particionadas es mostrada por cada paso. (b) Estructura correspondiente al quadtree. (c) Descomposición regular del quadtree por comparación. El número de pasos en (a) corresponde al número de nodos con división en (b).

Los nodos que no tienen división son numerados tomando en cuenta el orden de profundidad en (b) – los bloques correspondientes son mostrados en (a). Como todos los bordes son horizontales y verticales en la imagen sintética, las localizaciones particionadas calculadas usan el esquema indicado en la sección 3.4.3 en que se alinean exactamente con la intensidad de los bordes.

	DESCOMPOSICIÓN REGULAR	DESCOMPOSICIÓN IREGULAR
Número de nodos sin división	67	12
Codificación de bits de localizaciones particionadas	X	51
Codificación de bits de estructura del árbol	86	34

Si el movimiento o disparidad por cada nodo sin división necesita ser codificado, el equilibrio se inclina claramente a favor de la descomposición irregular.

Figura 3.21 Partición de un quadtree irregular de una imagen de prueba sintética.

6. Recursivamente para cada bloque de altura h y ancho w ;

a. Si ($h>S_{min}$), calcular la localización del borde horizontal dominante l_h.

Si (($h- l_h>S_{min}$) y ($l_h>S_{min}$)), permitir división horizontal.

Si ($w >S_{min}$), calcular la localización del borde horizontal dominante l_v.

Si(($w - l_v >S_{min}$) y ($l_v > S_{min}$)), permitir división vertical.

b. Para cada uno de los posibles subbloques en el paso a), calcular las disparidades de bloque[22]. El rango de búsqueda es independiente del nivel de resolución (se dice, ±2 píxeles alrededor de la estimación actual). Si el error absoluto medio (MAE) después de la compensación es un umbral predeterminado anteriormente, la actual estimación es ignorada y el emparejamiento de bloque es realizado nuevamente con el rango de búsqueda en el nivel l fijado a 2^{-l} ésimo rango de búsqueda en el nivel 0. Esto es hecho para prevenir la propagación de estimaciones erróneas a bajo de la pirámide.

c. Si (la diferencia entre las disparidades de sub-bloque > D_{max}) o ($h > S_{max}$) o ($w > S_{max}$), dividir el bloque en las localizaciones determinadas en el paso (a).

Caso contrario declare el bloque como un nodo sin división.

7. Si el actual nivel de resolución es el nivel de resolución mas alto, entonces calcular las disparidades exactas de medio píxel para los nodos sin división.

Caso contrario ir al paso 5.

---

[22] El borde dominante es ignorado durante la estimación de disparidad y se le asigna al subbloque con una mayor disparidad. Puesto que un borde en los límites del objeto corresponde a un objeto

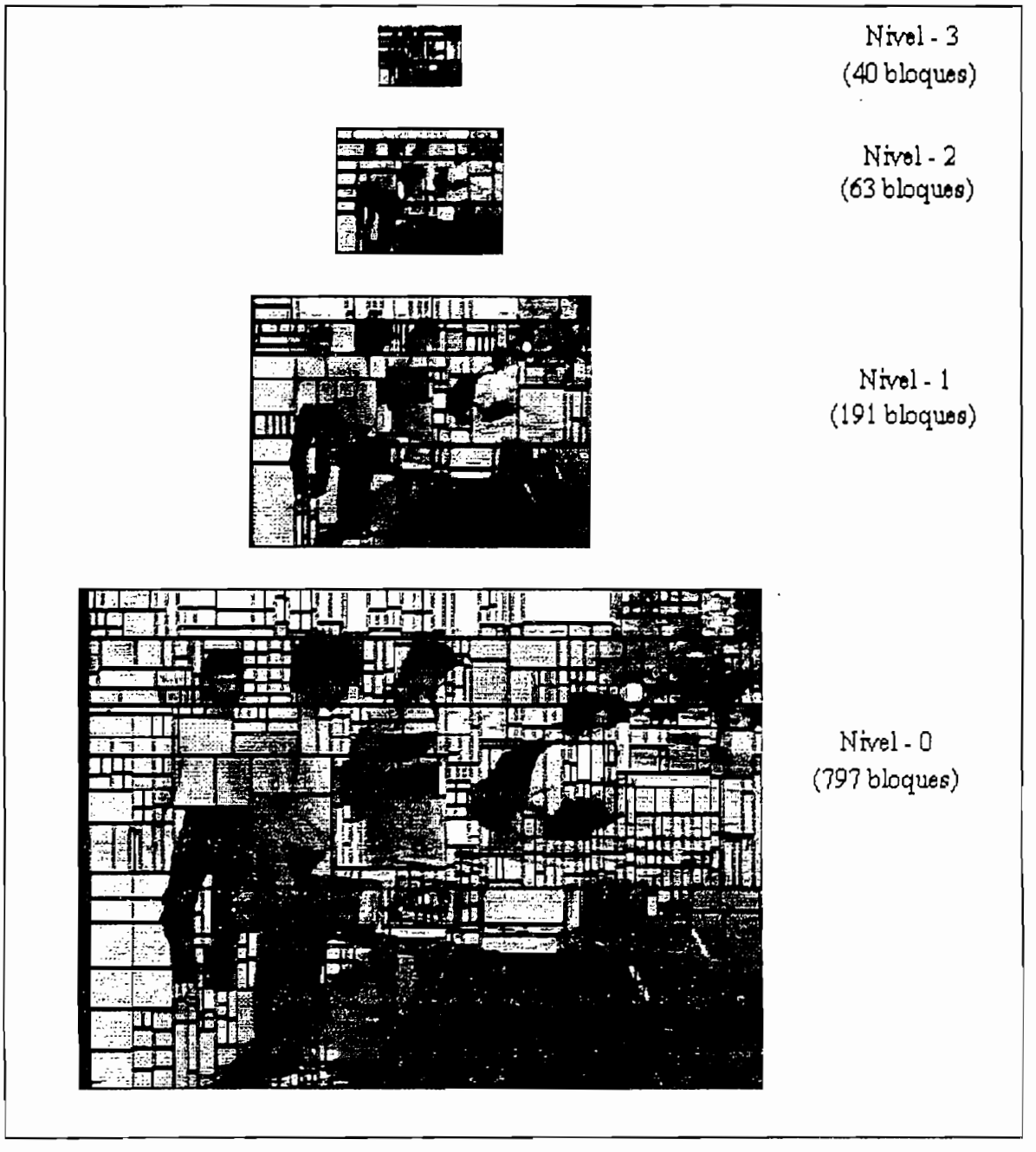


Figura 3.22 Ejemplo de algoritmo de segmentación basado en disparidad (aplicado a la imagen izquierda de un par estereoscópico de una secuencia de venta de libros)

en primer plano, el anterior paso evita que a este borde se le asigne erróneamente a un objeto en el fondo y mejora la exactitud de la estimación.

3.2.10 COMPRESIÓN DE SECUENCIAS ESTEREOSCÓPICAS.

En las secciones anteriores se considera el problema de comprimir pares de imágenes estereoscópicas y se afirma que la segmentación adaptiva de disparidad usando la descomposición de quadtree de multiresolución, es un método que ofrece un considerable incremento en la eficiencia de la codificación para la predicción compensada en disparidad.

En esta sección se extenderá la técnica de la segmentación anterior para encajar dentro de una estructura de codificación de secuencia, dirigiéndose a varios problemas críticos que afectan la compresión de secuencias estereoscópicas y se propondrán soluciones para:

1. Explotar las redundancias espacial (intraview) y temporal (interview) para incrementar la eficiencia de la codificación.
2. Ajustar el excesivo ancho de banda necesario para transmitir video estereoscópico a ser proporcionado con la demanda de video estereoscópico.
3. Explotar las propiedades del sistema visual humano específico para percepción estereoscópica y
4. Codificación conjunta de las secuencias para mejorar la escalabilidad de cálculo y eficiencias de codificación con múltiples vistas.

Anteriormente se describieron métodos típicos de compresión de secuencias de imágenes. Estos métodos explotan la redundancia espacial dentro de un cuadro, la redundancia temporal entre cuadros adyacentes y tolerancias del sistema visual humano para lograr proporciones de compresión muy altas. El método mas simple concebido de compresión de secuencia estereoscópica puede ser codificar cada una de las vistas usando tales métodos de compresión. En este caso, una secuencia de n vistas podría requerir de n veces la velocidad de bits necesaria para transmitir una secuencia simple. Para lograr una significante reducción del ancho de banda, comparada con tal codificación independiente, se necesita considerar varios factores adicionales tales como, la correlación del cruce de flujo

y factores psicofísicos asociados con la percepción estereoscópica. El problema se hace mas difícil debido a otras consideraciones prácticas, tales como la necesidad de un mapa de disparidad en el decodificador para sintetizar vistas intermedias (ver sección 3.1.1.2) sin utilizar un excesivo ancho de banda, un codificador moderado, los bajos requerimientos de complejidad del decodificador y la necesidad de una compatibilidad de calidad con los esquemas de transmisión monoscópicos existentes. En la siguiente subsección, se presenta una estructura de trama para codificación de secuencias estereoscópicas que permitirán explotar la correlación de flujo cruzado mientras retienen algunas de las características deseables de los métodos de compresión de secuencias monoscópicas.

3.2.10.1 Compresión de secuencias estereoscópicas para estructuras de cuadro.

La estructura de cuadro recomendada por el estándar de codificación de video MPEG tiene varios rasgos atractivos. La intra codificación independiente de cuadros I habilita el acceso aleatorio, editabilidad y decodificabilidad independiente de diferentes segmentos de una secuencia codificada. Los cuadro I y P sirven como referencia periódica de la que los cuadros intermedios B son predecidos. Para prevenir la acumulación de errores de predicción sobre el tiempo debido a la predicción progresivamente mas baja en calidad de cuadros, los cuadros I y P se codifican típicamente con una mayor calidad que los cuadros B. La eficiencia de codificación para los cuadros B es mejorada empleando la predicción bidireccional, aunque a costa de incrementar la carga computacional, puesto que las regiones ocluidas en un cuadro de referencia pueden predecirse de otro cuadro de referencia.

Comparada la codificación independiente de las secuencias multivistas, la compresión adicional puede lograrse aprovechando las redundancias temporal y espacial que existe. Supongamos que una de las secuencias es codificada independientemente, mientras las otras secuencias son codificadas con respecto a esta secuencia codificada independientemente. A esta codificación de escenario se la conoce como *codificación dependiente.*

Asumiendo una codificación MPEG como estructura de cuadro en cada una de las vistas, los cuadros I de estas otras vistas pueden ser predecidos usando compensación de disparidad con respecto a el cuadro I de la primera secuencia de codificación independiente. Consecuentemente, la intracoding típicamente ayuda a disminuir del 20 al 30 % de la velocidad de bits global, la mas significativa reducción de velocidad de bits podría venir de este paso. Además los cuadros P en estas vistas pueden ser predecidos bidireccionalmente con respecto a un cuadro pasado de referencia dentro de esta vista y con respecto al cuadro correspondiente en la secuencia codificada independientemente.

Ya que la correlación con el cuadro correspondiente en la otra vista es probablemente mayor que la correlación con el cuadro de referencia previa dentro de una vista (para una secuencia con una cámara y moderados movimientos de objetos y para una típica separación de cuadros P a P), este paso también. contribuirá a una reducción en la velocidad de bits. La reducción en la velocidad de bits puede también atribuírsele al hecho de que una región ocluida en el cuadro de referencia temporal puede ser predecida de la vista correspondiente (con tal de que esta no este ocluida también en perspectiva). Similarmente, los cuadros B pueden predecirse tridireccionalmente. Estos modos de predicción son ilustrados en la figura 3.23.

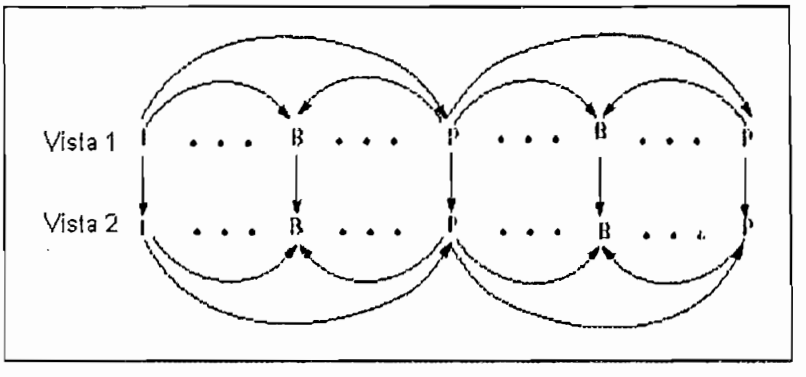


Figura 3.23 Codificación dependiente – modos de predicción para los diferentes cuadros (Se supone una estructura de cuadro MPEG)

3.2.10.2 Factores que influyen en los modos de predicción

En las sección anterior, no se considero específicamente la calidad de los cuadros de referencia. Sin embargo la demanda para video estereoscópico nunca podrá ser bastante alta para garantizar unas n veces o cercano a n veces en el incremento de ancho de banda en una aplicación de tipo broadcast. Puesto que es probable que la mayoría de los espectadores miren monoscopicamente en cualquier momento dado, por lo menos una secuencia dentro de las secuencias de múltiples vistas debe ser codificada con una alta calidad. Dicha secuencia es conocida como la secuencia principal. Las otras secuencias que se codifican con una calidad correspondiente con la demanda para el video estereoscópico y las ventajas funcionales que el video estereoscópico ofrece, serán referidas como secuencias auxiliares. Los cuadros en las secuencias auxiliares que corresponden a los cuadros I - P y B de las secuencia principal, se denotan como los cuadros IA – PA y BA respectivamente.

La diferencia en niveles de calidad entre los diferentes cuadros dentro de una secuencia y a través de vistas, tiene una considerable influencia en el modo de predicción particular que podría ser favorable durante la codificación de secuencias estereoscópicas. Por ejemplo, si la secuencia auxiliar es codificada con una calidad significativamente mas baja que la secuencia principal, entonces la predicción compensada de disparidad sería favorecida sobre las secuencias con predicción compensada en movimiento para los cuadros PA y BA. Similarmente, como los cuadros B en una secuencia son codificados a una calidad mas baja que los cuadros I y P, si la secuencia auxiliar es codificada en una proporción similar al de la secuencia principal, entonces la predicción compensada en movimiento podría ser favorable sobre la predicción compensada en disparidad para los cuadros BA.

Aunque la calidad reducida de codificación de cuadros auxiliares ha sido considerada, el exceso de ancho de banda es escogido arbitrariamente. Así mismo el impacto de la calidad del cuadro de referencia no ha sido dirigido por otras investigaciones. La elección entre la DCP y MCP para codificar un cuadro auxiliar también depende de los siguientes factores:

1. Movimiento Intercuadro (magnitud, componentes rotacionales y cambios de escala).
2. La magnitud de la disparidad o la distancia de los objetos a las cámaras.
3. Disparidad que es un escalar (como oposición a ser un vector de dos componentes como el movimiento) para unos ejes paralelos a la geometría de la imagen.
4. El emparejamiento entre las cámaras izquierda y derecha en términos de brillo, contraste y balance de color, y
5. La necesidad para sintetizar vistas intermedias en el decodificador.

3.2.10.3 Configuraciones para compresión de secuencias estereoscópicas.

Mientras la mayoría de los factores anteriormente descritos influencian la elección de DCP vs. MCP como una base por bloque, la necesidad de un completo mapa de disparidad en el decodificador evita la posibilidad del uso por completo de MCP. Por esta razón se consideran dos configuraciones básicas, configuración-1 y configuración-2, para codificar el flujo auxiliar.

En la configuración-1, los cuadros de la secuencia auxiliar se estiman a través de DCP. Aquí un mapa de disparidad completo estará disponible en el decodificador para cada cuadro. Regiones subcompensadas debido a oclusiones y errores de DCP son compensados posteriormente a través de MCP con respecto a cuadros de referencia pasados y posteriores en la secuencia auxiliar.

En la configuración-2, los cuadros auxiliares son estimados a través de predicción bidireccional con respecto al correspondiente cuadro de secuencia principal y el cuadro de referencia mas cercano en la secuencia auxiliar. Esta configuración así tiene la capacidad para escoger adaptivamente entre DCP y MCP. Sin embargo, el decodificador ya no tiene mapa de disparidad completo y aquí la síntesis de vistas intermedias no es posible. Estas dos configuraciones básicas son ilustradas en la figura 3.24.

3.2.10.4 Codificador residual.

Anteriormente al considerar los esquemas de codificación de secuencias estereoscópicas, se describió brevemente un codificador residual.

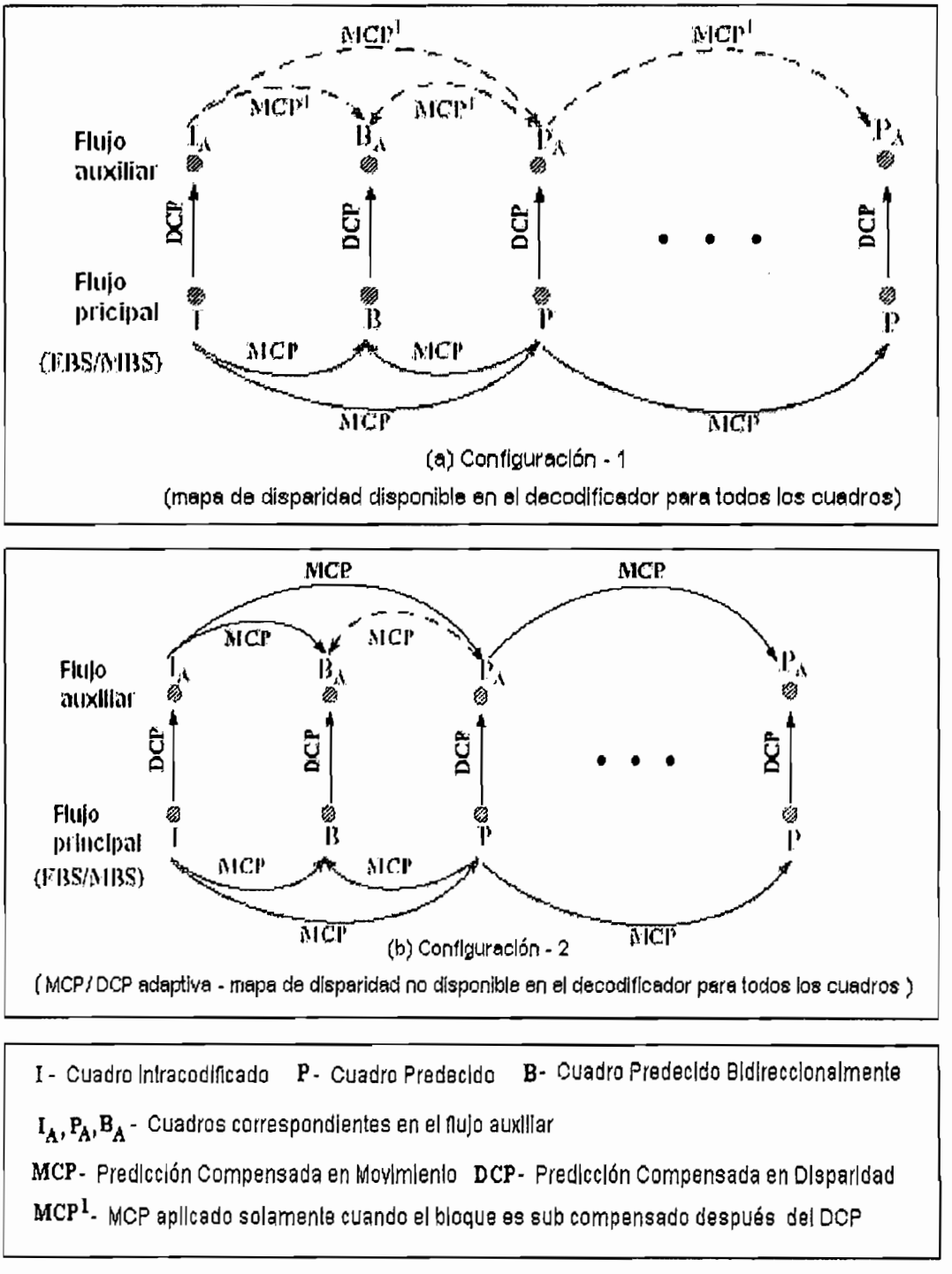


Figura 3.24 Compresión de secuencia estereoscópica – dos configuraciones básicas

Aunque la predicción compensada en disparidad o movimiento típicamente provee una aceptable compensación para la mayoría de regiones en un cuadro de imagen, errores significativos pueden estar presentes en algunas regiones debido a la falla de las suposiciones detrás del bloque basado en compensación, por ejemplo, fallas de desplazamiento traslacional o constante de disparidad sobre un bloque asumido y oclusión parcial de un bloque.

Residuos significantes, si se parte del decodificador, pueden producir degradación severa en la calidad percibida de una imagen y debido a las predicciones intercuadro, los errores aumentarán también con el tiempo. Sin embargo, debido a la alta entropía de los residuos, incluso su codificación lossy típicamente constituye una fracción significante del bit global presupuestado. El estándar MPEG recomienda un codificador residual basado en transformada coseno discreto. Sin embargo los residuos no contienen una estructura especial en el dominio de la transformada, lo cual puede ser explotado para codificarlos eficazmente. De hecho, si los residuos son esparcidos dentro de un bloque (lo cual es mas probable), el número de valores significantes diferentes de cero en el dominio de la transformada será mas alto que el número de residuos significantes en el dominio espacial. Debido al reducido número de bits presupuestado para la frecuencia auxiliar, necesitamos un codificador residual que pueda designar bits específicamente a regiones con errores significantes, así que la mayoría de errores perturbantes pueden ser codificados dentro de un limitado bit presupuestado.

La codificación residual selectiva requiere codificar las localizaciones de los residuos más significantes en adición a la codificación de los valores de los residuos.

Una combinación de cuantificación vectorial/escalar es usada para codificar los valores de error en los diferentes tamaños de bloque del quadtree. Cada cuadro residual es dividido en bloques de tamaño 16 x 16, denominado como macrobloques (como en el estándar MPEG). Dos medidas de distorsión son

usadas decidiendo si un bloque necesita ser codificado o no. Uno es el MAE definido como:

$$\sum_{k\in\eta}\left|I_{act}(k)-I_{est}(k)\right| \qquad \text{Ec. 3.16}$$

donde I_{act} es la imagen actual, I_{est} es la imagen estimada y η es el conjunto de todos los píxeles en el bloque.

El otro es la cuenta de error significante (N_T) definido como el número de píxeles para el cual,

$$\left|I_{act}(k)-I_{est}(k)\right| > T, (k\in\eta) \qquad \text{Ec. 3.17}$$

donde T es algún error significante pre especificado. Dos umbrales, conocidos como el máximo MAE aceptable (E_{max}) y la máxima cuenta de error significativo aceptable (N_{max}), típicamente 0 o 1, se especifica para cada cuadro. Si (MAE > E_{max}) o ($N_T > N_{max}$) para un bloque, entonces ese bloque es considerado para la codificación residual.

El tamaño de macrobloque es escogido como 16 x 16 para guardar la profundidad del menor quadtree y para habilitar un cierto grado de paralelismo. También para imágenes típicas, un tamaño de bloque mas grande tiene una probabilidad mas alta de contener errores significantes. El quadtree basado en algoritmos de codificación residual VQ / SQ para cada macrobloque es resumido en la tabla 3.2.

Los codebooks son generados usando el algoritmo LBG. El vector dimensional 16 del codebook se obtiene por entrenamiento sobre un conjunto de vectores de código residual derivado de secuencias típicas. Un subconjunto de vectores de entrenamiento con errores en el rango de (-32 , 32) en los niveles de gris son escogidos para entrenamiento real y se relegan vectores con errores mas grandes a los niveles subsecuentes del quadtree. La entropía de cada código

vector sobre el conjunto de entrenamiento es usado para asignar un código de longitud variable (VLC) a ese vector código. El codebook del vector de 4 dimensiones se obtiene de manera similar con un rango mas grande para los residuos. Los niveles de cuantización escalar se diseñan para la distribución Laplaciana de los errores obtenidos de las ejecuciones del codificador residual incorporando los dos vectores de cuantización (VQ's) anteriores.

La codificación de estructura superior quadtree (1 bit por nodo) y los códigos de longitud variables de los estados del vector de cuantización y cuantizador escalar constituyen la codificación residual superior para un macrobloque.

Tamaño del Bloque	Paso 1 : Si (MAE > E_{max}) o (N_T>N_{max})	Paso 2 : Si (MAE >E_{max}) o (N_T>N_{max})
16 x 16	Divide en 4 sub bloques de 8 x 8	
8 x 8	Ejecuta MCP con respecto a a un cuadro de referencia, si es necesario. Si el MAE resultante es menor que el 70% del MAE previo, codifica el vector de movimiento. Calcula N_T.	Divide en 4 sub bloques de 4 x 4
4 x 4	Calcula el error cuadrado medio (MSE) con respecto a los vectores código de un codebook vector dimensión 16. Escoge el vector código que produce el menor MSE. Calcula el nuevo MAE y N_T.	Divide en 4 sub bloques de 2 x 2. Codifica el VLC correspondientemente al mejor emparejamiento del vector código.
2 x 2	Calcula MSE con respecto a los vectores código de un vector codebook de una dimensión 4. Escoge el vector código que produce el menor MSE. Calcula nuevo MAE y N_T.	Divide en cuatro píxeles simples. Codifica el VLC correspondientemente al mejor emparejamiento del vector código.
1 x 1	Estima el nivel de cuatización mas cercano en un cuantizador escalar. Codifica el VLC correspondiente a ese nivel.	

Tabla 3.2 Resumen de Quadtree y VQ / SQ basado en codificación residual

El control de este codificador residual se logra mediante dos medidas de calidad, denominadas, MAE y N_T (para un T específico). Estas medidas solamente aseguran calidad constante; el control de velocidad preciso no es posible. Sin embargo es posible conseguir una velocidad de bits cercana a la deseada adaptablemente poniendo los umbrales en el codificador, basado en el conocimiento de la velocidad de los bits de cuadros previamente codificados. Los umbrales para las medidas de calidad pueden aumentarse con un umbral para la proporción entre la variación del error dentro de un bloque y la variación de la intensidad (o actividad espacial) dentro de ese bloque. Comparando umbrales basados en solo el error estático, semejante umbral aprovecha los efectos enmascarados de acuerdo con el inherente sistema visual humano para designar los bits de codificación residual. Por ejemplo, una particular variación del error que es aceptable en un bloque con una alta actividad espacial puede ser inaceptable en bloques homogéneos.

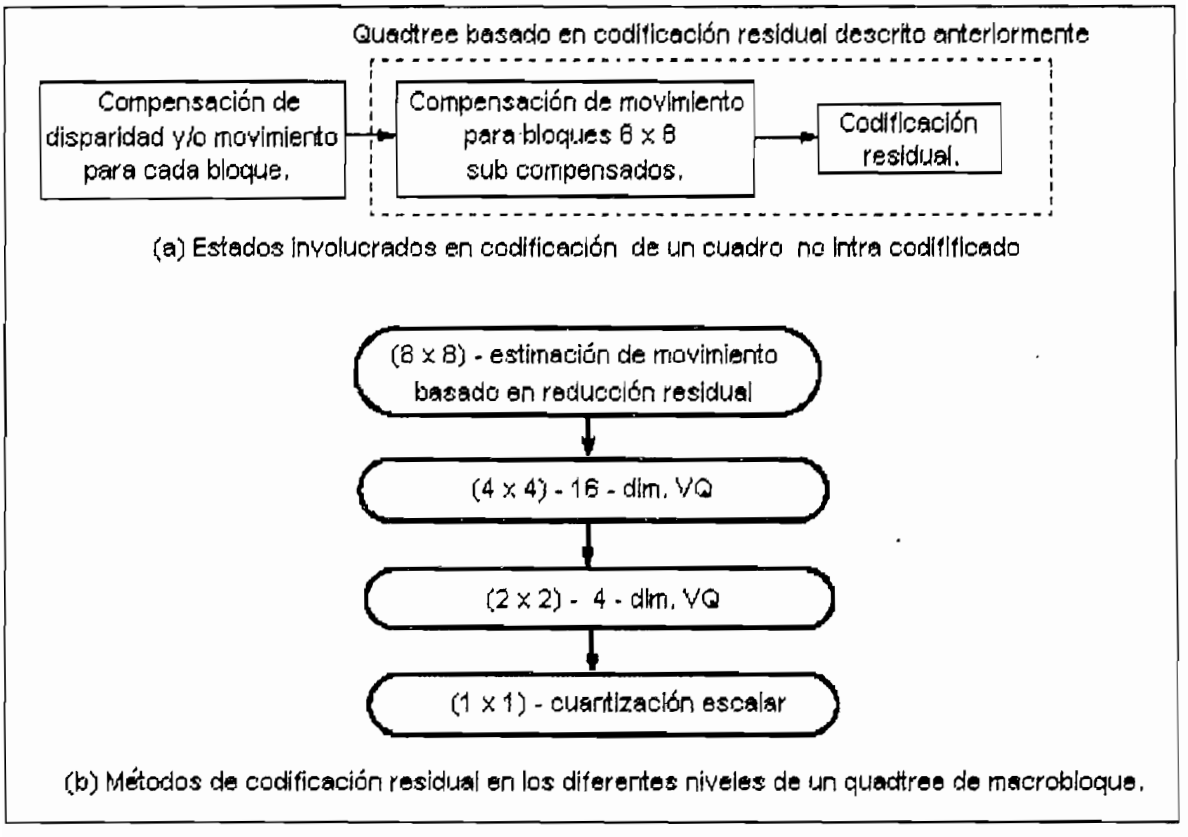


Figura 3.25 Quadtree y VQ / SQ basado en codificación residual.

3.2.10.5 Esquemas básicos.

Inicialmente se introducieron dos esquemas de compresión de secuencias estereoscópicas, uno para cada configuración. Los esquemas básicos emplean tamaño de bloque fijo basado en compensación de disparidad y movimiento (como se muestra en la figura 3.24 (a) y (b)) y son representativas de las normas MPEG). Estos esquemas básicos llamados esquemas FBS-1 y FBS-2 para denotar el uso de bloques de tamaño fijo y las configuraciones de codificación, son usadas para delinear los detalles detrás de la codificación de secuencias estereoscópicas; estos también sirven como referencia contra las cuales las extensiones basadas en MR-QTD son presentadas en las secciones mas tarde. Las secuencias de entrada a los esquemas de codificación están en formato 4:2:0. La secuencia principal es codificada independientemente a través de un tamaño de bloque fijo basado en predicción compensada en movimiento con una estructura MPEG como estructura de cuadro.

Los componentes Y, U y V de los cuadros I son codificados usando DCT basado en intracodificacion de bloques de 8x8, descrito en la sección 3.2.3 y mostrado en la figura 3.8. La tablas de Huffman de las recomendaciones MPEG-2 son usadas para ejecutar la longitud del código de los coeficientes de cuantificación DCT después del escaneo en zig-zag.

El emparejamiento de bloque jerárquico como se describió en la sección 3.2.7.5 y figura 3.16 es empleado para MCP y DCP. Puesto que es difícil lograr una configuración de la cámara absolutamente paralela, un pequeño rango de búsqueda se permite en la dirección vertical durante la compensación de disparidad. Los residuos son codificados seteando los parámetros umbrales de E_{max} y N_{max} para el codificador residual (descrito en la última sección).

3.2.10.6 Multiresolución con base en descomposición quadtree basados en extensiones de codificación dependientes.

En esta sección se consideran dos extensiones mas simples que incorporan el acercamiento de descomposición de múltiples resoluciones basadas en quadtree dentro de dos configuraciones básicas.

3.2.10.6.1 Extensión -1 (DBS-1)

El esquema básico FBS-1 (extensión de codificación de secuencias estereoscópicas usando FBS-BMA - configuración 1) puede extenderse de una manera directa para incorporar la MR-QTD como una aproximación por reemplazo del tamaño de bloque fijo basado en compensación de disparidad con el algoritmo DBS desarrollado. Haciendo referencia a esta extensión como DBS-1. Todos los resultados que se aplicaron para una codificación de cuadro simple usando DBS podría aplicarse a codificar los cuadros de secuencia auxiliar.

3.2.10.6.2 Extensión –2 (DBS-2)

El algoritmo DBS (**S**egmentación **B**asado en **D**isparidad) sólo es aplicable para predicción compensada en disparidad. Las diferentes partes de un objeto a una profundidad particular (de la cámara) pueden sufrir diferentes desplazamientos con el tiempo, por ejemplo un objeto que rueda sobre un eje paralelo al eje de la cámara. El esquema FBS-2 involucra una predicción bidireccional que usa compensación en movimiento y disparidad. Para incorporar el método MR-QTD dentro de este esquema, el algoritmo DBS tiene que ser extendido para incluir movimiento también basado en segmentación. Esto es hecho estimando ambos movimientos (con respecto al cuadro de referencia en la secuencia auxiliar) y la disparidad para cada segmento en el algoritmo DBS de la sección 3.2.5. El criterio de particionamiento en el paso 6(c) del algoritmo se modifica como sigue:

> Si ((la diferencia entre las disparidades de sub bloque > D_{max}) y (la diferencia entre los componentes de desplazamiento de sub bloque > M_{max})) o ($w>S_{max}$),

entonces se divide el bloque en la ubicaciones predeterminadas (donde M_{max} máxima diferencia absoluta aceptable en un componente de desplazamiento entre sub bloques).

Puesto que un buen emparejamiento es necesario en cualquiera de los dos cuadros de referencia, un bloque es dividido solamente si los movimientos del sub bloque así como las disparidades del sub bloque son diferentes. De esta segmentación típicamente resulta en menos segmentos que con DBS-1. Se hace referencia a esta codificación de secuencias estereoscópicas como DBS-2.

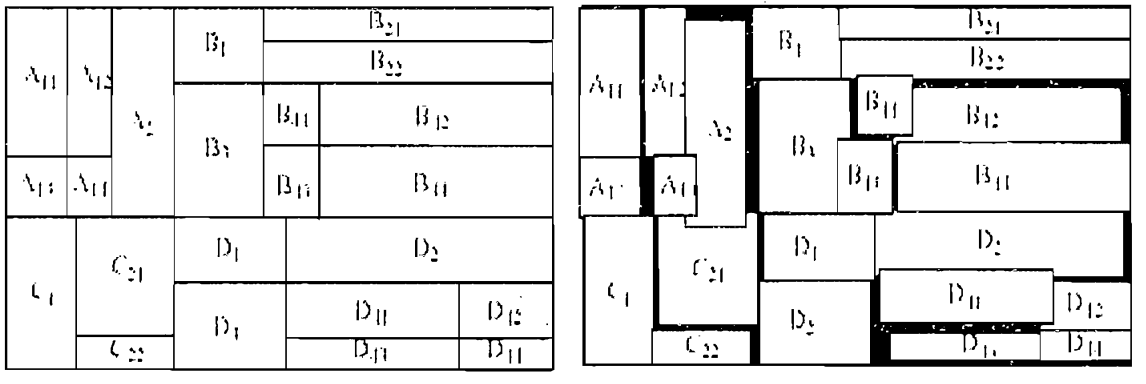
3.2.11 MR-QTD BASADO EN EXTENSIONES DE CODIFICACIÓN CONJUNTA

3.2.11.1 Inversión de dirección de la predicción.

La segmentación usando MR-QTD una levada codificación así como un elevado cálculo. Las extensiones DBS-1 y DBS-2 exigen segmentar cada cuadro. También, la secuencia principal en estas extensiones son codificadas independientemente usando un tamaño de bloque fijo basado en predicción compensada en movimiento. Esta secuencia también puede ser codificada usando segmentación adaptable de movimiento. Tal segmentación adicional incrementaría la carga computacional. Ahora esto sería preferible si la misma segmentación pudiera ser usada para codificar varios cuadros a lo largo de la dimensión de visión o a lo largo de la dimensión temporal, así que el elevado cálculo y la segmentación de codificación elevada pueden ser compartidos por todos estos cuadros. Sin embargo, la representación basada en quadtree es una representación espacial donde un juego de objetos encajan dentro de uno mayor y no puede ser usada cuando sus nodos sin división están sufriendo desplazamientos espaciales independientes. Esto evita la posibilidad de usar la misma representación quadtree para todos los cuadros mientras realizan movimiento o estimación de disparidad en dirección hacia delante. Compartir segmentación superior entonces requiere una inversión en la dirección de la predicción. En otras palabras, los segmentos en un cuadro pueden rastrearse a otros cuadros. Esto constituye un cambio significante en el paradigma comparado

con la estimación convencional. En la estimación convencional, el cuadro a ser codificado se divide en bloques no solapados y la mejor juntura para cada uno de estos bloques es buscada en los cuadros de referencia. En este caso, alguna predicción razonable (no necesariamente significante) se obtiene para todos los bloques. Sin embargo, la inversión de la dirección de la predicción resulta en un cuadro predecido con algunas regiones que no tiene ninguna predicción (agujeros) y algunas regiones que tiene múltiples predicciones. Esto se ilustra usando la figura 3.26.

Como los objetos dentro de la escena sufren desplazamiento, nuevas regiones pueden exponerse y pueden ocluirse regiones actualmente expuestas. Si un segmento en el cuadro – A es ocluido parcialmente en el cuadro – B (en la figura 3.26), entonces la mejor unión para ese segmento puede ocurrir en la localización correcta, o un falso emparejamiento puede ser generado, dependiendo de la magnitud de la oclusión y la existencia de oportunidad de buenos emparejamientos. Cuando el emparejamiento ocurre en la posición correcta, la región ocluida tiene dos posibilidades de emparejamiento – una corresponde a la región ocluida y la otra corresponde a la región ocluyente. por ejemplo, en el cuadro – B, una porción del segmento B41 ocluye a la porción de segmento B22.



Cuadro segmentado - A (particion no solapada) Cuadro predecido por rastreo de bloques - B

Los segmentos correspondientes en B solapan debido a las oclusiones de movimiento y errores en la estimación de movimiento.

Nuevas regiones (mostradas en negro) que fueron ocluidas en A son expuestas en B

Figura 3.26 Impacto en la inversión de la dirección de la predicción.

La región común entre estos dos segmentos tiene dos posibles elecciones de emparejamiento. Cuando un falso emparejamiento ocurre, el segmento correspondiente deja atrás una región vacía y también se agrega como un candidato estimado para la localización de un falso emparejamiento. Las regiones descubiertas mientras se rastrean los segmentos, por definición, no tienen predicciones.

Para codificar el cuadro bajo consideración, se necesita:
- Escoger el correcto emparejamiento entre múltiples emparejamientos y
- Obtener predicciones convenientes para las regiones expuestas.

Cuando las estimaciones de disparidades fiables para los segmentos en el cuadro – A están disponibles, se puede usar el orden de profundidad proporcionado por estas disparidades estimadas (es decir el hecho de que un segmento que esta mas lejano no pueda ocluir otro segmento que esta mas cercano a la cámara) para quitar la ambigüedad entre múltiples emparejamientos. Las regiones sin predicción pueden ser intra codificadas. Pero debido a la situación arbitraria y a las formas irregulares de estas regiones, la intra codificación podría ser alta. La interpolación basada en el relleno de estas regiones podría resultar en un perdida de la calidad.

3.2.11.2 Esquema RDBS

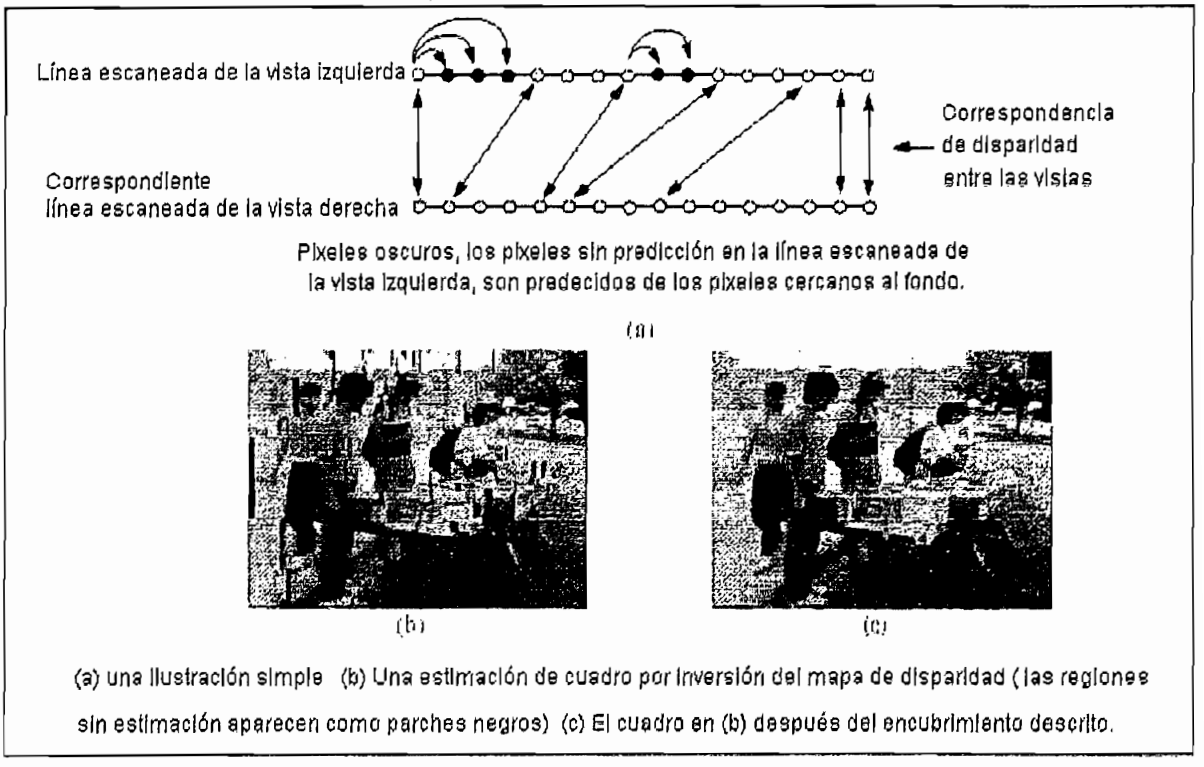
El esquema RDBS es el esquema de extensión en la codificación de secuencias estereoscópicas usando segmentación basada en disparidad inversa – configuración – 1. El que considera un esquema de codificación conjunta en el cual la secuencia principal es también codificada usando segmentación adaptiva de movimiento. Cada cuadro en la secuencia principal es segmentado usando el algoritmo DBS (segmentación basada en disparidad). Así el principal flujo de cuadros no tiene particiones solapadas. La compensación de movimiento para los cuadros P y B se llevan a cabo sobre estos bloques de tamaño variable. Para considerar desplazamientos independientes de subbloques dentro de un bloque, estos bloques son particionados teniendo en cuenta el error después de la

compensación de movimiento como el criterio de fraccionamiento. El mapa de disparidad, calculado durante la segmentación, se invierte para predecir la secuencia auxiliar de cuadros del flujo de cuadros principal. Así cada par estereoscópico de cuadros comparten la codificación superior de segmento. Las particiones no solapadas en el flujo principal de cuadro, solapan en la secuencia predecida de cuadros auxiliar, surgiendo agujeros en los lugares que corresponden a regiones ocluidas y regiones con errores de estimación de disparidad. Múltiples posibilidades de emparejamiento durante la inversión son verificados usando la disparidad. Sin embargo, el costo de codificar lo agujeros (regiones donde ninguna predicción es disponible) puede compensar parcialmente la ganancia en velocidad de bits conseguido a través de la codificación conjunta.

Predicción espacial para regiones no cubiertas

Dado que los pares estereoscópicos de cuadros son capturados al mismo tiempo, el mapa de disparidad depende solamente de las profundidades de los diferentes objetos en ese momento y la geometría de la cámara binocular fijada. Así , las oclusiones debido al paralaje binocular son mas estructuradas que las oclusiones basadas en movimiento las cuales dependen de los desplazamientos de los diferentes objetos en la escena.

Un algoritmo de escaneo de línea para llenar los agujeros puede ser desarrollado, dado que los ejes de las cámaras son paralelos. Asumiendo que una región no cubierta es parte de un objeto que esta en una profundidad mayor que el objeto que se expuso en esa región, una predicción espacial para las regiones no cubiertas puede ser formulada. Operando a lo largo de las líneas escaneadas y usando el mapa de disparidad estimado, la dirección (izquierda o derecha) del objeto del fondo cerca de una región expuesta puede encontrarse. El valor de la intensidad en las cercanías del píxel en el fondo sirve en la predicción para todos los píxeles expuestos en una línea escaneada. Semejante predicción unidireccional asegura que una interpolación erronea no se lleve a cabo sobre dos regiones con diferentes disparidades.



(a) una ilustración simple (b) Una estimación de cuadro por inversión del mapa de disparidad (las regiones sin estimación aparecen como parches negros) (c) El cuadro en (b) después del encubrimiento descrito.

Figura 3.27 Predicción espacial para regiones no cubiertas durante la inversión de la dirección de predicción

Para escenas típicas, el llenado de los agujeros, en valor esta cerca al valor de intensidad actual para la mayoría de píxeles en la mayoria de estos. No se incurre en ninguna codificación elevada para semejante esquema de predicción. Sin embargo, como el decodificador también tiene que realizar la detección y predicción de agujeros, su complejidad se aumenta. Este esquema se ilustra en la figura 3.27 (a); la efectividad del método se muestra para un cuadro auxiliar de la secuencia de booksale en (b) y (c).

Una estimación de movimiento con exactitud de medio píxel se lleva a cabo para cada uno de los bloques rastreados en los cuadros de referencia. Al contrario de la estimación típica de una sola dirección donde la exactitud de medio píxel puede ser codificado usando un bit adicional para cada dirección, en este caso, se necesitan dos bits por dirección para codificar las tres posibilidades de – ½ , 0 y + ½ de desplazamientos de píxeles. Después de rellenar las regiones expuestas, los residuos son codificados usando un codificador residual. Ya que el flujo de

cuadros auxiliares es obtenido invirtiendo la dirección de la predicción, se hace referencia a este esquema como RDBS (reversed DBS). Los diferentes modos de predicción son ilustrados en la figura 3.28. Este esquema pertenece a la configuración – 1 ya que el decodificador tiene un mapa de disparidad completo para cada cuadro.

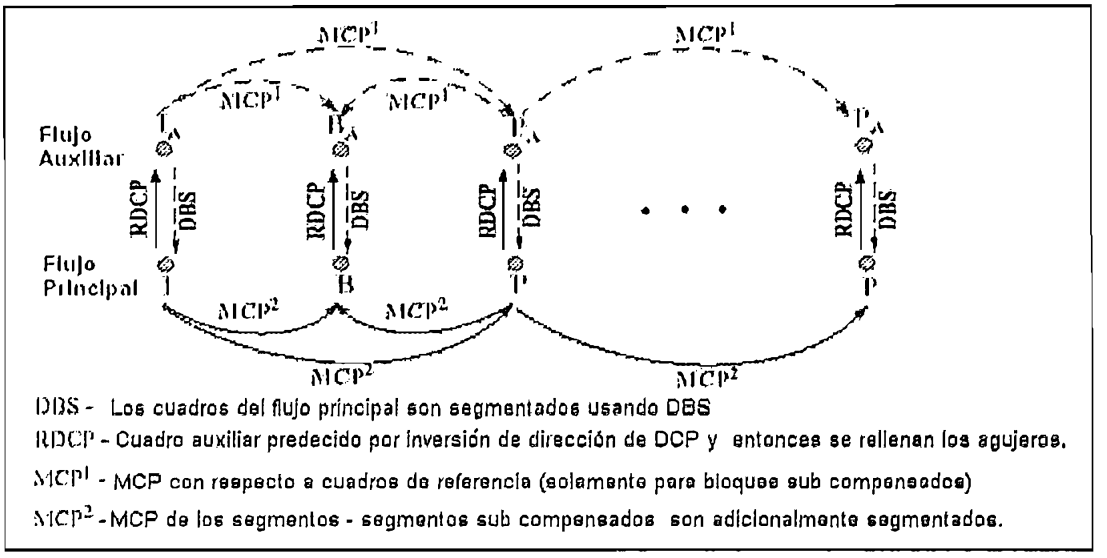


Figura 3.28 esquema RDBS – configuración 1

3.2.11.3 Rastreo de segmento (ST-1)

En el esquema RDBS, la segmentación tiene que ser repetida para cada par de cuadros estereoscópicos. El cálculo y la elevada codificación asociada con la segmentación pueden reducirse más allá si un grupo de pares estereoscópicos de cuadros comparten la misma segmentación. Esto puede lograrse segmentando un cuadro de referencia y rastreando los segmentos en ambos flujos al próximo cuadro de referencia. Desde que la segmentación se lleva a cabo para la compensación de movimiento y disparidad, nuevamente una unión de movimiento y disparidad basada en segmentación (MDBS) es requerida como se dijo en la sección 3.2.10.6.2, pero con la siguiente modificación.

La segmentación adaptiva de movimiento se realiza con respecto al cuadro de referencia mas cercano dentro de la secuencia y el criterio para dividir un bloque (paso 6 (c) de la sección 3.2.9.5) es:

Si (la diferencia entre las disparidades del sub bloque > D_{max}) o (la diferencia entre los componentes del desplazamiento de sub bloque > M_{max})) o ($h>S_{max}$) o ($w>S_{max}$), entonces divida el bloque en las ubicaciones predeterminadas (donde M_{max} es la diferencia absoluta máxima permitida en un componente de desplazamiento entre sub bloques).

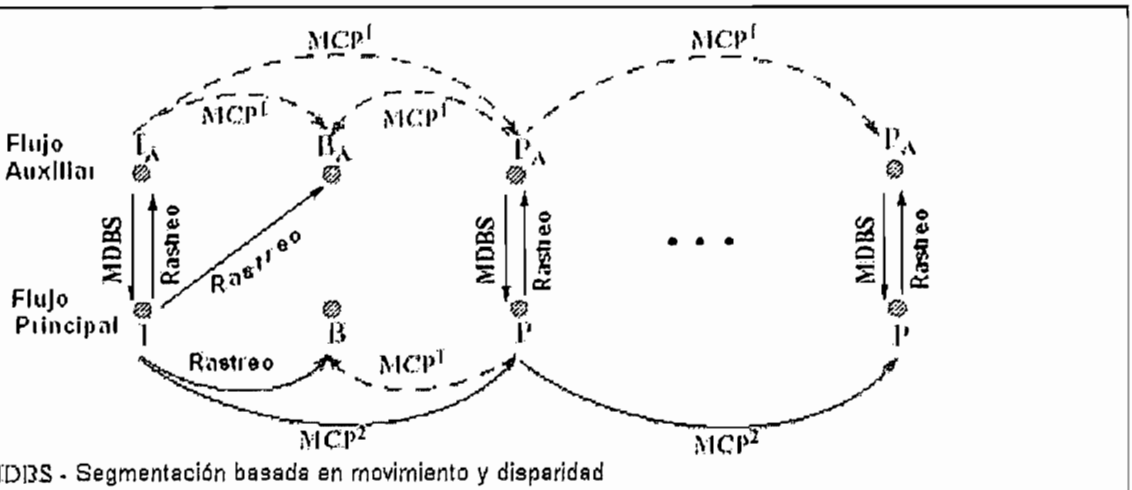
Tal segmentación típicamente resulta en mas segmentos que con el algoritmo DBS, requiriendo de una buena compensación ambos dominios, temporal y de la perspectiva. Los cuadros de referencia del flujo principal son segmentados usando MDBS. Los flujos principales de cuadros –B son estimados rastreando los segmentos del cuadro de referencia y entonces invirtiendo la dirección de la predicción. Puesto que el mismo segmento se rastrea, la estimación de movimiento de la segmentación con una escala apropiada puede ser usada como estimaciones iniciales para emparejar el bloque. Los cuadros del flujo auxiliar pueden ser estimados de dos maneras usando la ecuación de coherencia siguiente:

$$v_m + \delta_t = v_a + \delta_{t+k} \qquad \text{Ec. 3.18}$$

donde v_m es el vector de movimiento del flujo principal de un segmento entre los cuadros en los instantes t y (t+k), v_a es el vector de movimiento del flujo auxiliar entre los cuadros en los instantes t y (t+k), δ_t es la disparidad izquierda-derecha en el instante t, y δ_{t+k} es la disparidad izquierda-derecha en el instante (t+k). El cuadro auxiliar correspondiente al cuadro segmentado puede estimarse por la inversión del mapa de disparidad obtenido durante MDBS. Los otros cuadros del flujo auxiliar se estiman mediante DCP. Para cada segmento en el cuadro (t+k) - esimo del flujo principal, se encuentra un buen emparejamiento en el correspondiente cuadro auxiliar. Entonces la dirección de predicción se invierte para estimar el cuadro auxiliar. Para un pequeño k, δ_t puede ser usado como una buena estimación inicial para δ_{t+k}. Ya que el mapa de disparidad para cada cuadro esta disponible en el decodificador, este caso se considera bajo la

configuración – 1 y se referirá a este esquema como ST-1 (rastreo de segmento - configuración 1). La estructura del cuadro se muestra en la figura 3.29. Una extensión similar también puede realizarse mediante el uso de compensación de movimiento para predecir los cuadros BA.

Puesto que todos los cuadros B y cuadros BA son estimados por la inversión de la dirección de la predicción, estos cuadros tendrán regiones sin las predicciones y las múltiples posibilidades de predicción en los solapamientos.



MDBS - Segmentación basada en movimiento y disparidad

Rastreo - Los segmentos en el cuadro de referencia son rastreados y los vectores de movimiento/disparidad son invertidos para predecir el cuadro. Los agujeros son entonces llenados.

MCP^1 - MCP con respecto a cuadros de referencia (para bloques sub compensados)

MCP^2 - MCP de los segmentos

Figura 3.29 Esquema de rastreo de segmento ST-1 – configuración 1

Los múltiples emparejamientos pueden nuevamente ser resueltos basados en las estimaciones de disparidad. Sin embargo, el procedimiento de relleno no es simple como en RDBS. Esto es porque los cuadros ahora son compensados en tiempo y aquí una simple predicción 1D a lo largo de las líneas escaneadas no es posible. Además, desde que la secuencia principal tiene que ser codificada a una calidad mas alta, el simple encubrimiento no es suficiente. El incremento en una codificación residual elevada a una alta velocidad de bits puede mas que compensar las ventajas ganadas por la distribución de una elevada segmentación sobre un grupo de cuadros. Sin embargo, computacionalmente este esquema es bastante atractivo. Esto es porque la frecuencia de segmentación es

significativamente reducida y las complejidades de compensación de disparidad y movimiento son también considerablemente reducidas usando adecuadamente escaladas las pasadas estimaciones como estimaciones iniciales. Lo último es posible solamente porque el mismo segmento es rastreado sobre el tiempo a través de las vistas.

Además, los cuadros B y cuadros BA no necesitan ser descompuestos en múltiples resoluciones como los refinamientos de las estimaciones iniciales y pueden llevarse a cabo al nivel de resolución mas fino. Así este esquema será ideal en situaciones donde una muy alta calidad de flujo principal no sea requerida, o en casos donde se desea una muy baja complejidad de cálculo.

Una predicción con precisión de medio píxel en el cuadro de referencia se obtiene para cada uno de los segmentos rastreados. Como en RDBS, se necesitan dos bits por dirección para representar esta estimación exacta de medio píxel. Los agujeros en este caso se rellenan primero extrayendo sus localizaciones y entonces realizando MCP. Esto se hace para aprovechar el hecho de que los agujeros típicamente son bastante largos en una dirección, así que solamente unos pocos vectores de movimiento son necesarios. Además, si un orden particular es empleado en la extracción de bloques, el decodificador puede repetir ese orden sin ninguna ambigüedad y así ninguna localidad incurre en una codificación elevada. Los residuos en los cuadros auxiliares son codificados usando el quadtree basado en la combinación VQ/SQ, con estimación de movimiento bidireccional en el bloque de tamaño 8x8 explota las redundancias temporales que no fueron explotadas durante el rastreo del segmento.

3.2.12 RESOLUCIÓN MIXTA BASADA EN CODIFICACIÓN

Codificación de baja velocidad de bit es deseada para la secuencia auxiliar para reducir el exceso de ancho de banda. Esto restringe el número de bits que pueden asignarse para la codificación residual. Los residuos significantes que son izquierdos no codificados pueden resultar en artefactos de distracción visual. Los

artefactos notable pueden ser suprimidos cambiandolos fuera de la resolucion y codificando los cuadros del flujo auxiliar en un resolución reducida. Los estudios psicofísicos han mostrado que la satisfacción de la percepción estereoscópica se logra cuando una de las secuencias estereoscópicas es presentada a un observador con una resolución reducida. Basados en experimentos psicofísicos con estereogramas de punto aleatorio, se ha reportado que esta estereopsis puede ocurrir aún cuando las similitudes espaciales solo existan en una banda de frecuencia particular. Basado en un experimento donde una imagen delineada se presento al ojo derecho y una imagen significativamente nublada fue presentada al otro ojo, se reporta que el par de imagen estereoscópica es fácil de fundir y la percepción binocular aparece no solamente en profundidad sino tambien parece tan detallada como la imagen bien delineada. La mezcla de resolución basada en la codificación de imágenes estereoscópicas describe que cada bloque de 4x4 en una vista es promediado para obtener un píxel en la resolución reducida. Durante el despliegue, una interpolación bilineal es aplicada para estirar el tamaño. El submuestreo y el sobremuestreo se hacen asi en una manera conveniente para cada objeto, sin cualquier consideración sobre el aliasing o la calidad de reconstrucción. Una pirámide Gaussiana (como la vista en la sección 3.2.7.1) basada en submuestreo y sobre muestreo es usada para resolución reducida de codificación. Desde el empleo de una estructura de múltiple resolución para segmentación y estimación de movimiento/disparidad, la mezcla de resolución basada en codificación automaticamente se ajusta en esta estructura. La estimación de múltiples resoluciones de movimiento o disparidad necesita ser llevada fuera solamente sobre la resolución deseada. La figura 3.30 muestra las modificaciones necesarias en el codificador y decodificador para codificación de resolución mezclada, con la secuencia auxiliar que es codificada a la mitad de la resolución horizontal actual. Ya que la elevada codificación residual es más pequeña en una resolución reducida que en la resolución original, los bits disponibles para la codificación residual pueden usarse para suprimir artefactos significantes. También, como el cuadro intracodificado típicamente descarta los componentes de frecuencia mas altos, la pérdida de información en comparación a la codificación de resolución completa puede esperarse que sea pequeña. Sin embargo, la reducción en la resolución horizontal puede producir una reducción

en la resolución del plano de profundidad o "agudeza estereo".

Para evitar esto, se emplea una estimación de disparidad con exactitud de subpíxel a la resolución reducida que es equivalente a una estimación de disparidad con exactitud de medio píxel en la resolución original.

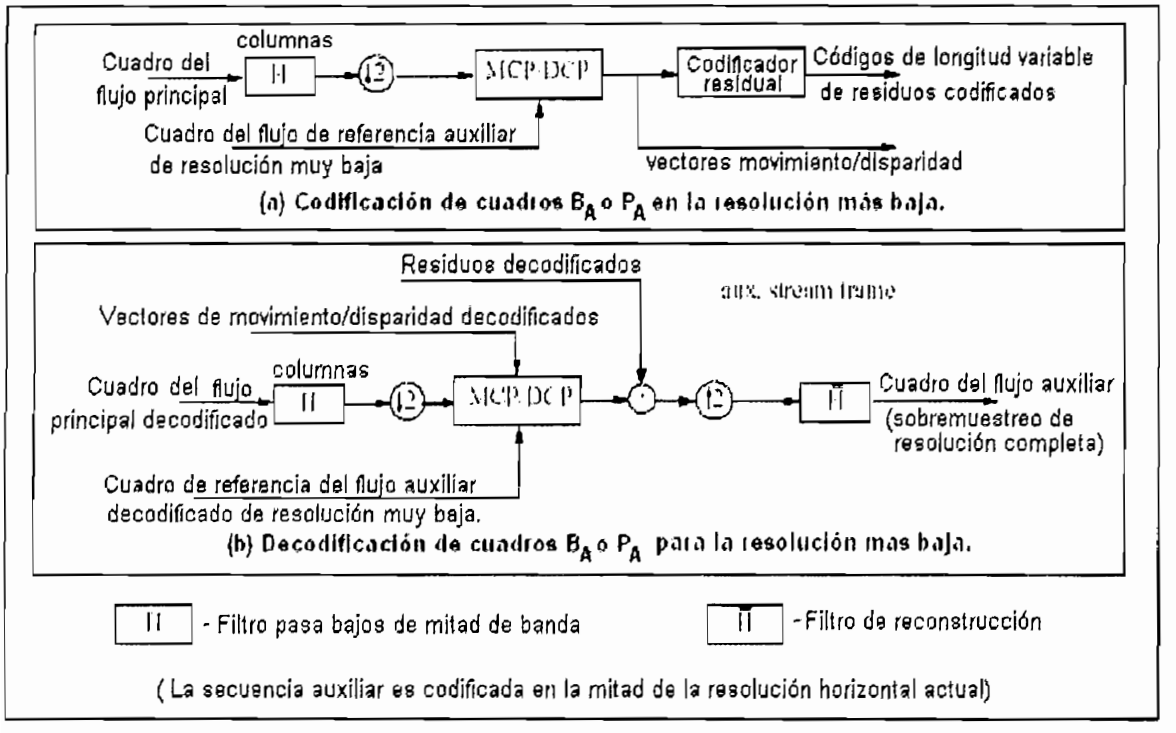


Figura 3.30 Esquema de codificación basado en mezcla de resolución.

Desde que los filtros no son ideales y los componentes de alta frecuencia se pierden, la reconstrucción puede contener componentes de alta frecuencia con aliasing (si la imagen original tuviera energía significante en las altas frecuencias). La complejidad del decodificador aumenta debido a la necesidad para el filtrado en el submestreo y sobremuestreo. Sin embargo, como mencionamos en la sección 3.2.7.6, conseguir bancos de filtros de multiples proporciones de escalabilidad espacial y temporal es en general deseable en decodificadores. El recurso del hardware disponible para este propósito puede usarse para el codificador de resolución mixta. Así, la codificación de resolución mixta proporciona un método para cambiar resolución por percepción de calidad en una manera controlada, la cual puede ser un factor significante haciendo práctica la transmisión de video estereoscópico.

3.3 DESPLIEGUE DE LAS IMÁGENES

En el despliegue de la imágenes estereoscópicas, el objetivo principal es que cada ojo vea la imagen que le corresponde, es decir, el ojo derecho debe ver la imagen derecha y el ojo izquierdo la imagen izquierda, para conseguir este objetivo como ya se explicó en el capítulo 1, existen dos sistemas:

a. El que utiliza algún dispositivo especial o visor sobre los ojos y
b. El que prescinde de dispositivos o visores especiales.

En ésta sección se dará especial atención al sistema que no utiliza dispositivos o visores especiales, a este tipo de dispositivos se les conoce como monitores o displays autoestereoscópicos, los cuales han tenido gran aceptación y demanda en aplicaciones donde dispositivos montados sobre la cabeza o gafas estereoscópicas son inaceptables debido a que reducen la visibilidad ya sea del ambiente circundante o de la cara del usuario.
Se debe mencionar que existen varios formatos para estereoplexar[23] una imagen estereoscópica sobre pantallas o displays electrónicos.

3.3.1 FORMATO ESTEREOSCÓPICO DE VISIÓN

Un *formato estereoscópico de visión* es la técnica usada para asignar píxeles (líneas o campos) para las imágenes izquierda y derecha, permitiendo de esta manera obtener en la pantalla o display, una imagen con verdadera estereopsis binocular. Entre los formatos más importantes se tienen:

- *Campo secuencial.-* esta técnica es también conocida como de campo alternado o de multiplexación de tiempo y consiste en mostrar alternadamente los campos izquierdo y derecho como lo indica la figura 3.31. Los campos pueden ser de barrido entrelazado o progresivo.

[23] **Estereoplexar.-** se refiere a la multiplexación de pares estereoscópicos para conseguir la estereopsis visual.

Los productos que utilizan el formato de campo secuencial en el mercado son principalmente las gafas shutters.

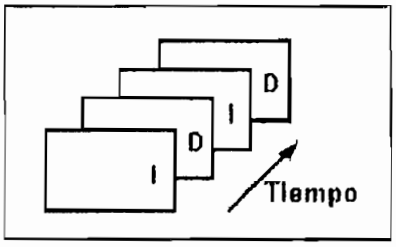


Figura 3.31 Formato de campo secuencial

Se debe tener en cuenta que el término "entrelazado" se lo ha estado utilizando mal para describir un despliegue multiplexado en tiempo, esto se verá mas en detalle al explicar el formato de despliegue estereoscópico entrelazado.

- *Entrelazado estéreo.*- es conocido como el formato original y básico de televisión estereoscópica, este aprovecha la estructura de entrelazado par e impar para poner en la pantalla las imágenes izquierda y derecha en campos alternados. Este es un método que aun en la actualidad se utiliza y que tiene la ventaja de usar los estándares de televisión convencional y equipo de demultiplexación de bajo costo. De hecho, el corazón del sistema es un interruptor simple que desvía la mitad de los campos a un ojo y la otra mitad al otro ojo.

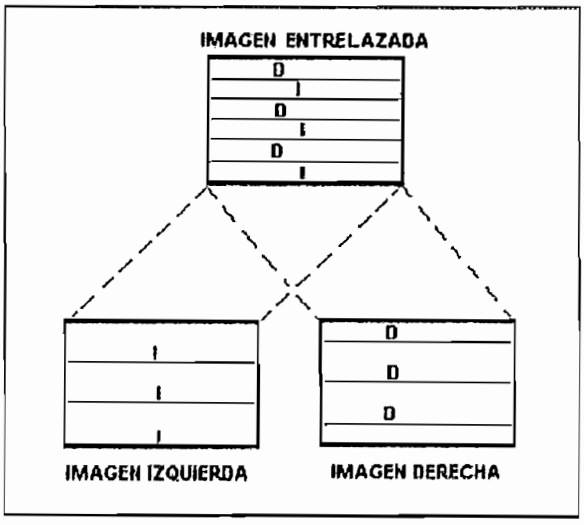


Figura 3.32 Separación de la imagen entrelazada para obtener un estéreo par con vistas independientes izquierda y derecha

Debido a la baja velocidad de despliegue de cuadros o campos, el método ocasiona parpadeo de la imagen. Otro problema es que debido a que cada ojo ve solamente la mitad del número de líneas disponibles normalmente, la resolución disminuye a la mitad. Este formato es utilizado en sistemas que utilizan HMD con displays LCD. La figura 3.32 muestra el esquema de un estéreo par con independencia de los canales izquierdo y derecho .

- *Segmento o línea secuencial.*- es conocido como una variante del entrelazado estéreo en el cual se despliegan primero todas las líneas impares correspondientes a la imagen izquierda, para posteriormente desplegar todas las líneas pares correspondientes a la imagen derecha, este sistema utiliza gafas con una alta velocidad de despliegue en los displays LCD y obturadores que permiten ver solo la imagen cuando se hayan desplegado en su totalidad los segmentos. La figura 3.33 muestra una imagen instantánea de la segmentación secuencial con el último segmento de la vista izquierda encima y debajo todos los segmentos correspondientes a la vista derecha, el origen de este campo es secuencial.

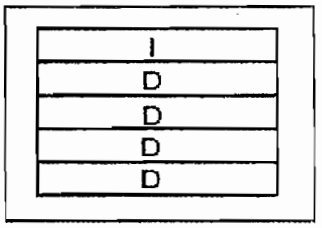


Figura 3.33 Formato de despliegue de segmento secuencial.

El formato de línea secuencial es interesante debido a que de esta manera se podría conseguir una imagen libre de parpadeo.

- *Imágenes interdigitales.*- también conocidas como de píxel secuencial, este formato se lo utiliza para aplicaciones estereoscópicas y auto estereoscópicas. El método de ínter digitalización estereoscópica utiliza el formato ínter lazado pero con una interesante técnica de selección

diferente. Al contrario de la multiplexación por división de tiempo que usa ínter lazado para visualización a través de gafas o HMD utilizando un display doble LCD estereoscópico, el método de ínter digitalización utiliza ínter lazado entre bits correspondientes a las vistas izquierda y derecha. Este sistema utiliza un panel LCD con un pedazo de matriz denominada micropol, compuesta de píxeles o tiras anchas de elementos polarizados en yuxtaposición con filas alternas de píxeles LC (cristal líquido). El panel LC, debido a la localización fija de sus píxeles garantiza una buena yuxtaposición con los campos par e impar y las tiras polarizadas asociadas, el gran retardo de la imagen a lo largo del LCD ha sido usado para suprimir el efecto de parpadeo que podría verse en displays con pequeños elementos de despliegue de imagen. Esta técnica es utilizada para proyectar ambas vistas y en una forma directa.

Otro tipo de ínter digitalización de imágenes se obtiene con columnas verticales en lugar de filas horizontales. Estas columnas, típicamente con las imágenes izquierda y derecha posicionadas lada a lado en tiras, se alinean con un dispositivo de selección apropiado tal como cubiertas lenticulares colocadas sobre la pantalla. Se utiliza un rastreo o barrido invertido en el cual se crean delgadas columnas de iluminación posterior para dirigir la raya de la imagen apropiada dentro de una columna al ojo correcto.

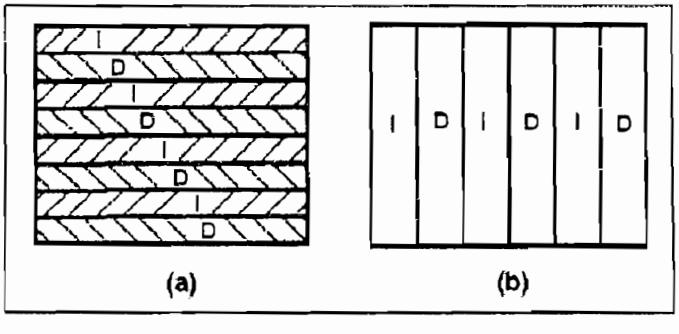


Figura 3.34 Píxel secuencial en filas y columnas

La figura 3.34a muestra filas alternadas de franjas correspondientes a las vistas izquierda y derecha, cuyo origen puede ser un campo secuencial, la

figura 3.34b muestra columnas de franjas correspondientes a las vistas izquierda y derecha, utilizada generalmente en formato auto estereoscópico.

- *Formato encima / debajo.-* este formato fue creado con la finalidad de a la vez de crear imágenes estereoscópica sobre una pantalla, aprovechar la infraestructura existente en sistemas de video y graficación en computadora, sin necesidad de hacer modificaciones del hardware o los procedimientos básicos del funcionamiento.

Se fundamenta en desplegar el par estereoscópico alternado mostrando al observador la imagen derecha e izquierda secuencialmente. En un primer intento por poner en práctica esta técnica se pensó en utilizar el sistema de 60 cuadros por segundo, codificando alternadamente los cuadros de información izquierdo y derecho, lo que resultaba en una reducción de la mitad del número de campos que alcanzan a ver cada ojo, esto produce un intolerable parpadeo (flicker).

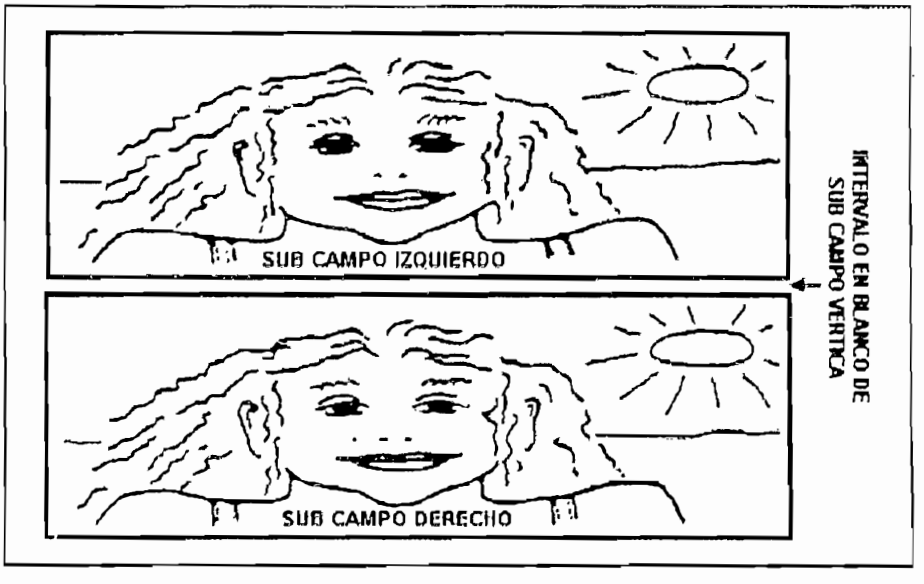


Figura 3.35 Imágenes de las vistas izquierda y derecha, con deformidad vertical posicionadas una encima de otra

Como solución a esto se pensó en duplicar el número de campos por segundo, duplicando la velocidad de rastreo vertical, con lo que se obtienen

120 campos por segundo. Mientras el número de campos se duplica, el número de líneas por campo se reduce a la mitad.

En un monitor con el estándar de 60Hz , dos imágenes, los campos izquierdo y derecho, serían observadas posicionadas una encima de la otra y con una deformidad vertical (como si estuvieran aplastadas) figura 3.35.

Estos dos campos, en el formato de 120 Hz, son conocidos como sub campos, los mismos que al ser desplegados en un monitor de 120 campos por segundo muestran las dos imágenes en secuencia de la siguiente manera: izquierda--derecha—izquierda—derecha. Note que con este sistema el requerimiento de ancho de banda permanece igual para el sistema de 120 Hz como para sistemas de video generalmente empleados en computadoras por lo que existen computadoras que no requieren de modificación de hardware para utilizar este sistema.

- *Formato lado a lado.*- esta técnica nace como solución al problema que presentaba el formato arriba – debajo de no tener suficientes líneas de rastreo. La solución del formato arriba – abajo es buena para aplicaciones de gráficos por computadora debido a que estas presentan un mayor número de líneas de rastreo que en televisión.

  Para mostrar los pares estereoscópicos en formato lado a lado, las imágenes izquierda y derecha de la cámara estereoscópica se almacenan para ser reproducidas al doble de la velocidad de la que fueron almacenados. Además, los campos se concatenan o se revuelven para conseguir el modelo izquierdo-derecho necesario. El resultado es una señal del doble de ancho de banda que lo normal, la cual preserva las características originales de la imagen y adicionalmente es estereoscópica. Lo anteriormente descrito es utilizado para ver imágenes en tiempo real sobre monitores con una frecuencia de despliegue de 120 cuadros por segundo. La figura 3.36 muestra como quedarían las imágenes del par estereoscópico en el formato lado a lado.

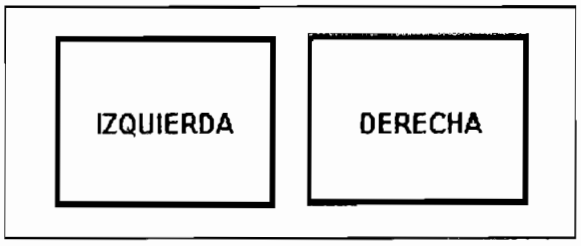


Figura 3.36 Estéreopar en formato lado a lado.

- *Flujo doble.*- en este formato se alimenta con un flujo individual de imágenes a cada una de las pantallas correspondientes a cada ojo, teniendo por separado las imágenes correspondientes al ojo izquierdo y derecho, una ilustración del formato de flujo doble es presentada en la figura 3.37

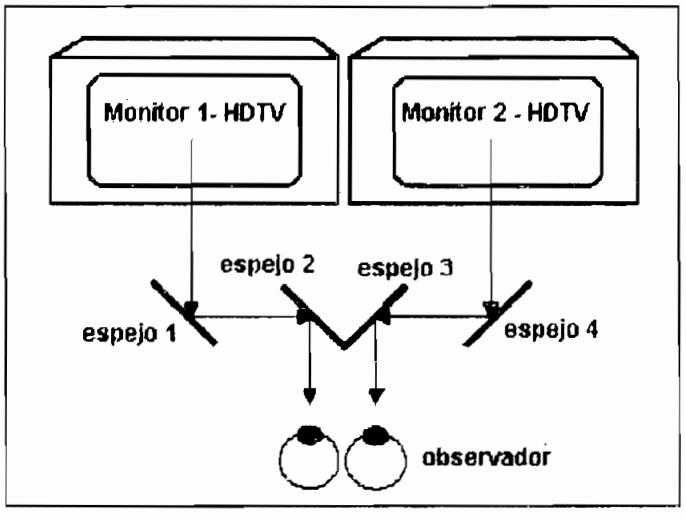


Figura 3.37 Ejemplo de utilización del formato de doble flujo

- *Código de línea blanca.*- conocido como sistema WLC por sus siglas en inglés (White-Line-Code), es usado por computadores del tipo Pentium y ofrece una solución al problema de desplegar imágenes estereoscópicas con una alta calidad a un bajo costo. Para este formato es de poca importancia si los campos de las vistas izquierda y derecha son de barrido entrelazado o progresivo y la velocidad del campo no es cosa de

preocupación. Es más en el modo entrelazado cualquiera de las líneas sean estas pares o impares puede asignársele a cualquiera de las dos vistas en perspectiva. El formato WLC fue creado para ofrecer el sistema de estéreo visión más flexible que satisfaga a los proveedores, diseñadores y usuarios. Los componentes del hardware WLC permiten una instalación rápida por parte del usuario.

Al final de cada campo para la última línea de video, se agregan líneas en blanco para indicar si el campo corresponde a una imagen derecha o izquierda. La última línea de video se escoge debido a que esta dentro del dominio del diseñador del sofware que se agregará. Cuando se reconoce al campo como izquierdo o derecho, la electrónica del hardware adicional indica al obturador de las gafas shutter que el pulso de sincronía vertical ha sido censado y cual imagen debe ser mostrada en las gafas. El WLC es universal en el sentido de que simplemente no se preocupa de si el rastreo es entrelazado o progresivo o de la velocidad de la resolución del cuadro. Si el formato WLC está allí, las gafas shutter operaran sus obturadores en sincronía con los campos y se podrá ver una imagen estereoscópica.
Los modos más populares de funcionamiento del WLC son:

1. El modo de página alternada, el cual se lo utiliza más a menudo en los juegos de acción bajo DOS que corren a una velocidad de entre 70 a 80 campos por segundo.

2. El modo multimedia o internet el cual corre a una proporción de por lo menos 90 cuadros por segundo entrelazados con una resolución de 1024 x 768 píxeles.

3.3.2 DISPLAYS AUTOESTEREOSCÓPICOS

Como ya se mencionó anteriormente un display autoestereoscópico es aquel que provee al observador de una imagen tridimensional sin necesidad de gafas

especiales, estos displays pueden ser de múltiples vistas o de rastreo de cabeza (head tracked).

Los displays autoestereoscópicos combinan dos importantes tipos de información del mundo real que nos permite obtener una apreciación estereoscópica del entorno, estas son:

- Paralaje estéreo, que se refiere a la capacidad de ver una imagen diferente con cada ojo y
- Paralaje en movimiento, que es la posibilidad de ver imágenes diferentes cuando se mueve la cabeza.

La figura 3.38(a) muestra a un observador mirando una escena, él ve una imagen diferente de la escena con cada ojo y diferentes imágenes cada vez que el mueve su cabeza, siendo capaz de ver potencialmente un infinito número de diferentes imágenes de la escena. En la figura 3.38(b) se muestra al mismo observador, viendo dividido el espacio en un finito número de ventanas horizontales.

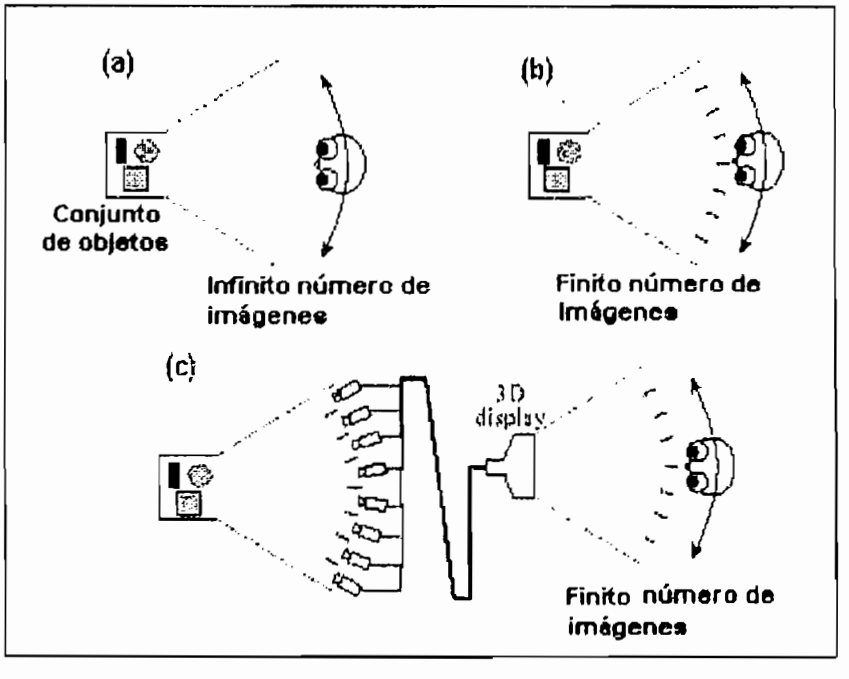


Figura 3.38 Número de vistas provistas a un observador

En cada ventana solamente una imagen o vista de la escena es visible. Sin embargo el observador ve dos imágenes diferentes y las imágenes cambian cada vez que mueve su cabeza aunque con saltos ya que el observador mueve su cabeza de ventana en ventana. De esta manera ambos movimientos de paralaje pueden proveer con un pequeño número de vistas.

Esto no es una restricción fundamental para que solo se pueda mover la cabeza horizontalmente, se podría mover la cabeza verticalmente pero se debería de proveer de vistas para el movimiento de paralaje vertical con lo cual se cuadriplica el número de vistas.

El finito número de vistas requerido en la figura 3.38 (b) permite reemplazar la escena por un display 3D que provee una diferente imagen a cada ventana como se muestra en la figura 3.38 (c). Este es el principio del display autoestereoscópico de múltiples vistas.

Los displays de rastreo de cabeza por otro lado trabajan desplegando solamente dos vistas y rastreando la cabeza del observador de tal manera que cada ojo vea solamente la vista correcta. Si el proceso de generación de imágenes toma en cuenta la posición de la cabeza entonces el efecto de paralaje en movimiento puede ser simulado. Por otra parte un display de rastreo de cabeza solamente provee paralaje estereoscópico.

3.3.3 TIPOS DE DISPLAYS ESTEREOSCÓPICOS

Se pueden identificar tres tipos de displays estereoscópicos :

- displays de dos vistas
- displays de rastreo de cabeza, normalmente de dos vistas
- displays de múltiples vistas con tres o más vistas.

Los cuales para desplegar los pares estereoscópicos pueden tener como fundamento tecnológico el uso de :

Displays de barrido de paralaje.- son los que utilizan un arreglo de aberturas ópticas cada una de las cuales es alineada con por lo menos dos columnas de

píxeles del LCD, las aberturas pueden ser incluidas como aberturas en una máscara o como líneas de luz de la imagen. De esta manera es posible producir arreglos de aperturas con una alta calidad óptica los cuales pueden ser mejorados cubriendo la superficie del display con una superficie antireflejante con lo que la superficie del display puede ser sustancialmente mejorada, la figura 3.39 muestra la estructura de este tipo de display.

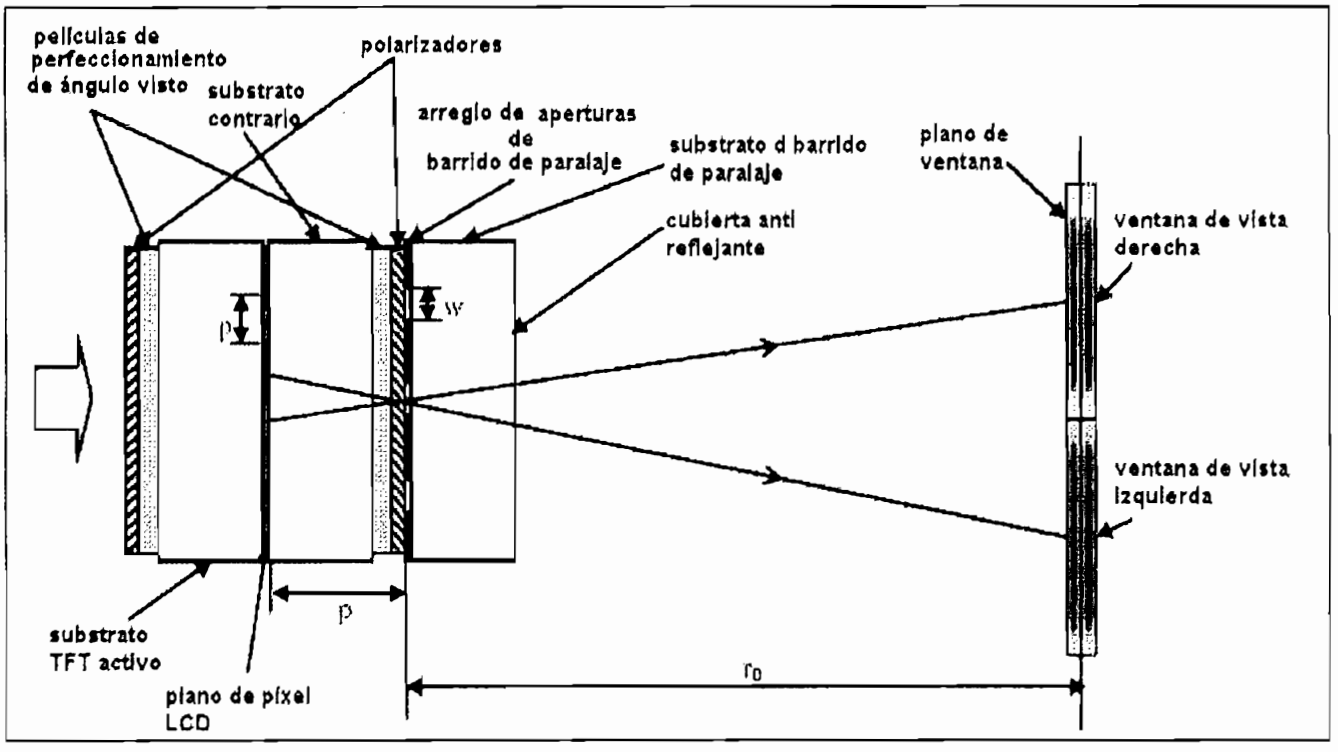


Figura 3.39 Estructura de un display de barrido de paralaje

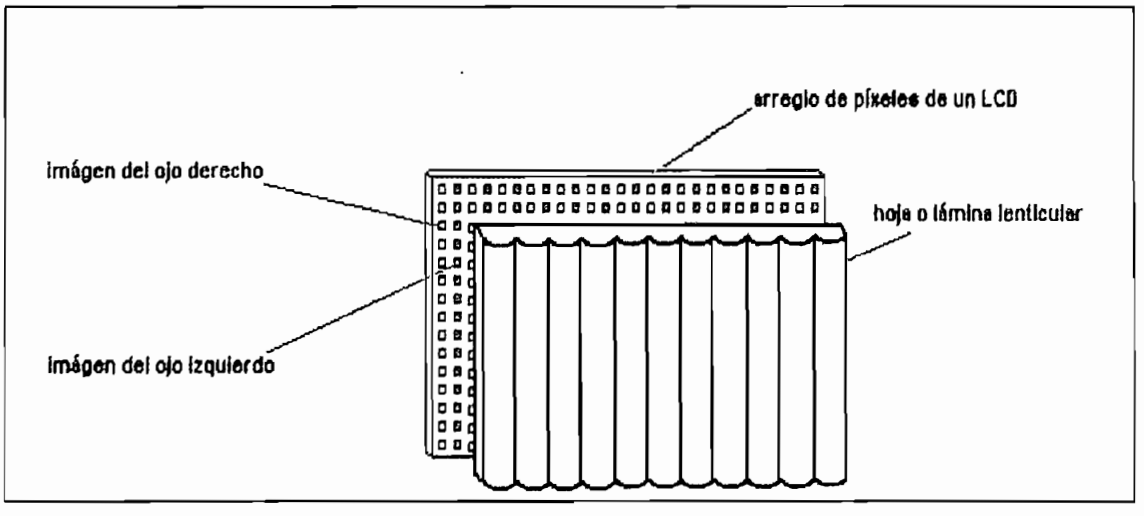


Figura 3.40 Estructura de display lenticular

Displays lenticulares.- usan un arreglo de lentes cilíndricos refractivos, cada uno de los cuales es alineado con por lo menos dos columnas de píxeles LCD. Los displays lenticulares tienen brillo total y la resolución óptica de estos elementos puede ser superior al de la apertura del barrido de paralaje, los cuales pueden en principio dar ventanas claramente definidas. La figura 3.40 muestra la estructura básica de un display lenticular.

3.3.3.1 Displays de dos vistas

Para un display autoestereoscópico de dos vistas se pueden usar dos clases de tecnología, como son el barrido de paralaje o la tecnología de hoja lenticular, las cuales dividen en dos juegos de imágenes en la resolución horizontal típicamente mostradas por displays de cristal líquido, donde una de las dos imágenes visibles esta formada por cada píxel de la segunda columna y la segunda imagen corresponde a los píxeles de la otra columna. Las dos imágenes son capturadas o generadas de tal manera que una es apropiada para el ojo izquierdo del observador y la otra apropiada para el ojo derecho.

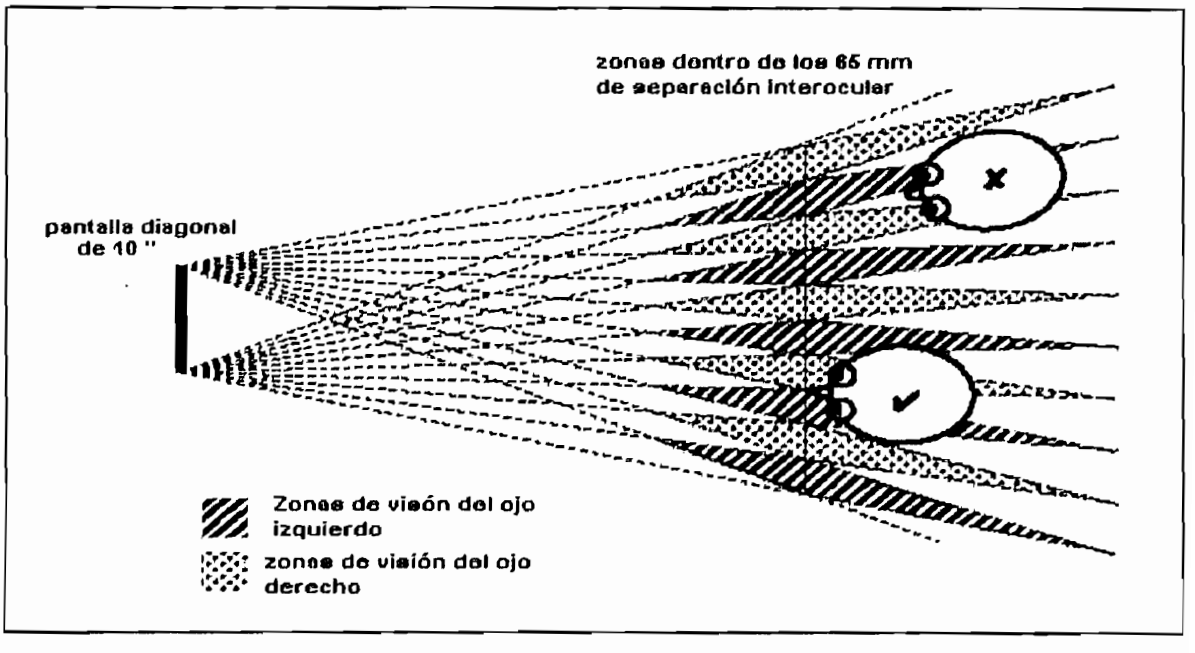


Figura 3.41 Espacio de visualización de un sistema de display de dos vistas.

Las dos imágenes desplegadas son visibles en múltiples zonas en el espacio como se puede apreciar en la figura 3.41 . Si el espectador esta de pie a la

distancia ideal y en la posición correcta él percibirá una imagen estereoscópica, la desventaja de esto es que existe el 50% de probabilidades de que el observador este en una mala posición y vea una incorrecta pseudo imagen estereoscópica. Al moverse muy por delante o por detrás de la distancia ideal se incrementa la posibilidad de ver una imagen incorrecta. Esta seria limitación que hace necesario el uso de otra solución autoestereoscópica, la cual sería incrementar el número de vistas o introducir un sistema de rastreo de cabeza.

3.3.3.2 Displays de rastreo de cabeza

Como ya se indicó anteriormente la tecnología de dos vistas funciona correctamente pero solo dentro de un rango, sin embargo al saber la posición de la cabeza del observador, entonces las apropiadas imágenes izquierda y derecha pueden ser desplegadas en las zonas apropiadas, previniéndose así cualquier vista seudo estereoscópica como se indica en la figura 3.42

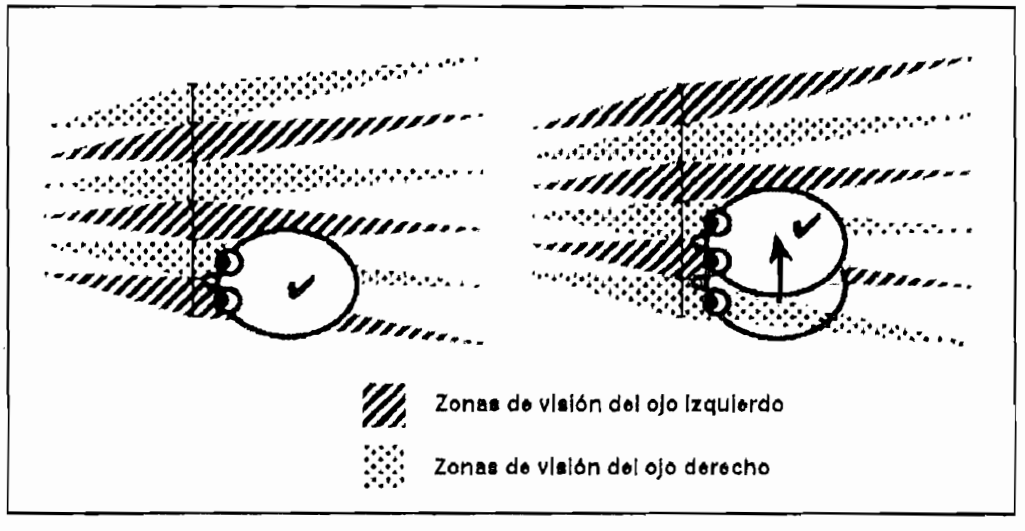


Figura 3.42 Despliegue de las vistas apropiadas al conocer la posición de la cabeza

. Alternativamente una tecnología completamente diferente podría ser usada, la cual permite que solo dos zonas se desplieguen y estas sean movidas físicamente como se puede apreciar en la figura 3.43. La principal dificultad con este método es el propio rastreo de cabeza, ya que se debe utilizar algún mecanismo que no requiera que el usuario lleve puesto algún implemento especial, ya que sería en vano reemplazar las gafas especiales con otro tipo de

dispositivo especial que sirva para el rastreo de la cabeza. Recientemente se han desarrollado tales mecanismos y se ha alcanzado la fase de utilidad comercial. La otra limitación de la mayoría de sistemas de rastreo de cabeza es que son construidos para un solo espectador, lo cual es aceptable en algunas aplicaciones pero no en otras en las cuales se hace necesaria el considerar la alternativa de múltiples vistas.

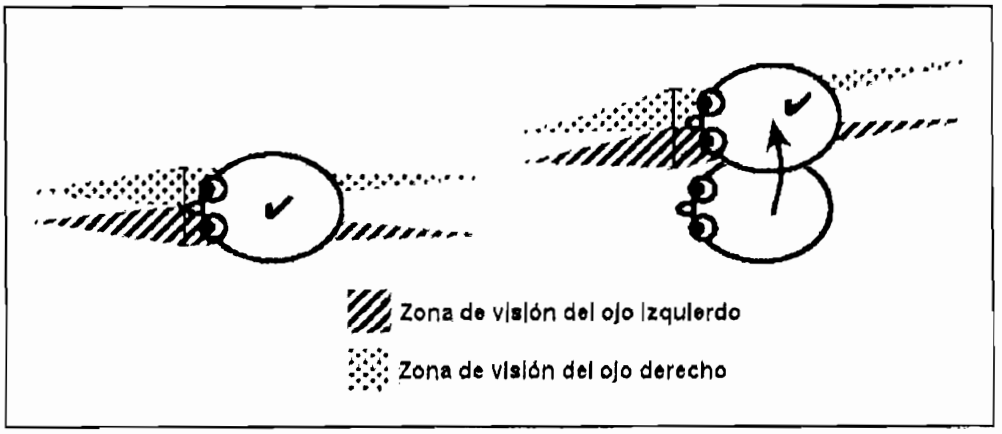


Figura 3.43 Despliegue de dos zonas que se mueven según el movimiento de la cabeza

3.3.3.3 Displays de múltiples vistas

Estos displays despliegan múltiples imágenes diferentes a múltiples zonas en el espacio como se ilustra en la figura 3.44 y 3.45. Esto tiene las ventajas que:

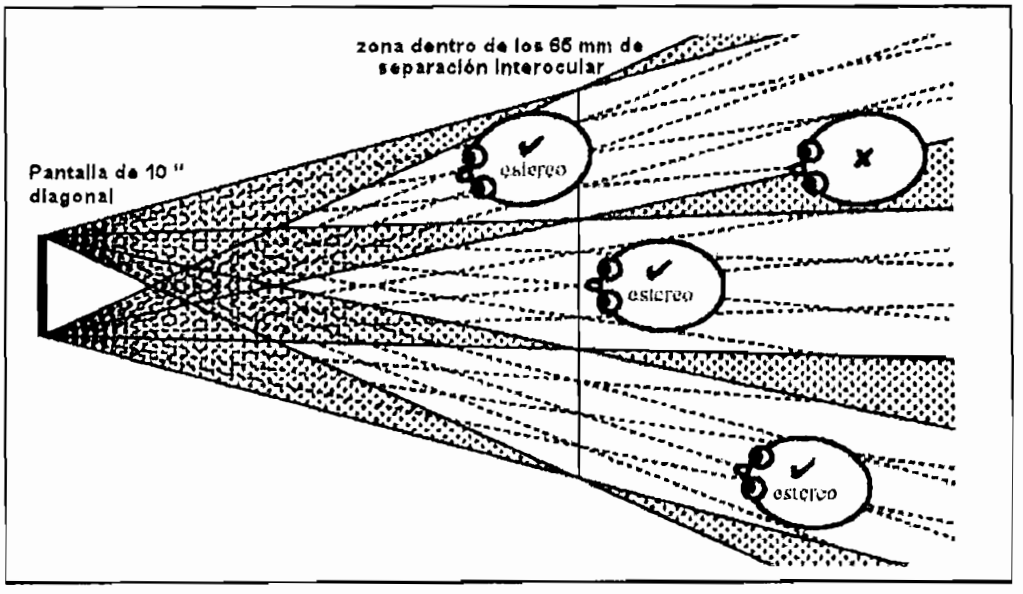


Figura 3.44 Cuatro vistas de display autoesteroscópico con tres lóbulos

- El observador es libre de poner su cabeza en cualquier parte dentro del lóbulo de visión mientras todavía se perciban imágenes estereoscópicas.
- El espectador tiene la capacidad de mirar alrededor objetos en la escena simplemente moviendo su cabeza.
- El sistema soporta múltiples observadores, cada uno ve una escena estereoscópica desde su propio punto de vista (figura 3.48), y no se requiere de un rastreador de cabeza con toda la complejidad requerida asociada a tal sistema.

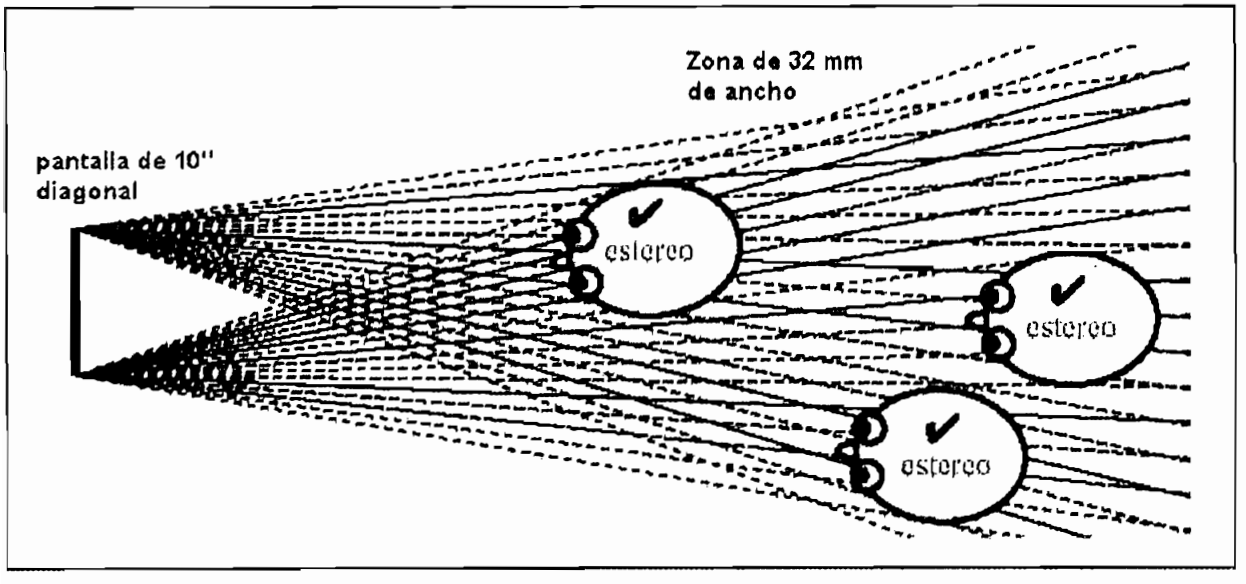


Figura 3.45 Dieciséis vistas de display autoesteroscópico con un solo lóbulo

Las desventajas del sistema de múltiples vistas son:

- La dificultad de construir un display con muchas vistas
- El problema de generar todas las vistas simultáneamente, ya que cada vista esta desplegándose en todo momento, pudiendo ser vista por alguien o no.

CAPITULO IV

4. DESCRIPCIÓN DE LAS PRINCIPALES RECOMENDACIONES PARA LA TRANSMISIÓN DE SEÑALES ESTEREOSCÓPICAS.

La televisión digital en el futuro tendrá un gran mercado de expansión y los estándares creados en los últimos años se han popularizado en la actualidad. En la mayoría de los casos estos estándares se han basado en las necesidades de aplicaciones específicas. Sin embargo existen normas muy importantes que se deben de tomar en cuenta como las siguientes.

4.1 JPEG (JOINT PHOTOGRAPHIC EXPERTS GROUP)

Este grupo había sido formado por la Organización de Estándares Internacionales (ISO) y por la Comisión Electrotécnica Internacional (IEC) para formular un estándar que permitiera disminuir la cantidad de espacio de almacenamiento requerido para las imágenes fijas. Fue aprobado en 1992 y es válido para imágenes con tonos de gris como para imágenes en color. El formato JPEG se lo utiliza generalmente para mostrar catálogos de fotografías, o también en documentos de lenguaje HTML en la World Wide Web.

Existen cuatro modos de funcionamiento en la codificación JPEG: básico/secuencial, sin pérdidas, progresivo y jerárquico. Sin embargo el modo básico/secuencial es el de mayor uso.

- *Modo básico/secuencial,* el proceso de codificación se lo muestra en la figura 4.1 y se lo resume de la siguiente forma:
  - La imagen de entrada se divide en subimágenes o bloques de 8x8 píxeles.

- Se resta la componente continua (DC) del bloque y se cuantifica la diferencia de su valor respecto al término DC del bloque anterior.
- Una vez que se elimina la componente continua de cada bloque, se transforma mediante la DCT y se cuantifican los coeficientes transformados mediante un cuantificador escalar uniforme. Los pasos de cuantificación están definidos para cada uno de los 64 coeficientes en una matriz de cuantificación de 8x8. Generalmente se utiliza una matriz de cuantificación para la luminancia y otra para la crominancia.
- Se ordenan los coeficientes transformados en zigzag de forma que quedan ordenados de menor a mayor frecuencia y se cuantifican con mayor precisión los coeficientes de baja frecuencia del bloque transformado. Una vez ordenados, se codifican mediante un código run length (RLE) que tiene dos campos: longitud y valor, de los cuales la longitud indica el número de repeticiones consecutivas de un mismo caracter y el campo valor indica cuál es el caracter que se repite.
- La salida del codificador RLE y el término DC se codifican mediante un código de longitud variable tipo Huffman.

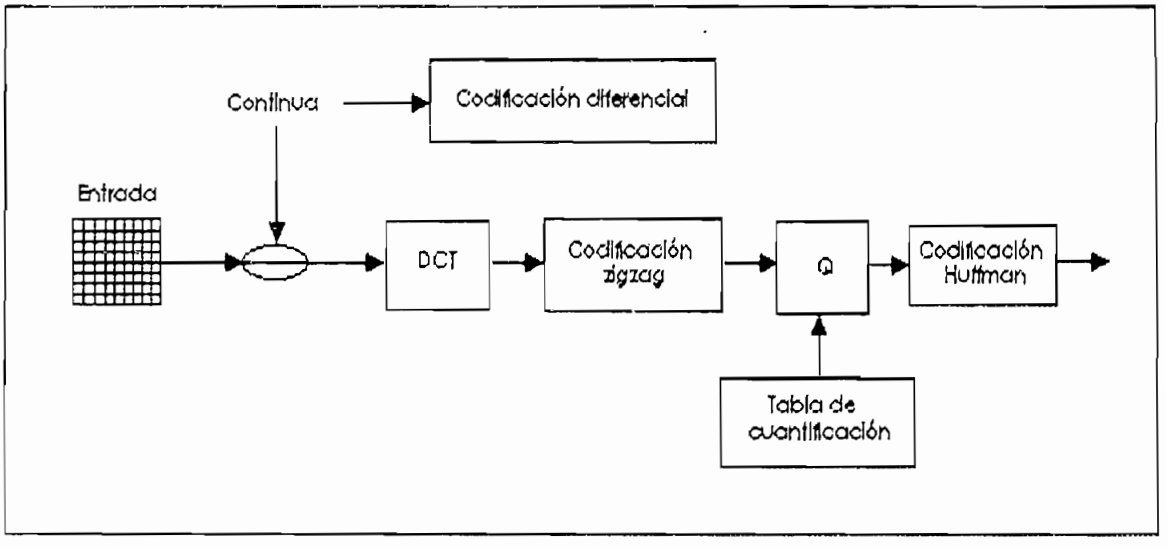


Figura 4.1 Diagrama de bloques de la codificación JPEG

- *Modo sin pérdidas,* para reducir la tasa de bits de la imagen original sin que aparezca error en el proceso de decodificación, se realiza un proceso predictivo, el cual consiste en formar una combinación lineal de píxeles vecinos ya codificados anteriormente. Para ello se utilizan siete posibles predictores, que son los que se muestran en la figura 4.2. La diferencia entre la imagen original y la predicción es la parte no predecible que, para conseguir una compresión sin error, tendrá que ser codificada de forma exacta. El estándar JPEG en el modo sin error utiliza una codificación Huffman.

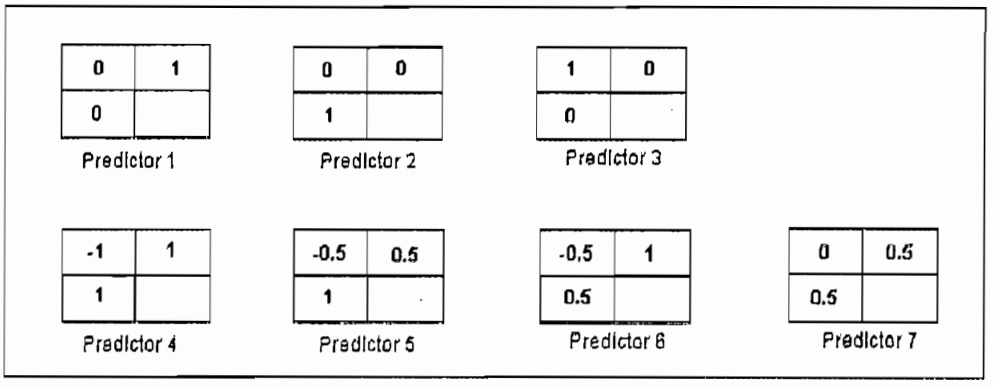


Figura 4.2 Esquemas de predicción (el píxel a predecir es el inferior derecho)

- *Modo progresivo,* el objetivo de este modo es visualizar inicialmente la imagen en un modo de baja calidad, para ir aumentándola progresivamente. Existen dos maneras para lograr este objetivo:
  - Selección de coeficientes transformados.
  - Aproximaciones sucesivas.

- *Modo jerárquico,* se trata de un algoritmo semejante a la codificación piramidal descrita en el capitulo anterior y se lo puede resumir de la siguiente forma:
  - Diezmar la imagen por un factor 2 en cada dirección.
  - Codificar la imagen resultante usando otro método.
  - Decodificar la imagen comprimida y restaurar el tamaño original, interpolándola por el factor diezmado.

- Codificar la diferencia entre la imagen original y la interpolada.

4.2 ESTÁNDARES DE CODIFICACIÓN MPEG (MOVING PICTURES EXPERTS GROUP)

La organización ISO/IEC crea el Comité Técnico Unido sobre Tecnologías de la Información, Subcomité 29, Grupo de Trabajo 11 (ISO/IEC JTC1/SC29/WG11), más conocido como MPEG, el que se encarga del desarrollo de estándares para la representación codificada de imágenes en movimiento, la información del audio asociado, y su combinación para la grabación y lectura en un medio de almacenamiento digital, es decir MPEG desarrolla un conjunto de estándares para compresión de video digital que estén en concordancia con las 7 capas del modelo OSI Existen numerosas versiones que han ido apareciendo a lo largo de los años, las cuales se irán describiendo a continuación en lo que se refiere a imagen.

4.2.1 ESTÁNDAR MPEG-1

El estándar MPEG-1 (numerado como ISO/IEC 11172) fue creado en el año de 1993 con el objetivo de leer discos compactos con imágenes en movimiento a una velocidad de transmisión cercana a 1.5 Mbps. MPEG-1 soporta velocidades mayores que la recomendación UIT-T H 261. dado que soporta un amplio rango de aplicaciones, el usuario puede fijar un gran número de parámetros, tales como el número de imágenes por segundo, el tamaño de la imagen, etc.

El formato MPEG-1 trabaja con 3 tipos de imágenes, las imágenes tipo I, tipo P y tipo B.

Imagen intracodificada (I) se codifica utilizando solamente información de sí misma;

Imagen con codificación predictiva (P) es una imagen que se codifica utilizando predicción con compensación de movimiento a partir de una trama de referencia pasada o de un campo de referencia pasado;

Imagen codificada con predicción bidireccional (B) es una imagen que se codifica utilizando predicción con compensación de movimiento a partir de una trama (o tramas) de referencia pasada y/o futura[24].

La secuencia de imágenes de más uso es I, B, B, P, B, B, P, B, B, I, B, B, P, B, B, P, B, B, etc. Sin embargo, puede variarse según decida el decodificador, y no tiene por que ser regular.

En cada uno de los modos de imagen se tienen relaciones de compresión como se detalla en la tabla 4.1.

TIPO DE IMAGEN	COMPRESIÓN
I	7 : 1
P	20 : 1
B	50 : 1

Tabla 4.1 Compresión para cada tipo de imagen

Otra característica de MPEG-1 es el "relleno condicional", que consiste en actualizar la información de un macrobloque en el receptor sólo si es necesario, es decir que si la información cambia se volverá a enviar la información correspondiente, caso contrario no.

4.2.2 ESTÁNDAR MPEG-2

Establecido en 1994 para ofrecer mayor calidad con mayor velocidad de transmisión (típicamente de 3 a 10 Mbits). En esa banda, proporciona 720x486 píxeles de resolución, es decir, calidad TV. Fue diseñado para tener compatibilidad con MPEG-1.

MPEG-2 puede describirse como una "caja de herramientas" de compresión más compleja que MPEG-1, por lo tanto, también puede ser considerada como una unidad superior: en efecto, toma todas las herramientas anteriores y le añade otras.

[24] Recomendación UIT-T H 262

Perfiles y niveles MPEG-2

MPEG-2 se puede utilizar en un vasto rango de aplicaciones, requiriendo diferentes grados de complejidad y desempeño. Para un propósito práctico el estándar MPEG-2 es dividido en perfiles y cada perfil es subdividido en niveles, los cuales permiten establecer las compatibilidades entre diversos equipos MPEG-2.

Un «perfil» es un subconjunto definido de toda la sintaxis de tren de bits definida por esta especificación. Dentro de los límites impuestos por la sintaxis de un perfil dado, es posible aún requerir una variación muy grande en el funcionamiento de los codificadores y decodificadores según los valores tomados por los parámetros en el tren de bits[25]. En otras palabras un perfil es básicamente el grado de complejidad esperada en la codificación. Cada perfil define un nuevo conjunto de algoritmos a añadir a los del perfil inmediatamente inferior.

Un nivel es un conjunto definido de restricciones impuestas a los parámetros en el tren de bits. Estas restricciones pueden ser simples límites de números. Como otra posibilidad, pueden adoptar la forma de restricciones en combinaciones aritméticas de los parámetros (por ejemplo, la anchura de trama multiplicada por la altura de trama multiplicada por la velocidad de trama)[2].

El nivel especifica el margen de valores que puede soportar cada uno de los parámetros, entonces se puede describir que un nivel describe el tamaño de la imagen, la resolución de ésta o la velocidad de transferencia de bits usada en ese perfil. Un codificador MPEG cuando entrega un perfil y un nivel determinado, debe además ser capaz de decodificarlo a perfiles y niveles inferiores. La tabla 4.2 describe las características principales de los niveles y la tabla 4.3 de los perfiles.

---

[25] Recomendación UIT-T H 262, pag iv

NIVEL	PARÁMETROS
Alto	1920 muestras/línea, 1152 líneas/imagen, 60 imágenes/segundo, 80 Mbps
Alto 1440	1440 muestras/línea, 1152 líneas/imagen, 60 imágenes/segundo, 60 Mbps
Principal	720 muestras/línea, 576 líneas/imagen, 30 imágenes/segundo, 15 Mbps
Bajo	352 muestras/línea, 288 líneas/imagen, 30 imágenes/segundo, 4 Mbps

Tabla 4.2 Características de los diferentes niveles de un perfil.

PERFIL	ALGORITMOS
Alto	3 capas y modo 4:2:2
Escalable espacialmente	Añade la escabilidad espacial (2 capas), 4:0:0
Escalable SNR	Añade la escabilidad de SNR al perfil anterior, 4:2:0
Principal	Soporta imágenes B, 4:2:0
Simple	No soporta imágenes B, Modo 4:2:0

Tabla 4.3 Funcionalidades soportadas en cada perfil.

En la tabla 4.4 se menciona las características más preponderantes del formato MPEG-2.

Aplicación	TV digital y HDTV
Resolución espacial	4 CIF, 16 CIF
Resolución temporal	50-60 100-120 campos/segundo
velocidad de transmisión	4 - 20 Mbps
Calidad	TV (NTSC o PAL)
Tasa de compresión	30 - 40

Tabla 4.4 Características MPEG-2.

En la actualidad existen cuatro modos escalables: escalabilidad espacial, particionamiento de datos, escalabilidad SNR y escalabilidad temporal. Estos modos dividen al video en diferentes capas (base, media y alta) con la finalidad de prioritizar los datos de video.

El propósito de la escalabilidad es para divisiones complejas. A continuación un detalle breve de los modos de escalabilidad:

- *Escalabilidad espacial:* Este método de dominio espacial codifica la capa base a una dimensión de muestreo bajo (por ejemplo resolución) que las capas superiores. Las capas bajas (base) reconstruidas delñ muestreo son usadas como predicción de las capas superiores.

- *Particionamiento de datos:* es un método de dominio de frecuencia que rompe los bloques de 64 coeficientes cuantizados de la transformada dentro de las cadenas binarias. La primera, cadena de alta prioridad contiene los coeficientes más críticosde las frecuencias bajas e información (tales como valores DC, vectores, etc.), la segunda, cadena binaria de baja prioridad lleva datos AC de las altas frecuencias.

- *Escalabilidad SNR:* es un método de dominio espacial donde los canales son codificados a velocidades de muestreo idénticas, pero con diferentes calidades de imágenes. La cadena binaria de alta prioridad contiene datos de la capa base que pueden ser **añadidos** a la capa de refinamiento de baja prioridad para construir un imagen de alta calidad.

- *Escalabilidad temporal:* es un método de dominio temporal usado por ejemplo en video estereoscópico. La primera, la cadena binaria de alta prioridad codifica video a una baja velocidad de tramas, y las tramas intermedias pueden ser codificadas en una segunda cadena binaria usando la reconstrucción de la primera cadena binaria como predicción. Por ejemplo en una visión estreoscópica, el canal de video izquierdo puede ser predecido del canal derecho[26].

4.2.3 ESTÁNDAR MPEG-3.

El estándar MPEG-3 fue creado en un inicio para controlar la televisión digital de alta definición (HDTV), la cual usa imágenes de 1920 x 1080 píxeles. Posteriormente fue incluido dentro del estándar MPEG-2.

4.2.4 ESTÁNDAR MPEG-4

En un futuro muy cercano la convergencia del mundo de las computadoras y el consumo de productos audiovisuales estarán acompañados por grandes avances en las telecomunicaciones. Las redes de ordenadores y la industria cinematográfica ofrecen un potencial para la explotación de estas aplicaciones.

Este estándar fue desarrollado para un amplio rango de aplicaciones, desde tasas de bits de 5 a 64 Kbps para aplicaciones telefónicas a velocidades hasta 4 Mbps para aplicaciones de televisión digital.

MPEG-4 incluye un concepto nuevo denominado escalabilidad basada en el contenido, la cual proporciona los mecanismos necesarios para interactuar y modificar el contenido de las imágenes. Para ello se utiliza los planos de objeto de video (VOP: **V**ideo **O**bject **P**lanes), el cual consiste en segmentar cada una de las imágenes en un número de regiones de la imagen de forma arbitraria. Cada una de estas regiones puede contener una parte en concreto de la imagen. De esta forma, a diferencia de los estándares MPEG anteriores, no se divide la imagen en bloques cuadrados, sino en objetos. La forma y posición de cada uno de los objetos puede variar de una imagen a otra, y los VOP pertenecientes a un mismo objeto físico de la imagen se denominan objetos de video (VO: video objects). Para cada uno de los VO se codificará su textura y movimiento. Además se añade información de cómo se recompone la imagen original, para poder eliminar o añadir nuevos objetos en la imagen del receptor, así como es posible manejar prioridades en los objetos, de tal forma que los más importantes se representen con resoluciones espaciales y/o temporales mayores.

---

[26] http//neuton.ing.ucv.ve/revista-e/No1/Mpeg2.htm

4.2.5 ESTÁNDAR MPEG-7

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.

Mientras que buscar texto es relativamente fácil con un ordenador, resulta más difícil encontrar partes concretas de audio y video basadas en su contenido. MPEG-7 pretende describir los diferentes objetos de forma que sea posible una búsqueda eficiente de los mismos.

4.3. RECOMENDACIÓN UIT-R BT.1438: EVALUACIÓN SUBJETIVA DE LAS IMÁGENES DE TELEVISIÓN ESTEREOSCÓPICA.

Siendo la Televisión Estereoscópica un servicio de radiodifusión del futuro, en su diseño y construcción se debe tomar muy en cuenta las evaluaciones subjetivas como un elemento vital, debido a que las características del ojo humano del observador son muy frágiles y se pueda tener la correcta percepción de profundidad sin que se produzca la fatiga del observador.

Las condiciones de evaluación comunes para los sistemas de Televisión Estereoscópica deberían de incluir: métodos de evaluación, las condiciones de filmación, las condiciones de visualización, los materiales de prueba a utilizar en la evaluación y los métodos de cribado que sirven para asegurar que los observadores tienen una percepción de profundidad normal.

Factores de evaluación

Entre los principales factores que se deben de tener en cuenta tenemos:

- *Relación de profundidad:*

Resolución espacial en profundidad. Una resolución reducida en profundidad puede reducir la calidad de la imagen de la televisión estereoscópica.

- *Movimiento en profundidad:*
  Factor que establece el movimiento en el sentido de la profundidad se reproduce sin discontinuidades.
- *Efecto teatro de marionetas.*
  Describe un tipo de distorsión en imágenes 3-D. A veces los objetos estereoscópicos se perciben como anormalmente grandes o pequeños.
- *Efecto papel de cartón.*
  Describe otro tipo de distorsión en la reproducción de imágenes 3-D. Las posiciones 3-D de objetos estereoscópicos se perciben de manera estereoscópica, pero estos parecen ser anormalmente delgados[27].

Además se tomaría en cuenta los factores de evaluación que normalmente se aplican a la televisión monoscópica tales como: resolución, representación del color, representación del movimiento, calidad general, nitidez de perfiles, profundidad, etc.

Condiciones de visualización.

En las condiciones de visualización se deben de tomar en cuenta dos condiciones:

- Efecto del marco de visualización.
- Inconsistencia entre acomodación y convergencia.

Materiales de prueba.

Se describen ocho pruebas de visión (VT, vision tests) principales. Los observadores deben tener estereopsis normal, es decir deben de haber superado la prueba VT-04 y VT-07. Las pruebas de visión son las siguientes:

[27] Recomendación UIT-T BT.1438

- VT-01: Percepción simultánea.
- VT-02: Fusión binocular.
- VT-03: Estereopsis gruesa.
- VT-04: Estereopsis de detalle.
- VT-05: Límite de fusión cruzada.
- VT-06: Límite de fusión sin cruce.
- VT-07: Estereopsis dinámica.
- VT-08: Agudeza binocular.

4.4. RECOMENDACIÓN UIT-R BT.2017: PERFIL MULTIVISIÓN MPEG-2 PARA TELEVISIÓN ESTEREOSCÓPICA

En esta recomendación se introduce además del estándar MPEG un nuevo perfil que es el Perfil Multivisión (MVP) útil para aplicaciones que necesitan muchos puntos de visión en el contexto de la norma de video MPEG-2, como es el caso de la Televisión Estereoscópica. El MVP admite imágenes estereoscópicas como imágenes fuente para una amplia gama de resoluciones y calidades de imagen, que dependen de las necesidades de las aplicaciones de que se trate.

Entre sus principales características tenemos:

Codificación monoscópica en su capa base a efectos de compatibilidad y predicción híbrida de movimientos y disparidad a fin de aumentar la eficacia de la compresión. Para codificar una capa de mejora se utilizan herramientas de escalabilidad temporal. A la capa base se le aplica una codificación monoscópica con las mismas herramientas que el perfil principal (MP, Main Profile). Se asigna una capa base de MVP a la visión izquierda y una capa de mejora a la visión derecha. La capa de mejora se codifica utilizando herramientas de escalabilidad temporal y en la capa mejorada puede aplicarse la predicción híbrida de movimientos y disparidad. Se prevé una mayor compresión de la visión derecha

del video estereoscópico a causa del parecido entre la visión izquierda y la visión derecha. Un ejemplo se describe en la figura 4.3.

Los niveles del MVP son: alto, alto-1440, principal y bajo. Las características de los niveles son las mismas que se detallaron en el estándar MPEG-2.

Es importante mencionar que el perfil multivisión MPEG ofrece una base para codificación y compresión de las secuencias de Televisión Estereoscópica.

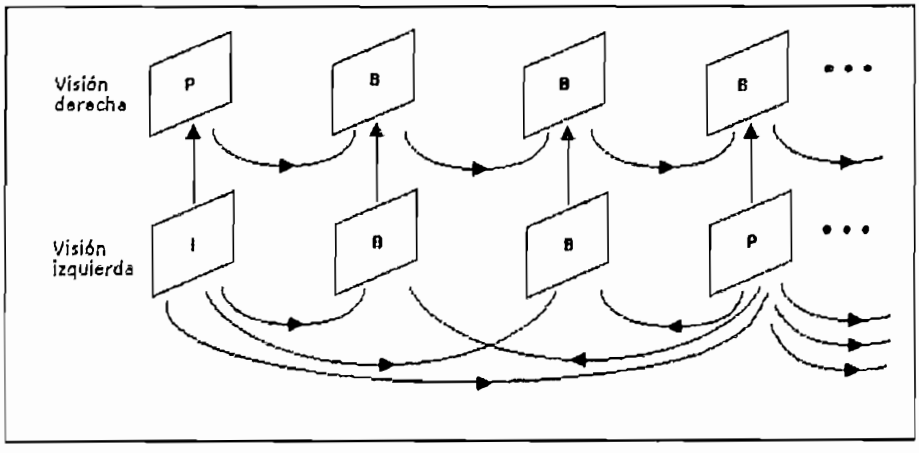


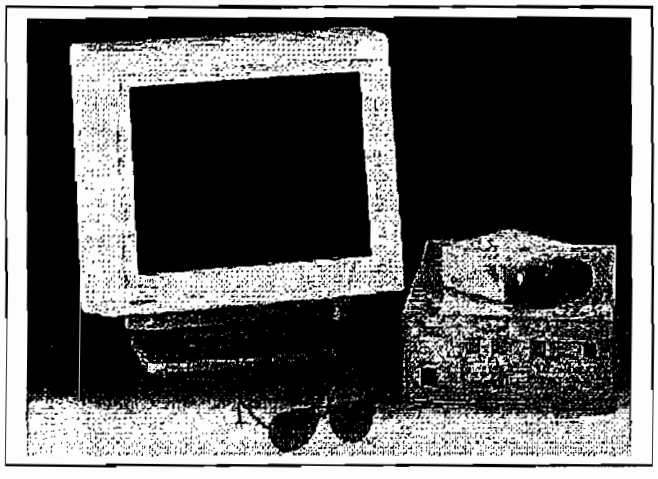
Figura 4.3 Ejemplo de configuración de predicción de la visión izquierda, imagen de trama de la visión derecha codificada mediante predicción de disparidad respecto a la visión izquierda y predicción de movimientos con respecto a si misma.

CAPÍTULO V

5. PRODUCTOS EXISTENTES PARA LA VISUALIZACIÓN DE IMÁGENES ESTEREOSCÓPICAS.

El interés por el ser humano de conseguir un sistema artificial que simule de una manera muy parecida el sistema visual natural ha sido un reto desde hace mucho tiempo atrás, pero es hoy en día que con el acelerado progreso tecnológico se han conseguido resultados realmente sorprendentes en el campo de la visualización estereoscópica. Es de esta manera como varias empresas están continuamente introduciendo al mercado equipo para visualización estereoscópica que día a día nos sorprende mas con sus resultados y calidad de imagen ofrecida. Es así como aquí se muestran los siguientes productos.

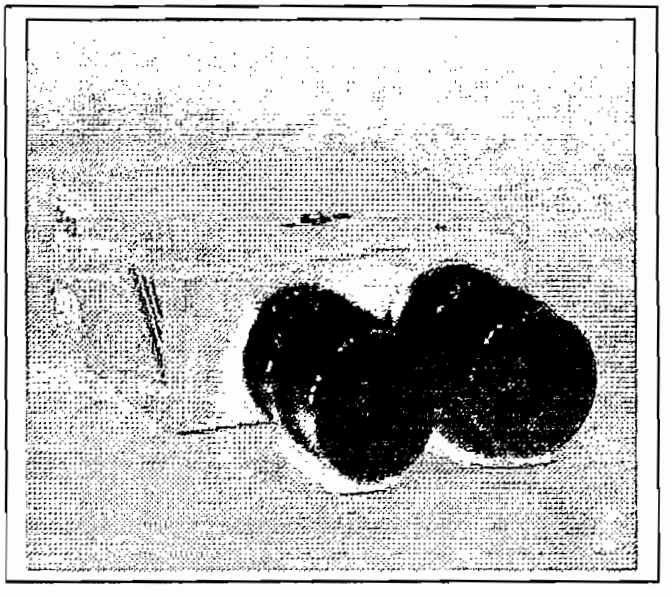
5.1 SISTEMA DE VIDEO 3D ESTEREOSCÓPICO KAPPA



El sistema entrega una verdadera imagen tridimensional en tiempo real. El sistema consiste de una cámara estereoscópica CF23/CF44 que adquiere imágenes separadas izquierda y derecha, la cual entrega una señal de TV convencional que contiene la información para cada ojo (en campo secuencial), además tiene un sistema conversor de barrido de video SM 100 que elimina el parpadeo, ya que dobla la frecuencia de video regular a 120 Hz (PAL: 100 Hz) proporcionando un despliegue de imágenes a 60 Hz a ambos ojos (PAL: 50 HZ). El resultado es una clara y estable imagen 3-D. Esta imagen puede ser congelada con la utilización

de una memoria interna. El monitor con una contraventana de cristal líquido polariza las dos imágenes sobrepuestas. Las imágenes izquierda y derecha llegan al ojo correcto con la ayuda de unas gafas polarizadas que completan la separación entre las imágenes completando de esta manera el despliegue de una imagen estereoscópica.

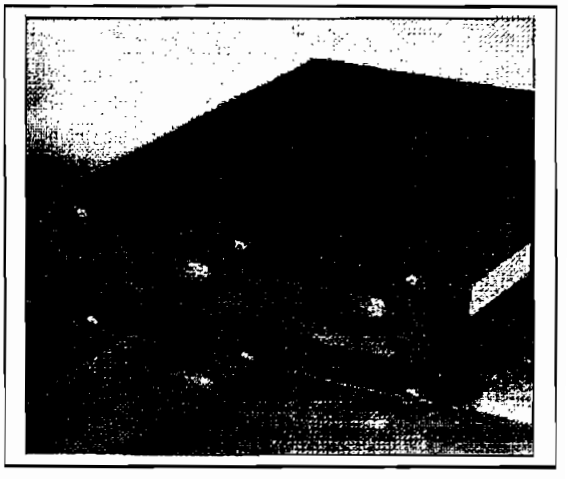
5.1.1 CÁMARA ESTEREOSCÓPICA A COLOR CF 23:



Los datos técnicos de la cámara versión PAL (NTSC) son:

- Dispositivos de adquisición: Dos censores CCD con transferencia de interlínea e integradores de filtro de mosaico.
- Número de píxeles:752(H) x 582(V) efectivos (768(H) x 494(V) efectivos)
- Resolución horizontal: >450 líneas de TV.
- Sensitividad a la luz: 10 lux (9.5 lux)
- Señal de salida: compuesta de video o Y/C (S-VHS) conmutable, 1Vpp, a 75 ohm.
- Lentes montados: 2 x cada cámara.
- Estereo básico: 55 mm
- Interruptor de cámara: izquierdo, derecho, estereo
- Dimensiones: 130 x 50 x 110 mm
- Peso aprox. 390 g
- Voltaje/ corriente: 12VDC, 410 mA.

5.1.2 CÁMARA ESTEREOSCÓPICA CON ZOOM CF 44:



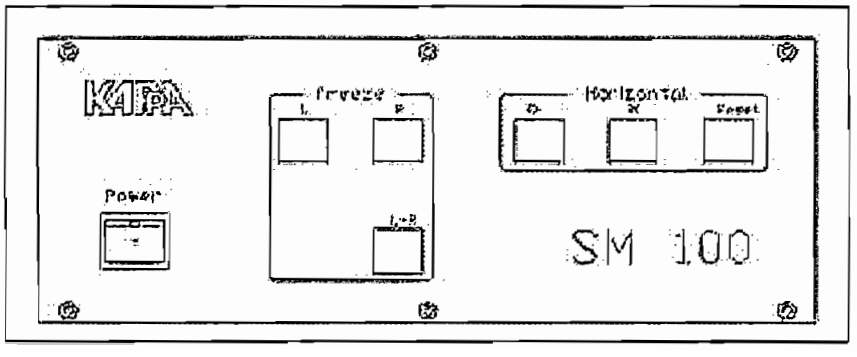
Esta cámara de video 3-D tiene convergencia motorizada y enfoque automático integrados dentro de una caja protectora. La unidad entrega imágenes observadas sobre un gran rango de distancia sin presentar esfuerzo para el observador. Equipo conveniente para el control remoto de vehículos, manipulación remota y macroscopía 3D. Todas estas funciones pueden ser controladas por computadora mediante una interface RS232.

Sus datos técnicos son:

- Dispositivos de adquisición: sensores CCD con 1/3" de ineterlineado, filtro de mosaico de color, micro lentes.
- Número de píxeles:752(H) x 582(V) efectivos o 768(H) x 494(V) efectivos
- Resolución horizontal: >470 líneas de TV.
- Salida de video: video compuesta y Y/C (S-VHS)
- Iluminación mínima: 6 lux (F1.4)
- Óptica: motor para Zoom 8x
- Características especiales:

  muy compacta, lentes de Zoom motorizados con enfoque automático selectivo. Las funciones de la cámara pueden ser controladas remotamente.

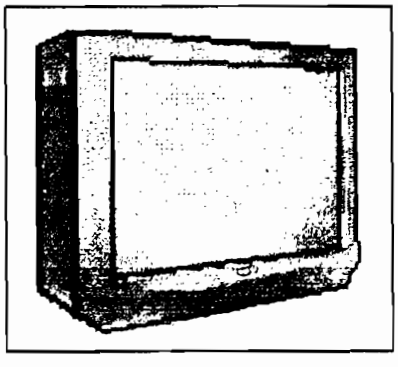
5.1.3 CONVERSOR DE BARRIDO: SM100



Para obtener imagen 3D sin parpadeo, es necesario una alta frecuencia de despliegue de imágenes, mayor que la disponible en los estándares de televisión convencional. La mejor manera de lograr esto es duplicar la frecuencia básica para obtener la misma velocidad de 50Hz (60 Hz) para cada ojo. Por esta razón el SM100 duplica la frecuencia de la imagen entrante a 100 Hz (PAL) o 120 HZ (NTSC) para obtener una imagen 3D libre de parpadeo, especial para aplicaciones profesionales como conducción de vehículos, robots, manejo remoto de sustancias peligrosas o tareas de inspección. Siendo posible trabajar por horas sin dolor de cabeza al que induce normalmente el efecto de parpadeo de la imagen.

El sistema acepta entrada de señales de video PAL y NTSC las cuales pueden ser de video compuesto, Y/C (super-VHS) o RGB. Este sistema trabaja con alimentación de 110 V/60Hz o 220 V/50Hz.

5.2 MONITORES 3D LIBRES DE PARPADEO

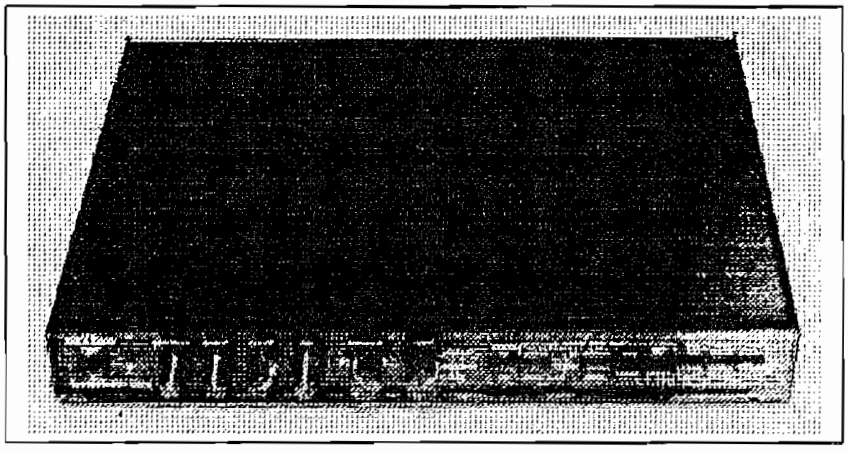


Multiestándar: NTSC, PAL, SECAM. 110/220V.
Frecuencia de despliegue: 100 Hz/PAL, 120 Hz/ NTSC.
Incluyen dos pares de gafas inalámbricas.

	29"	34"	38"
Modelos	3DTV29	3DTV34	3DTV38

Se ofrece una lista de precios en la página web:
www.3dmagic.com/catalog/price_list/price_list.html#TV

5.3 CONVERSOR DE IMÁGENES 2D / 3D SOLIDIZER PRO™



Solidizer Pro™ - este es un conversor de video 2D a 3D en tiempo real.
Entrada: estándar NTSC compuesta o S video.
Salida: estándar NTSC compuesta o S video, RGB.

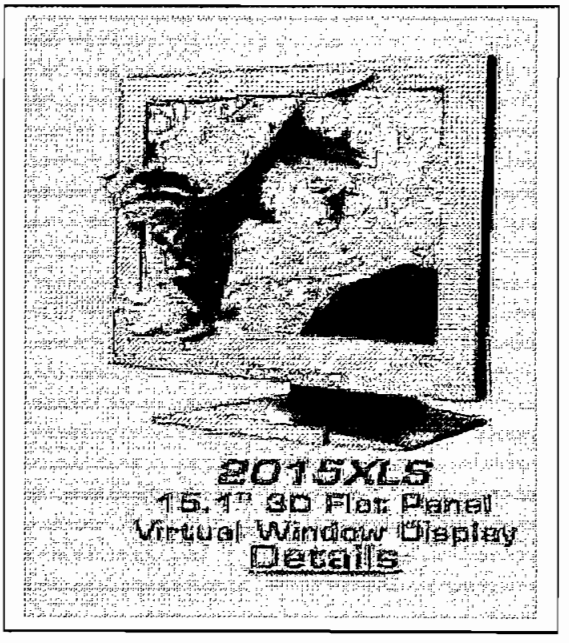
El campo o cuadro secuencial que produce a su salida puede ser visto con gafas shutter LCD o HMDs (head mounted displays tales como gafas o modelos Sony Glasstron Stereo). Así como con video proyectores dobles o 3DTV de pantalla estereo, la salida puede ser vista con gafas polarizadas. Con la adición de un sistema transcodificador SpaceSpex™ y el uso de gafas anaglifas, la salida puede ser vista a color. Su precio es de $22,000.

Para especificaciones adicionales se puede consultar el manual que esta en formato Word de la siguiente dirección: www.3dmagic.com/pdf/solidpro.doc

También disponible como versión solo para PC con total control sobre el software de conversión a un costo de $ 25,000.

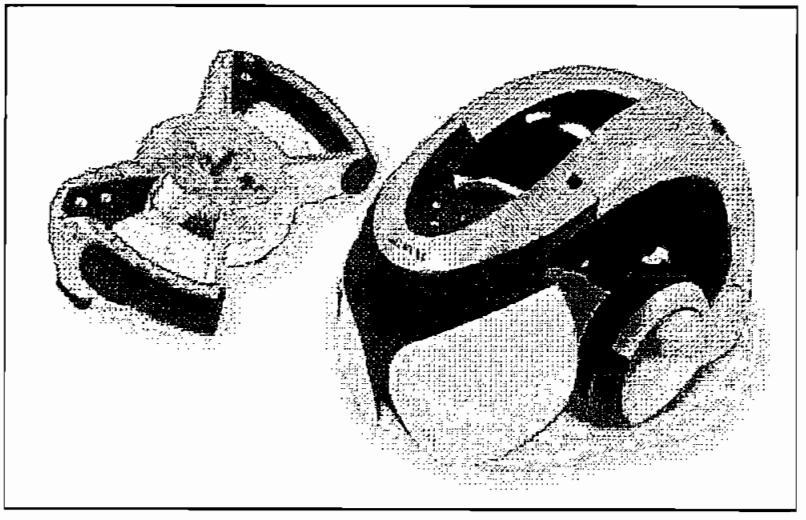
5.4 DISPLAY AUTOESTEREOSCÓPICO DE 15"

Este display es creado por DTI (Dimension Technologies Inc) tiene un precio en el mercado de $1,699. El display DTI soporta todos los formatos estereoscópicos comunes (4 en 1), pudiendo trabajar con virtualmente todas las aplicaciones estereo imágenes y animaciones.



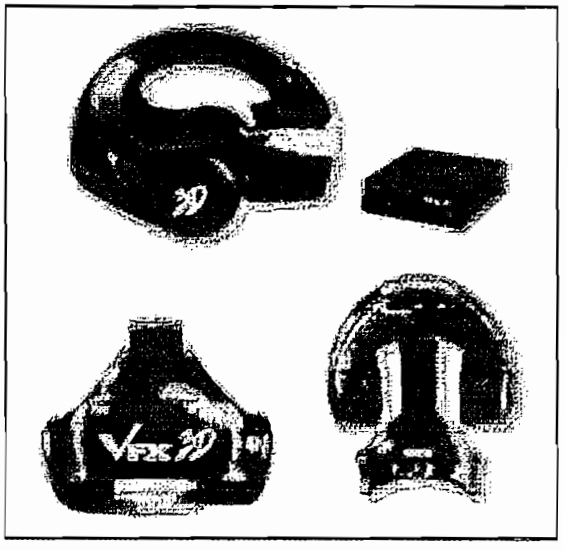
5.5 CASCOS ESTEREOSCÓPICO INALÁMBRICO

5.5.1 GLOBAL PLAYER



Este es un prototipo que Siemens saco al mercado en el 2001 conocido como el 'Global Pl@yer', su principal logro es proveer al usuario un ambiente estereoscópico con la mayor libertad posible, muy utilizado en juegos y simulaciones de vuelo.

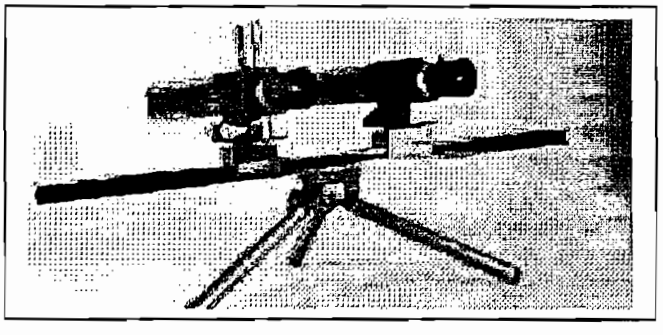
5.5.2 CASCO VFX3D



El casco estereoscópico de Interactive Imaging Systems, Inc. mejor conocido como VFX3D es un equipo de alto rendimiento en tiempo real. Utiliza un software que no necesita de ningún otro dispositivo especial de realidad virtual y además provee estereoscopia y rastreo de cabeza con tres grados de libertad para rotación horizontal, vertical e inclinación, este producto ha sido premiado en diversos campos que incluyen entrenamiento, educación diversión, arquitectura, industria y mucho más. Utiliza una interface con estándar VGA, entradas de audio y displays de color con 360,000 píxeles.

5.6 PRODUCTOS VREX

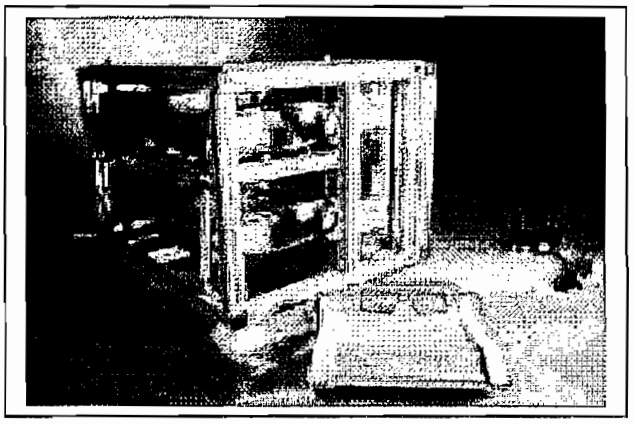
5.6.1 CÁMARA ESTEREOSCÓPICA CAM-4000



Combina perfectamente los rasgos de una cámara profesional o semi-profesional con las cualidades de una cámara estereoscópica. La CAM 4000 es producida por la compañía VREX incluye Zoom sincronizado, enfoque y apertura de iris. Su precio es de $7,495.00 más gastos de envío.

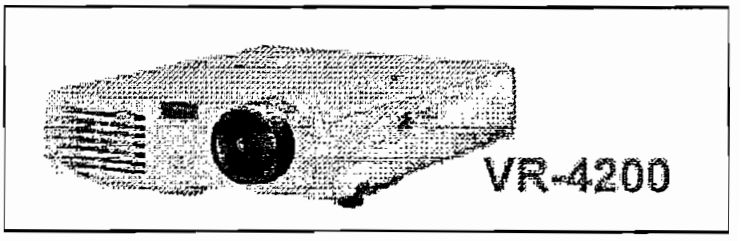
5.6.2 PROYECTORES 3D

La empresa VREX produce tres de los más conocidos proyectores estereoscópicos conocidos como el VR-Dual 1000, VR-4200 y VR-3100

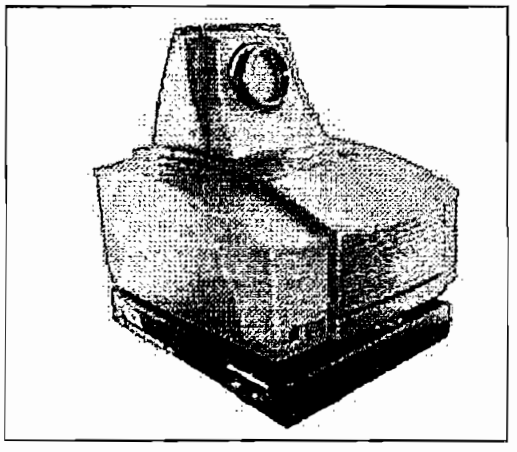


El VR-Dual 1000, tiene una resolución XGA (1024 x 768) y una mejora de brillo (2000 lúmenes por proyector). El VR-Dual 1000 está disponible también con una caja conversora XPO2

El proyector VR-4200 es el primer proyector digital estereoscópico portable basado en un simple chip con tecnología DLP™ que permite una mejora en la resolución y brillo.



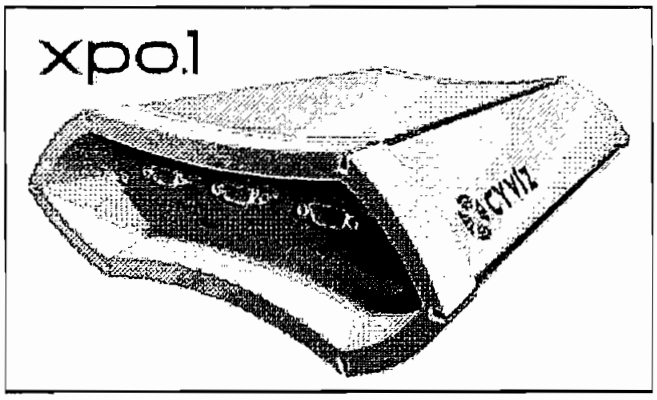
El VR-4200 ofrece una resolución en la imagen SVGA de 1024 x 768 píxeles, su precio en el mercado es de $15,995.



El VR-3100 es un proyector 3D económico basado en la tecnología uPOL™ que ofrece una resolución total de 800 x 600 tiene un peso de aproximadamente 10 lb, brillo de 350 lúmenes en una habitación con iluminación normal y su precio es de $9,995.

Cabe anotar que uPOL™ (pronunciado micropol) es la única tecnología patentada por VREX. El uPOL es un dispositivo óptico que cambia la polarización de la luz en una base línea por línea. El uPOL se produce con un proceso patentado usando avanzadas técnicas de micro fabricación.

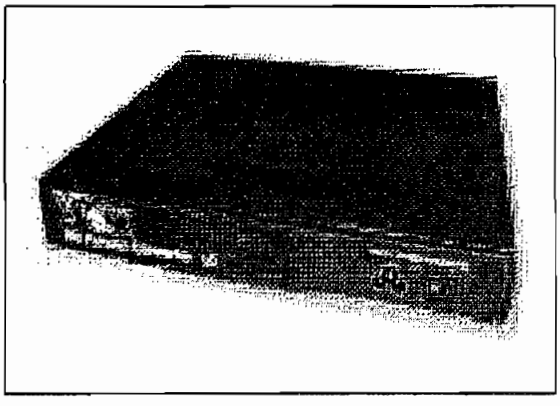
5.6.3 CONVERSOR ESTEREO XPO



El conversor estereoscópico XPO permite visualizar imágenes estereoscópicas usando un estándar normal de LCD o DLP y un PC regular.
Lo que proporciona una solución flexible a un bajo costo.

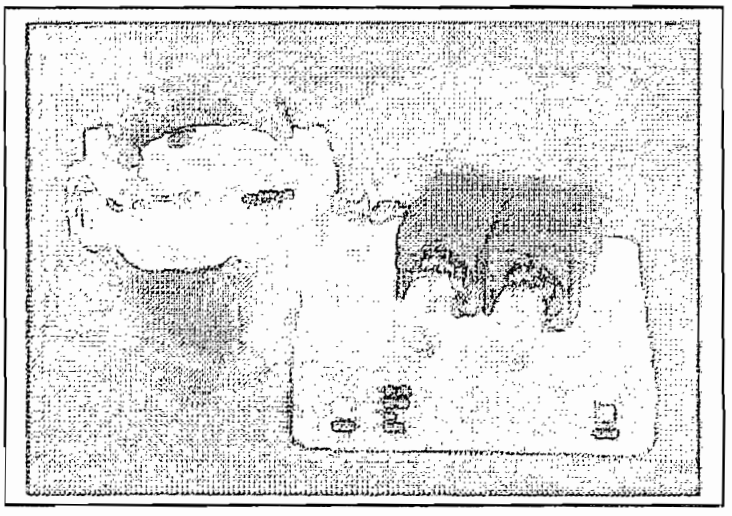
El XPO es sumamente flexible ya que trabaja con fuentes de video y proyectores independientes, lo que hace que sean compatibles con el estándar de salida de señales estereo de cualquier computador y pueden desplegarse con cualquier tipo de proyector 3D sin tener en cuenta la marca. Posibles futuras áreas de uso son displays montados en la cabeza (HMD), proyectores lado a lado, con proyección enfrente y detrás con pantallas de 500". El precio del **XPO.1** es de $8,500.

5.6.4 CONVERSOR DE VIDEO VR



Este es un demultiplexor de campo secuencial de video que puede tener conectado a su salida un proyector doble. El sistema es económico y proporciona características adicionales como ajuste horizontal y vertical de la imagen. Entre sus muchas aplicaciones, el conversor VR-video es usado para grandes y pequeñas presentaciones de video en eventos. Su costo es de $4,000.

5.7 PRODUCTOS DE VIDEO ESTEREOSCÓPICO DE 3-D IMAGE TEK CORP.

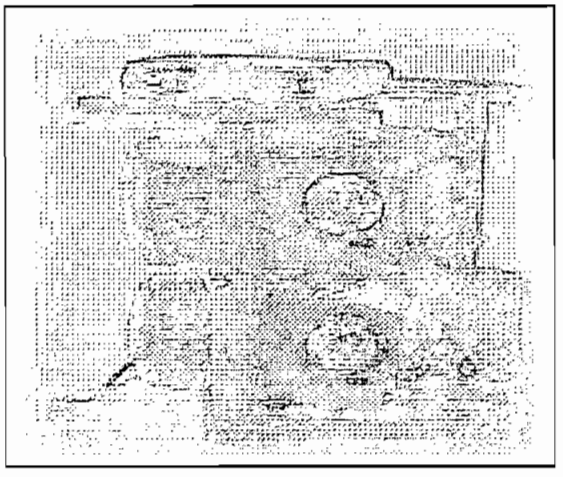


La compañía Image Tek Corp, presenta como su equipo a la venta mas representativo a:

3D-Video encoder: que es codificador de video 3D, cuya función es la de tornar dos señales de video en una señal de campo secuencial para HMD o VCR.
3D-Video decoder: el decodificador de video 3D, que tiene como función el cambiar una señal de campo secuencial en dos señales de video para ser mostradas por medio de proyectores dobles o HMD.
3D- Video encoder/decoder: el cual tiene ambas funciones anteriores en un solo dispositivo.

En el gráfico anterior se muestra el Codificador de video 3D, las cámaras y el HMD son opcionales

3DI TELEJECTOR™: compuesto por un decodificador y dos proyectores prepolarizados en un solo rack.



Aquí se muestra el TELEJECTOR serie 2000. muy utilizado en aplicaciones de :

- Educación
- Medicina
- Entretenimiento
- Simulacón
- Tele-robótica

Modelo #	Descripción	Precio
VIDEO ENCODERS / DECODERS		
Stereo Video 3DI-3000 Encoder	Codificador de video estereoscópico	$2595
Stereo Video 3DI-3000 Decoder	Decodificador de video estereoscópico	$2595
Stereo Video Encoder/Decoder Combination	Codificador/Decodificador de video estereoscópico	$3595
3DI Telejector Series 2000	Sistema de video proyección estereoscópico 3D	$24,995

CAPÌTULO VI

6. CONCLUSIONES Y RECOMENDACIONES

- En Esta tesis se pretende dar una visión global en el tratamiento de imágenes estereoscópicas, poniendo énfasis en mejora y compresión de imágenes, tratamiento de secuencias de imágenes en movimiento, y despliegue de las mismas.

- El estudio que hemos realizado resalta la importancia que tiene la Televisión Digital en todas sus aplicaciones, una de ellas la Televisión Estereoscópica que nos permite tener detalles que con la televisión convencional no se tiene, es decir visualizar las imágenes con profundidad y poder tener una idea real de las distancias de los objetos.

- En países donde la Televisión Digital esta muy desarrollada, la Televisión Estereoscópica tiene varios campos de aplicación como son: Medicina, Topografía, Ingeniería Molecular, Realidad Virtual, entre otros, en los cuales tener una buena apreciación de profundidad y volumen es de vital importancia, es por esto que esta técnica ha despertado gran interés.

- El presente trabajo presenta una alternativa para actualizar y dar a conocer los beneficios que se tienen utilizando la Televisión Estereoscópica y el fundamento teórico ayuda con los conocimientos básicos para personas que se interesen en el tratamiento digital de imágenes.

- Un sistema de Televisión Estereoscópica basado en las dos señales de ojo derecho y ojo izquierdo no debe de causar problemas en comparación con los sistemas de televisión monoscópica convencional,

como fátiga visual, parpadeo o el efecto de marionetas, y debería contener las medidas dirigidas a corregir dichas molestias.

- Se pretende desarrollar este sistema de televisión, de tal forma que se tenga la mayor compatibilidad posible con sistemas de televisión monoscópica ya existentes, y que la velocidad de transmisión adicional fuera la mínima posible .

- El sistema de Televisión Estrereoscópica presenta mayor calidad de definición que los actuales sistemas de televisión convencional.

- En esta tesis se expone el tratamiento de la imagen con un medio para obtener un uso eficiente de recursos, es decir extraer la información relevante de forma que se ocupe el mínimo ancho de banda posible en aplicaciones de transmisión y/o almacenamiento de imágenes.

- Cabe mencionar que para desarrollar los actuales sistemas de televisión estereoscópica se han realizado numerosas pruebas que han dado como uno de los principales resultados de que para que exista estereopsis visual no es necesario que la imagen que ven los ojos tengan la misma definición, pudiendo una de las dos imágenes ser de menor calidad que la otra, lo cual se traduce en una reducción del ancho de banda del canal de transmisión.

- Como una de las metas finales de la transmisión de una señal estereoscópica esta el poder mostrar sobre una pantalla imágenes que puedan ser vistas independientemente por cada ojo sin necesidad de utilizar implementos adicionales, por lo cual se han ideado varías soluciones muy ingeniosas que han sido puestas a consideración en este trabajo.

- Se ha llegado a considerar que el salto de la televisión convencional a lo que sería la televisión estereoscópica, tiene la misma o mayor importancia de lo que fue a su tiempo el cambio de la televisión en blanco y negro a lo que hoy es la televisión a color ya que con el constante avance tecnológico se pretende ofrecer al público televidente la mayor sensación de realidad virtual en sus hogares.

- Los sistemas de Televisión Estereoscópica que se desarrollen deben de cumplir con una serie de pruebas que sirven para evaluar las imágenes estereoscópicas y que se detallan en las recomendaciones de la UIT-R citadas en este trabajo .

- Se recomienda realizar estudios más profundos debido a que esta temática es una tecnología que se esta desarrollando, y tiene un amplio campo de estudio, es por esto que el estudio de esta tesis podrá servir como introducción para futuros trabajos en el tratamiento digital de imagen.

- Uno de los procedimientos que más ha aportado en el desarrollo de este tipo de tecnologías es la compresión, es por esto que debe tener posteriores estudios para su análisis y discusión.

- Se recomienda que los estudiantes de la facultad deberían de realizar prácticas profesionales en los principales estudios de televisión del país, para que de esta forma estén al tanto con las nuevas tecnologías que se implementan en este campo.

- Los principales centros de televisión nacional, deberían de ir cambiando de tecnología e ir implementando las nuevas técnicas existentes en el

campo de información visual, pero debiendo enmarcarse en los estándares internacionales que rigen la tecnología de Televisión.

BIBLIOGRAFÍA

- UNIÓN INTERNACIONAL DE TELECOMUNICACIONES, Recomendación UIT-R BT.2017, 1998.

- UNIÓN INTERNACIONAL DE TELECOMUNICACIONES, Recomendación UIT-R BT.1202, 1995.

- UNIÓN INTERNACIONAL DE TELECOMUNICACIONES, Recomendación UIT-R BT.2018, 1998.

- UNIÓN INTERNACIONAL DE TELECOMUNICACIONES, Recomendación UIT-R BT.601-5, 1995.

- UNIÓN INTERNACIONAL DE TELECOMUNICACIONES, Recomendación UIT-R BT.1438, 2000.

- UNIÓN INTERNACIONAL DE TELECOMUNICACIONES, Recomendación UIT-R BT.1198, 1995.

- UNIÓN INTERNACIONAL DE TELECOMUNICACIONES, Recomendación UIT-R BT.500-7.

- UNIÓN INTERNACIONAL DE TELECOMUNICACIONES, Recomendación UIT-T H.262, 1995.

- UNIÓN INTERNACIONAL DE TELECOMUNICACIONES, Recomendación UIT-T H.261, 1993.

- UNIÓN INTERNACIONAL DE TELECOMUNICACIONES, Recomendación UIT-T H.263, 1998.

- FAÚNDEZ Marcos, Tratamiento digital de voz e imagen, Alfaomega grupo editor, S.A.,2001

- Andrew J. Woods, John O. Merritt, Stephen A. Benton, Scott S. Fisher, Mark T. Bolas, Stereoscopic Displays and Applications, Recopilación de papers Vol 1 y 2. Spie org. 2001

- GONZALES C. , RICHARD E., Tratamiento digital de imágenes, Addison-Wesley Iberoamericana,S.A.,1996

- WATKINSTON, . Compresión in video & audio, British Library, 1995

- HERNÁNDEZ Oliver, descripción del estandar MPEG-2, Universidad Central de Venezuela, Escuela de Ingeniería Eléctrica, 1998

- CEPEDA Carmen, TERÁN Miriam, Diseño de un sistema integrado para estudios de estaciones del servicio de radio difusión de televisión digital, EPN – FIE, 2000

Direcciones de Internet:

www.3d-web.com/

www.spie.org/

www.3dmagic.com/catalog/price_list/price_list.html#TV

www.3dmagic.com/catalog/solidizerpro.html

www.stereographics.com/html/body_stereo_formats.html

www.paralax.com.mx/09a_Estereoscopia.html

www.users.red3i.es/~stereoweb/

www.paralax.com.mx/09a_estereoscopia.html

www.users.red3i.es/~stereoweb/historia.htm

www.users.red3i.es/~stereoweb/aplica.htm

www.tav.net/3d/

www.ejezeta.com

http://verona.fi-p.unam.mx/fardi/pagina/ESTEROS.htm

www.stereoscopy.com/

ANEXOS

ANEXO 1
LISTA DE ABREVIACIONES

LISTA DE ABREVIACIONES

ATM: Modo de transferencia asincrónico

ATSC: Advanced Television Systems Comittee

bpp: bits por pixel.

CAD: Diseño Asitido por Computador.

CAE:

COFDM: Coded Orthogonal Frequency Division Multiplexing.

COST: European **Co**operation in the **S**cientific and **T**ecchnical field

CRT: Tubo de rayos catódicos

DBS: Segmentación basada en disparidad.

DCP: Predicción compensada en disparidad.

DCT: Transformada del coseno discreta.

DISTIMA: **Di**gital **St**ereoscopic **Im**aging & **A**pplications

DPCM: (**D**ifferential **P**ulse **C**ode **M**odulation). Modulación de código de pulsos diferencial.

DVB: Digital Video Broadcasting System.

DVB-C: Sistema de difusión de video digital por cable.

DVB-S: Digital Video Broadcasting by Satellite.

DVB-T: Sistema de difusión de video digital terrestre.

DVB-MC/S: Sistema de difusión de video digital multipunto por microondas.

ETSI: Instituto Europeo para Normalización de las Telecomunicaciones.

FBS: (**F**ixed **B**lock **S**ize) Tamaño de bloque fijo.

FEC: Forward Error Correction.

HBM: (**H**ierarchical **B**lock **M**atching). Emparejamiento de bloque jerárquico.

HDTV: Televisión digital de alta definición.

HMD: **H**ead **M**ounted **D**isplay

HVS: Sistema visual humano.

IBCN: **I**ntegrated **B**roadband **C**ommunicate **N**etwork

ISDB: (**I**ntegral **S**ervice **D**igital **B**roadcasting). Transmisión digital de servicio integral.

LBG: Algoritmo **L**inde, **B**uzo, **G**ray.

LCS: Liquid **C**rystal **S**hutter glasses.

LCD: Liquid **C**rystal **D**isplay glasses.

MAD: (**M**inimun **A**bsolute **D**ifference). Diferencia mínima absoluta.

MAE: (**M**ean **A**bsolute **E**rror). Error absoluto medio.

ME: (**M**otion **E**stimation). Estimación del movimiento.

MF: (**M**odel **F**ailure). Modelo de fallo.

MIRAGE: Manipulation of Images in Real-time for the Creation of Artificially Generated Environments

MPEG:

MR: Multiresolución

NTSC: National Television Sistems Committee

PAL: Phase Alternating Line

PANORAMA: Package for New Operational Autostereoscopic Multiview Systems and Applications

QAM: Modulación de amplitud en cuadratura

QPSK: Quadrature Phase Shift Keying

RDBS: (Reversed **DBS**) extensión de codificación de secuencia estereoscópica usando DBS inversa. Configuración-1.

SECAM: Séquentiel Couleur A. Mémoire

SFN: Single Frequency Networks

S-MPEG: Estéreo MPEG

SQ: Quantizador escalar.

ST-1: Extensión de Codificación de Secuencias Estereoscópicas Rastreador de segmento- configuración 1. (**S**egment **T**racking).

VCR: Video Camera Recorder

VISIDEP: (**Vis**ual **I**mage **D**epth **E**nhancement **P**rocess) Proceso de perfeccionamiento visual de imagen y profundidad.

VLC: (**V**ariable **L**ength **C**ode). Código de longitud variable

VQ: (**V**ector **Q**uantization). Vector de cuantización.

VSB: Banda lateral vestigial

ANEXO 2
VOCABULARIO TÉCNICO BILINGÜE

VOCABULARIO TECNICO BILINGÜE

TERMINO ORIGINAL EN INGLES	TERMINO USADO EN LA TESIS	VOCABLOS ALTERNATIVOS DE USO COMUN
Aliasing	Aliasing	Alias, aliasing
Baseline	Línea base	Distancia
Bits per pixel	Bits por píxel	
Broadcast	Transmisión	Transmisión, emisión
Coarsest	Menor resolución	Tosco, rustico.
Convolution	Convolución	
Discrete Cosine	Transformada del Coseno	Transformada
Edge	Borde	
Entropy	Entropía	
Frame	Cuadro	Trama, marco,
Framework	Estructura	Armazón, esqueleto
Headtracker	Seguidor de cabeza	
Matching	Emparejamiento	Igualación
Motion copensated	Predicción compensada en	
Multiresolution	Multiresolución	
Neighboor	Vecino	Contiguo, colindante
Overlap	Traslapar	Sobreponer, recubrir.
Pel	Píxel	
Picture element	Píxel	Pel
Quadtree	Quadtree	Arbol cuádruple
Redundancy	Redundancia	
Resolution	Resolución	
Restoration	Restauración	
Sampling	Muestreo	Muestreado
Segmentation	Segmentación	
Shutter	Obturador	
Smoothing	Suavizado	Alisado uniforme
Stereo imaging	Imagen Estereoscópica	Imagen estéreo
Stereoscopic	Estereoscopia	Stereoscopy
Subsampling	Submuestreo	
Texture	Textura	
Threshold	Umbral	Comienzo, principio
Tree	Arbol	
Upsampling	Sobremuestreo	
Vector quantization	Vector de cuantización	
Wavelets	Wavelets	Ondículas
Zoom	Zoom	Acercamiento

ANEXO 3
LISTA DE FIGURAS

LISTA DE FIGURAS

ANEXO 4
LISTA DE TABLAS

Lista de tablas:

ANEXO 5
RECOMENDACIONES DE LA UIT

RECOMENDACIÓN UIT-R BT.1438

EVALUACIÓN SUBJETIVA DE LAS IMÁGENES DE TELEVISIÓN ESTEREOSCÓPICA

(Cuestión UIT-R 234/11)

(2000)

La Asamblea de Radiocomunicaciones de la UIT,

considerando

a) que se están realizando estudios para desarrollar la televisión estereoscópica como un potencial futuro servicio de radiodifusión;

b) que a los efectos de la televisión estereoscópica, basada en dos señales, el canal del ojo izquierdo y el canal del ojo derecho respectivamente, se ha adoptado la Recomendación UIT-R BT.1198;

c) que la evaluación subjetiva es un elemento vital en el diseño e introducción de los sistemas de televisión estereoscópica;

d) que las condiciones de filmación, de visualización y el tipo de pantalla pueden influir sobre la fatiga del observador;

e) que deberían establecerse condiciones de evaluación comunes adecuadas para los sistemas de televisión estereoscópica; que estas condiciones deberían incluir los métodos de evaluación, las condiciones de filmación, las condiciones de visualización, así como los materiales de prueba a utilizar en la evaluación y en los métodos de cribado para asegurar que los observadores tienen una percepción de profundidad normal,

recomienda

que se utilicen las condiciones descritas a continuación para la evaluación subjetiva de sistemas de televisión estereoscópica.

1 Factores de evaluación

Los factores de evaluación que normalmente se aplican a las imágenes de televisión monoscópica tales como resolución, representación del color, representación del movimiento, calidad general, nitidez de perfiles, profundidad, etc. pueden también aplicarse a los sistemas de televisión estereoscópica. Además, existen numerosos factores que son específicos de los sistemas de televisión estereoscópica. Aunque algunos de ellos se enumeran a continuación, es necesario realizar estudios adicionales para identificar otros y para establecer las definiciones físicas.

– *Resolución en profundidad*

  Resolución espacial en profundidad. Una resolución reducida en profundidad puede reducir la calidad de la imagen de la televisión estereoscópica.

– *Movimiento en profundidad*

  Factor que establece si el movimiento en el sentido de la profundidad se reproduce sin discontinuidades.

– *Efecto teatro de marionetas*

  Describe un tipo de distorsión en imágenes 3-D. A veces, los objetos estereoscópicos se perciben como anormalmente grandes o pequeños.

– *Efecto papel de cartón*

  Describe otro tipo de distorsión en la reproducción de imágenes 3-D. Las posiciones 3-D de objetos estereoscópicos se perciben de manera estereoscópica, pero éstos se parecen ser anormalmente delgados.

2 Métodos de evaluación

Los métodos que se describen en la Recomendación UIT-R BT.500 pueden utilizarse para evaluar la calidad general de imagen de los sistemas estereoscópicos, así como la nitidez y profundidad de la imagen (véase el Anexo 2). Si se dispone de una imagen de referencia, puede utilizarse el método de escala de calidad continua de doble estímulo o el método de escala de degradación de doble estímulo. Constituyen ejemplos de ello la comparación de los sistemas de visualización,

la evaluación de la calidad de los sistemas de codificación y otros. Si no se dispone de ninguna referencia, puede utilizarse el método de juicio categórico para identificar los méritos de los sistemas estereoscópicos. Los métodos de evaluación de factores específicos de los sistemas de televisión estereoscópica requieren estudios adicionales.

3 Condiciones de visualización

Deben tenerse en cuenta dos factores principales que son específicos de la representación estereoscópica, a saber, el efecto del marco de visualización y la inconsistencia entre acomodación y convergencia.

Las imágenes estereoscópicas parecen poco naturales cuando los objetos que se encuentran delante de la pantalla se acercan al marco de la misma. Este efecto antinatural se denomina efecto marco. Este efecto se reduce normalmente con pantallas grandes debido a que los observadores son menos conscientes de la presencia del marco cuando la pantalla es grande.

El ojo humano se enfoca sobre un objeto en función de la distancia al mismo. Al mismo tiempo, también se controla el punto de convergencia (punto de enfoque o de visión) sobre el objeto. Por lo tanto, en nuestra vida cotidiana no existe inconsistencia entra acomodación y convergencia. Sin embargo, cuando visualizamos imágenes estereoscópicas, el punto focal (acomodación) se fija siempre en la pantalla, con independencia el punto de convergencia que se obtiene de la disparidad de las señales. Dicho de otra forma, el observador no enfoca claramente. Por lo tanto, en los sistemas estereoscópicos se presenta una inconsistencia entre acomodación y convergencia.

Es algo generalmente aceptado que el valor mínimo de profundidad de campo del ojo humano es de ±0,3 D (dioptrías; valor inverso de la distancia (m)) [Hiruma y Fukuda, 1990]. Ello significa que puede percibirse la imagen sin desenfoque cuando el objeto se encuentra situado en el margen de ±0,3 D. Cuando se visualiza la televisión estereoscópica, el punto de acomodación permanece fijo en la pantalla y, por tanto, la imágenes estereoscópicas deben visualizarse preferentemente dentro de dicha gama. Dado que los programas ordinarios de televisión incluyen imágenes a una distancia infinita (es decir, $D = 0$), se considera que la gama deseable de profundidad que debe visualizarse con sistemas estereoscópicos se encuentra en el rango de 0 a 0,6 D. Por lo tanto, se considera que la distancia de visualización óptima es 0,3 D, es decir, 3,3 m.

Los parámetros de la cámara (separación de la cámara, ángulo de convergencia de la cámara, longitud focal de las lentes), la resolución del sistema y el efecto marco se deben tener en cuenta para determinar las condiciones de visualización (tamaño de la pantalla). En el caso de TVAD, cuando se mira a la distancia de visualización normalizada de 3 H (H es la altura de la pantalla), la distancia de 3,3 m. corresponde a una pantalla de 90 pulgadas (229 cm). En el caso de la televisión definición convencional (TVDC), cuando la distancia de visualización es la normalizada de 6 H, dicha distancia se corresponde con una pantalla de 36 pulgadas (91 cm). Utilizando un sistema de TVAD estereoscópico se realizó una evaluación subjetiva de la relación entre el tamaño de la pantalla y la percepción de profundidad, resultando que la percepción de profundidad más natural se obtuvo con una pantalla de 120 pulgadas (305 cm), que se corresponde con una distancia de visualización de 2,2 H [Yamanoue y otros, 1997].

4 Observadores

Los observadores deben gozar de una visión de agudeza normal (véase la Recomendación UIT-R BT.500). Además, deben tener una estereopsis normal. Para verificar la estereopsis, puede utilizarse el material de prueba que figura en el Anexo 1.

5 Materiales de prueba

En el Anexo 1 se enumeran el material de prueba utilizado con los observadores así como las secuencias estáticas o en movimiento de escenas naturales.

Los efectos en 3-D que se consiguen con las imágenes estereoscópicas dependen en gran medida de las condiciones de filmación, tales como la separación entre cámaras, el ángulo de convergencia de las cámaras y la longitud focal de las lentes. Las secuencias en movimiento fueron filmadas con una separación entre cámaras de 65 mm, que se corresponde con la separación media entre ojos, y la mayoría de ellas fueron producidas en condiciones de cámara no cruzada, lo cual permite disponer de condiciones ortoestereoscópicas [Yamanoue y otros, 1998].

REFERENCIAS BIBLIOGRÁFICAS

HIRUMA, N. y FUKUDA, T. [diciembre de 1990] Accomodation response to binocular stereoscopic TV images and their viewing conditions. *J. SMPTE*, 102, 12, p. 2047-2054.

YAMANOUE, H. y otros [octubre de 1997] Subjective study on the orthostereoscopic conditions for 3-D HDTV. ITE Tech. Report, Vol. 21, 63, p. 7-12.

YAMANOUE, H. y otros [1998] Orthostereoscopic conditions for 3-D HDTV. *Proc. SPIE*, 3295, *Stereoscopic displays and Applications IV*.

ANEXO 1

Material de pruebas para la evaluación subjetiva de imágenes de televisión estereoscópica

1 Prueba de visión

En el Cuadro 2 se enumeran las cartas o diagramas de prueba para las pruebas de visión. Las 12 cartas se han seleccionado de acuerdo a la jerarquía del sistema visual humano, desde los niveles inferiores a los superiores. Se describen a continuación ocho pruebas de visión (VT, *vision tests*) principales, quedando las otras cuatro para pruebas clínicas. Los observadores deben tener una estereopsis normal, es decir, deben haber superado la prueba VT-04 para la estereopsis fina y VT-07 para la estereopsis dinámica. Las seis pruebas restantes sirven para una caracterización más detallada. Las cartas de prueba deben ser visionadas a una distancia igual a tres veces la altura de la pantalla de visualización (3 *H*).

Las imágenes en miniatura situadas a derecha e izquierda se colocan una junto a otra con fines explicativos para una fusión sin cruce.

a) *VT-01*: Percepción simultánea (prueba del león)

Prueba la capacidad de percibir simultáneamente imágenes presentadas dicópticamante y en la posición correcta. Se presenta la imagen de una jaula en un ojo y la de un león en el otro, cuya posición se desplaza a razón de 12'/s. El tamaño de cada imagen se fija a 10°, de tal forma que los observadores pueden capturar las imágenes en sus paramáculas. Los observadores con una visión normal pueden ver al león dentro de la jaula durante un cierto tiempo del periodo de presentación.

FIGURA 1

Diagrama de prueba para VT-01

Imagen derecha

Imagen izquierda

1438-01

b) *VT-02*: Fusión binocular (prueba de los 4 puntos de Worth)

Prueba de la capacidad de percibir dos imágenes dicópticas en los ojos izquierdo y derecho como una sola imagen. La imagen para un ojo tiene dos puntos y la del otro ojo tiene tres puntos, con un punto común. Los observadores con una visión normal ven 4 puntos.

FIGURA 2

Diagrama de prueba para VT-02

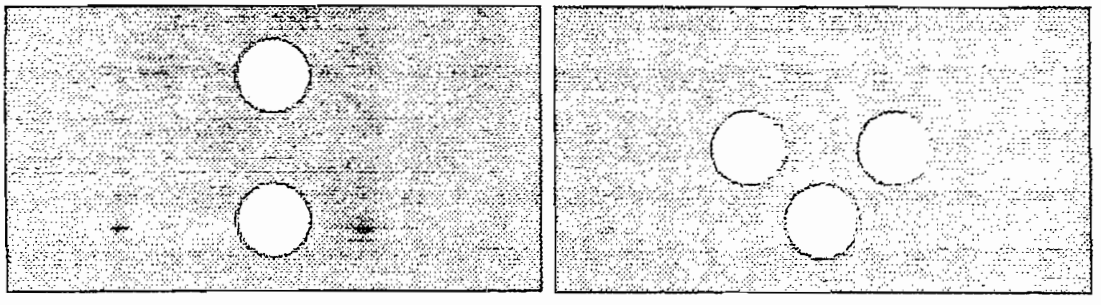


Imagen derecha Imagen izquierda

1438-02

c) *VT-03*: Estereopsis gruesa (prueba de la libélula)

Prueba de la capacidad de percibir imágenes que se presentan de forma dicóptica con un cierto paralaje como una sola imagen con una profundidad considerable. Las imágenes para ambos ojos son una estereopareja de imágenes de una libélula con sus alas extendidas. Los observadores con una visión normal perciben las alas delante de la pantalla de visualización.

FIGURA 3

Diagrama de prueba para VT-03

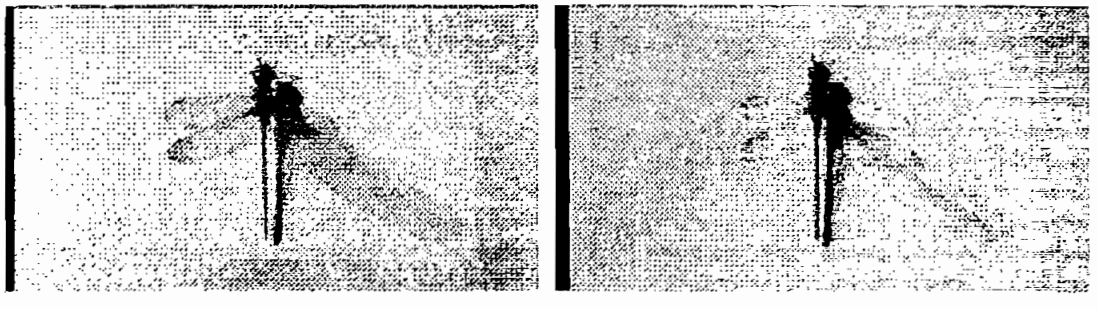


Imagen derecha Imagen izquierda

1438-03

d) *VT-04*: Estereopsis de detalle (prueba del círculo)

Prueba la capacidad de percibir imágenes que se presentan de forma dicóptica con un cierto paralaje como una sola imagen con una profundidad reducida. Se presentan nueve romboides de prueba, cada uno de los cuales tiene cuatro círculos, y sólo uno de los círculos tiene un pequeño paralelaje. Los observadores con visión normal pueden percibir el círculo con el pequeño paralelaje delante de la pantalla de visualización. El Cuadro 1 muestra el número de prueba, las respuestas correctas y el ángulo de estereopsis a 3 *H*.

CUADRO 1

Respuestas correctas y paralelaje

Número de prueba	Repuesta correcta	Ángulo de estereopsis a 3 *H* (")
1	Abajo	480
2	Izquierda	420
3	Abajo	360
4	Arriba	300
5	Arriba	240
6	Izquierda	180
7	Derecha	120
8	Izquierda	60
9	–	0

FIGURA 4

Diagrama de prueba para VT-04

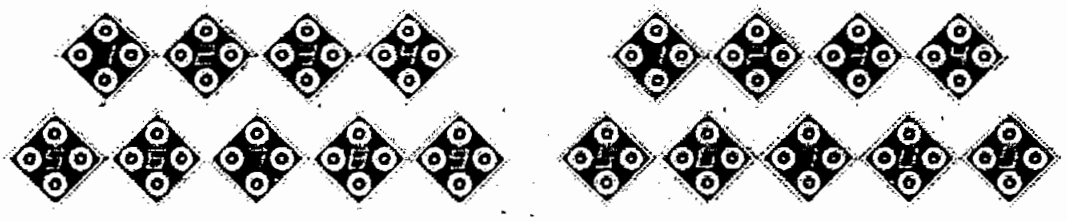


Imagen derecha Imagen izquierda

1438-04

e) *VT-05*: Límite de fusión cruzada (prueba de la barra)

Prueba la capacidad de percibir imágenes que se presentan de forma dicóptica con disparidades cruzadas como una sola imagen. Se presenta una estereopareja de barras cuyo paralelaje varía a razón de 10'/s. Pueden medirse los límites de fusión de las series ascendentes y descendentes. Se pide a los observadores que informen del momento en que detectan la ruptura de fusión, es decir, tan pronto como perciben imágenes dobles en las series ascendentes, así como de la recuperación de la fusión, es decir, tan pronto como perciben las imágenes dicópticas como una imagen única en las series descendentes.

FIGURA 5

Diagrama de prueba para VT-05

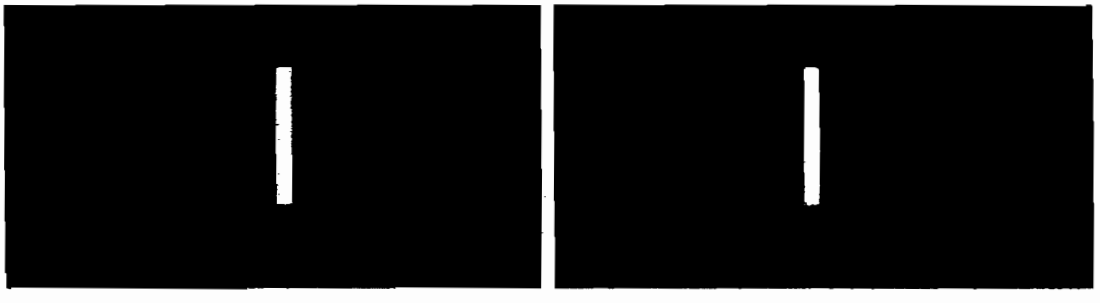


Imagen derecha Imagen izquierda

1438-05

f) *VT-06*: Límite de fusión sin cruce (prueba de la barra)

Prueba la capacidad de percibir imágenes presentadas de forma dicóptica con disparidades no cruzadas como una sola imagen. Las imágenes que se presentan son las mismas que en el caso cruzado anterior, pero se invierten las imágenes derecha e izquierda.

FIGURA 6

Diagrama de prueba para VT-06

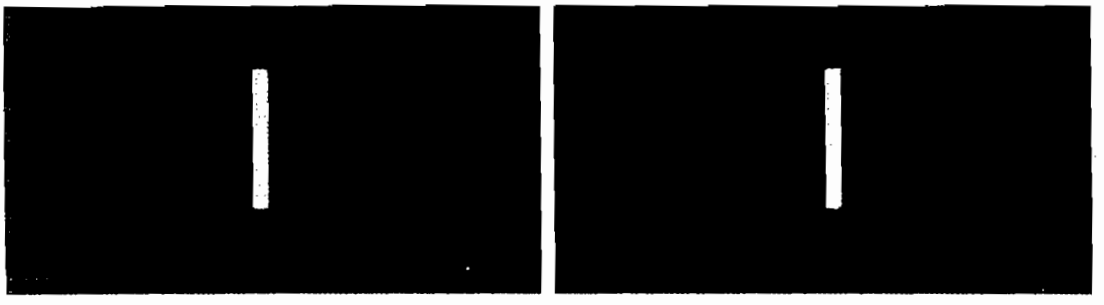


Imagen derecha Imagen izquierda

1438-06

g) *VT-07*: Estereopsis dinámica (prueba del estereograma de puntos aleatorios dinámicos)

Prueba la capacidad de percibir la profundidad en imágenes de un estereograma de puntos aleatorios en movimiento. Los observadores con visión normal pueden percibir una forma rectangular y un movimiento sinusoidal en profundidad en el estereograma de puntos aleatorios dinámicos.

FIGURA 7

Diagrama de prueba para VT-07

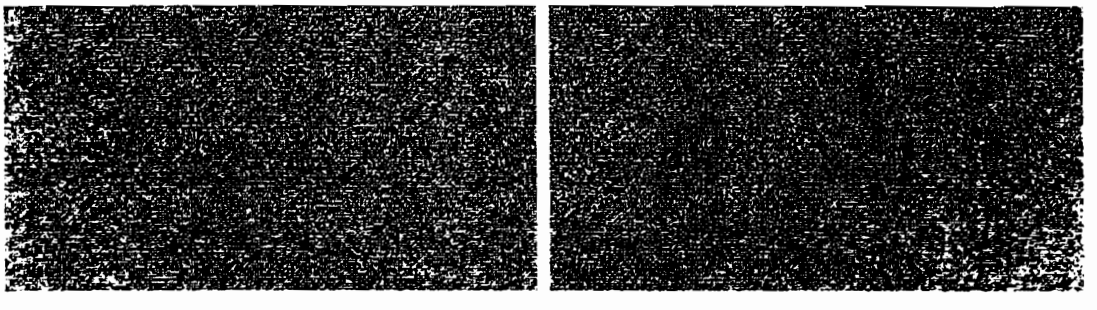


Imagen derecha Imagen izquierda

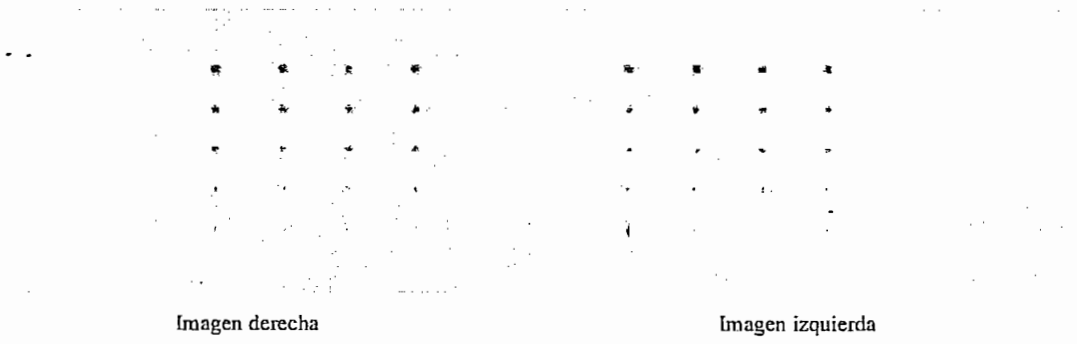
1438-07

h) *VT-08*: Agudeza binocular (prueba de agudeza)

Prueba la agudeza binocular con fusión binocular, incluyendo cualquier asimetría de la agudeza monocular que pueda impedir una estereopsis adecuada. Las imágenes tienen cuatro columnas y cinco líneas que consisten en caracteres E con diversas orientaciones y tamaños. Las dos columnas centrales pueden verse con ambos ojos; las dos columnas de la izquierda sólo pueden verse con el ojo izquierdo y las dos columnas de la derecha sólo pueden verse con el ojo derecho. Los observadores con una visión normal pueden decir cual es la orientación correcta de los caracteres E. Los tamaños de los caracteres se corresponden con agudezas de 1,0, 0,5, 0,33, 0,25 y 0,125 a 3 *H*.

FIGURA 8

Diagrama de prueba para VT-08



2 Imágenes naturales

Las imágenes naturales constan de 15 imágenes estáticas y 15 secuencias en movimiento, tal como se enumeran en los Cuadros 3 y 4. Algunas de ellas se ilustran en el Apéndice 1. Cada imagen se imprime de izquierda a derecha como imagen izquierda, imagen derecha, imagen izquierda: la imagen 3-D puede obtenerse fusionando la pareja de la izquierda (ojos no cruzados) o la pareja de la derecha (ojos cruzados).

3 Utilización del material de las pruebas estereoscópicas

La utilización del material de prueba debe limitarse a los propósitos siguientes:

- Evaluación técnica, incluyendo:
  - investigación y desarrollo de equipos y sistemas,
  - prueba de equipos en el proceso de desarrollo y producción,
  - prueba de las condiciones de transmisión para radiodifusión y telecomunicaciones,
  - mantenimiento del equipo.
- Demostración, incluyendo:
  - presentaciones en conferencias técnicas y talleres,
  - presentación de la calidad y funcionalidad de equipos, excluidas las promociones comerciales.

NOTA 1 – La presentación de la secuencia en movimiento N.° 10, Fútbol, SÓLO está permitida en recintos de investigación tales como universidades, institutos de investigación y laboratorios de fabricantes, pero no en lugares públicos.

CUADRO 2

Materiales para pruebas estereoscópicas – Pruebas de visión

N.°	Elemento	Prueba de	Contenido
1	Percepción simultánea	Capacidad de percibir imágenes presentadas simultáneamente de forma dicóptica y en su posición correcta	A un ojo se presenta una jaula y al otro un león
2	Fusión binocular	Capacidad de percibir dos imágenes dicópticas en los ojos izquierdo y derecho como una imagen	La imagen para un ojo tiene dos puntos y la del otro tres, con un punto en común
3	Estereopsis gruesa	Capacidad de percibir imágenes presentadas dicópticamente con cierto paralelaje como una sola imagen con una profundidad considerable	Las imágenes para los dos ojos son una estereopareja de imágenes de una libélula con sus alas extendidas
4	Estereopsis de precisión	Capacidad de percibir imágenes presentadas dicópticamente con cierto paralelaje como una sola imagen con una profundidad reducida	Nueve romboides cada uno con cuatro círculos, uno de los cuales tiene un pequeño paralelaje
5	Límite de fusión cruzada	Capacidad de percibir imágenes presentadas dicópticamente con disparidades cruzadas como una sola imagen	Una estereopareja de barras con paralelaje cruzado que varía a razón de 10'/s
6	Límite de fusión no cruzada	Capacidad de percibir imágenes presentadas dicópticamente con disparidades no cruzadas como una sola imagen	Una estereopareja de barras con paralelaje no cruzado que varía a razón de 11'/s
7	Estereopsis dinámica	Capacidad de percibir la profundidad en imágenes de un estereograma de puntos aleatorios en movimiento	Estereograma de puntos aleatorios dinámicos
8	Agudeza binocular	Agudeza binocular, incluyendo cualquier asimetría de la agudeza monocular que pueda impedir una buena estereopsis	Caracteres E con diversas de orientaciones y formas
9	Estrabismo horizontal	Desviación horizontal del ojo que el paciente no puede evitar	Líneas verticales y horizontales
10	Estrabismo vertical	Desviación vertical del ojo que el paciente no puede evitar	Líneas verticales y horizontales
11	Aniseikonia	Condición en la que la imagen ocular de un objeto visto por un ojo difiere en tamaño y forma respecto a como lo ve el otro ojo	La imagen izquierda consiste en caracteres [o y la derecha consiste en caracteres o], donde el carácter o tiene la misma posición en ambas
12	Cicloforia	Desviación de uno de los ojos alrededor del eje anteroposterior cuando se evita la fusión	La imagen izquierda consiste en la superficie de un reloj y la derecha en las manecillas del reloj marcando las seis en punto

NOTA 1 – Este material se ha grabado en formato VTR digital 1125/60/2:1 (véase la Recomendación UIT-R BT.709).

NOTA 2 – Este material puede obtenerse del Institute of Image Information and Television Engineers (ITE), 3-5-8 Shibakoen, Minato-ku, Tokio 105-0011, Japón. Tel.: +81-3-3432-4677, Fax: +81-3-3432-4675, e-mail: ite@ite.or.jp.

CUADRO 3

Material para pruebas estereoscópicas – Imágenes estáticas

N.º	Título	Contenido	Representativo de	Principales factores evaluados	Distorsión fundamental
1	Matices del otoño (Autumn tints)	Hojas otoñales rojas momiji a contra luz	Filmación en exteriores	Resolución estática y de profundidad	Ninguna
2	Matices del otoño y templo budista (Autumn tints and Buddhist temple)	Escena con hojas momiji rojas con luz directa y templo budista al fondo	Filmación en exteriores	Resolución estática y de profundidad	Ninguna
3	Atractivo kimono japonés en un templo budista (Attractive Japanese kimono in a Buddhist temple)	Mujer en kimono con un templo Daikakuji al fondo	Filmación en exteriores	Resolución estática y de profundidad	Ninguna
4	Hojas de otoño (Autumn leaves)	Mujer en kimono en un jardín japonés cubierto de hojas de otoño	Filmación en exteriores	Resolución estática y de profundidad	Ninguna
5	Cielo (Sky)	Escena de árboles con hojas con matices del otoño	Filmación en exteriores	Resolución estática y de profundidad	Ninguna
6	A la sombra de un árbol (Under the shade of a tree)	Mujer en una arboleda vestida con kimono	Filmación en exteriores	Resolución estática y de profundidad	Ninguna
7	Junto a los matices del otoño (By the side of the autumn tints)	Mujer en kimono y matices del otoño en un templo	Filmación en exteriores	Resolución estática y de profundidad	Ninguna
8	Jardín japonés (Japanese garden)	Jardín Eikando matizado por el otoño	Filmación en exteriores	Resolución estática y de profundidad	Ninguna
9	Belleza en kimono (Beauty in kimono)	Matices del otoño y mujer en kimono	Filmación en exteriores	Resolución estática y de profundidad	Ninguna
10	Escena 1 de ciudad (City scenery 1)	Edificio moderno y mujer	Filmación en exteriores	Resolución estática y de profundidad	Ninguna
11	Escena 2 de ciudad (City scenery 2)	Cascada artificial y mujer	Filmación en exteriores	Resolución estática y de profundidad	Ninguna
12	Escena 3 de ciudad (City scenery 3)	Paseo por un edificio y mujer	Filmación en exteriores	Resolución estática y de profundidad	Ninguna
13	En mi habitación 1 (In my room 1)	Mujer cómoda en su habitación	Producción de estudio	Resolución estática y de profundidad	Ninguna
14	En mi habitación 2 (In my room 2)	Mujer cómoda en su habitación	Producción de estudio	Resolución estática y de profundidad	Ninguna
15	Cenando (Dining)	Escena de una mujer cenando en una mesa	Producción de estudio	Resolución estática y de profundidad	Ninguna

NOTA 1 – Este material se ha grabado en formato VTR digital 1125/60/2:1 (véase la Recomendación UIT-R BT.709).

NOTA 2 – Todo el material está realizado en las condiciones de filmación siguientes: lente f = 75 mm, película EPR70 mm, separación de cámara 60 mm, con una disposición de cámara no cruzada.

NOTA 3 – Este material puede obtenerse del Institute of Image Information and Television Engineers (ITE), 3-5-8 Shibakoen, Minato-ku, Tokio 105-0011, Japón. Tel.: +81-3-3432-4677, Fax: +81-3-3432-4675, e-mail: ite@ite.or.jp.

CUADRO 4

Materiales para prueba estereoscópica – Secuencias en movimiento

N.º	Título	Contenido	Representativo de	Principales factores evaluados	Movimiento	Distancia focal (mm)	Distorsión fundamental
1	Jardín de tulipanes (Tulip garden)	Muchacha paseando por un jardín con tulipanes	Filmación en exteriores	Resolución estática y en profundidad	Lento	40	Ninguna*
2	Festival (Festival)	Templo portátil y tormenta de papel	Filmación en exteriores	Resolución estática y en profundidad	Lento	12	Ninguna*
3	Templos portátiles (Portable shrines)	Transporte de templos portátiles	Filmación en exteriores	Resolución estática y en profundidad	Panorámico lento	20	Ninguna*
4	Barcos cruzando (Crossing ships)	Barcos cruzando y público	Filmación en exteriores	Movimiento en profundidad	Panorámico lento	40	Ninguna*
5	Hojas rojas (Red leaves)	Mujer y hojas rojas	Filmación en exteriores	Movimiento en profundidad	Medio	12	Ninguna*
6	Jardines botánicos (Botanical gardens)	Cascada en jardín botánico	Filmación en exteriores	Ortoestereoscopia	Fijo	12	Ninguna*
7	Habitación de estar (Living room)	Mujer sentada en un sofá	Producción de estudio	Ortoestereoscopia	Fijo	12	Ninguna*
8	Almuerzo (A meal)	Gente almorzando en una mesa	Producción de estudio	Ortoestereoscopia	Fijo	12	Ninguna*
9	Parque de atracciones (Amusement park)	Muchachas jugando en un parque de atracciones	Filmación en exteriores	Movimiento en profundidad	Medio	12	Ninguna*
10	Fútbol (Football)	Partido de fútbol	Filmación en exteriores	Movimiento y resolución en profundidad	Medio	12	Ninguna*
11	Vocalista (A vocalist)	Vocalista en un auditorio	Producción de estudio	Ortoestereoscopia	Fijo	12	Ninguna*
12	Cromatismo (Chromakey)	Mujer y flores	Producción de estudio	Cromatismo	Fijo	12	Ninguna*
13	Maceta (Flower pot)	Muchacha y maceta	Filmación en exteriores	Movimiento en profundidad	Medio	12	Sí
14	Acuario (An aquarium)	Peces tropicales en un acuario	Filmación en exteriores	Movimiento en profundidad	Fijo	12	Sí
15	Jardín de flores (Flower garden)	Muchacha paseando en un jardín con flores	Filmación en exteriores	Movimiento y resolución en profundidad	Lento	12	Sí

NOTA 1 – Este material se ha grabado en formato VTR digital 1125/60/2:1 (véase la Recomendación UIT-R BT.709).

NOTA 2 – Todo el material señalado con * está producido con una disposición de cámara no cruzada.

NOTA 3 – Separación de cámara 65 mm en todo el material.

NOTA 4 – Este material puede obtenerse del Institute of Image Information and Television Engineers (ITE), 3-5-8 Shibakoen, Minato-ku, Tokio 105-0011, Japón. Tel.: +81-3-3432-4677, Fax: +81-3-3432-4675, e-mail: ite@ite.or.jp.

APÉNDICE 1
AL ANEXO 1

Ejemplos de secuencias de movimiento natural

FIGURA 9

N.º 1 - Jardín de tulipanes
(Tulip garden)

Imagen izquierda Imagen derecha Imagen izquierda

1438-09

FIGURA 10

N.º 2 - Festival
(Festival)

Imagen izquierda Imagen derechà Imagen izquierda

1438-10

FIGURA 11

N.º 5 - Hojas rojas
(Red leaves)

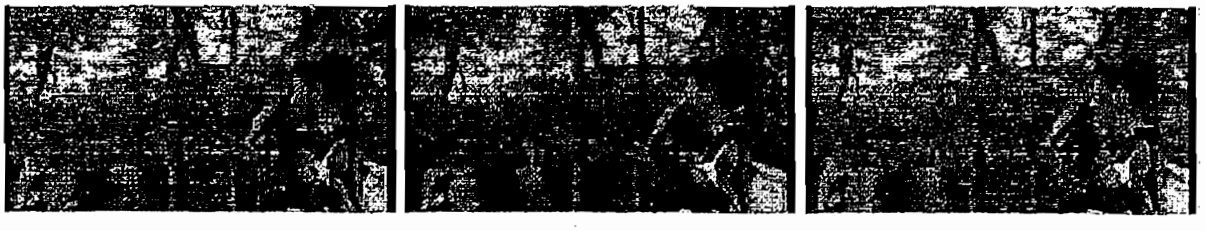


Imagen izquierda Imagen derecha Imagen izquierda

1438-11

FIGURA 12

N.° 7 - Habitación de estar
(Living room)

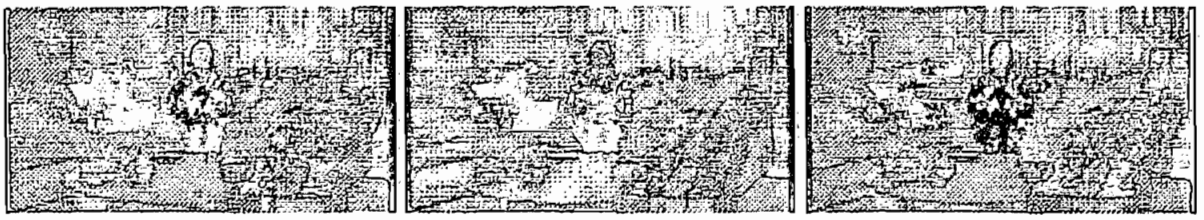


Imagen izquierda Imagen derecha Imagen izquierda

1438-12

ANEXO 2

Resultados experimentales utilizando el método de escala de calidad continua de doble estímulo (DSCQS, *double-stimulus continuous-quality scale*)

En este Anexo se documenta la aplicación del método para la evaluación subjetiva de imágenes. El método DSCQS se ha utilizado ampliamente y con éxito para evaluar la calidad de imagen subjetiva de imágenes monoscópicas. La experiencia ha demostrado que este método es igualmente adecuado para la evaluación de imágenes estereoscópicas y puede adaptarse fácilmente para medir otros atributos de la imagen además de la calidad, tales como la nitidez y profundidad percibida.

1 Medida mediante el método DSCQS de la nitidez y profundidad percibidas

El método DSCQS se ha adaptado fácilmente a la medida de otros atributos además de la calidad subjetiva de la imagen. Ello se ha conseguido realizando cambios específicos en las instrucciones dadas a los observadores. Por ejemplo, el método se ha adaptado a medir la nitidez percibida y la impresión general de la profundidad de las secuencias de imágenes estereoscópicas. En cada sesión sólo se midió un atributo (la calidad subjetiva de imagen percibida, la nitidez percibida o la profundidad percibida).

2 Estudio ilustrativo utilizando el método DSCQS con imágenes estereoscópicas

El método DSCQS se ha utilizado para medir la calidad subjetiva, la nitidez percibida y impresión general de profundidad de un conjunto de secuencias de imágenes estereoscópicas y monoscópicas. En el estudio ilustrativo, el objetivo era determinar si el procesamiento de un canal de una secuencia de imágenes estereoscópica afectaría a dichos atributos. A tal fin, la visión del ojo derecho de las secuencias de imágenes estereoscópicas se sometió a un filtrado paso bajo a tres niveles: no filtrado, resolución mitad y resolución de un cuarto. En las condiciones monoscópicas, ambos ojos observaban la imagen filtrada. Una revisión de la literatura científica [Julesz, 1971; Pastoor, 1991; Pastoor y otros, 1995; Perkins, 1992 y Berthold, 1997] inducía a esperar que el filtrado de un canal de una imagen estereoscópica tuviera un efecto mucho menor sobre los índices subjetivos que el filtrado de ambos canales, y que el índice subjetivo estuviera dominado por el canal no filtrado.

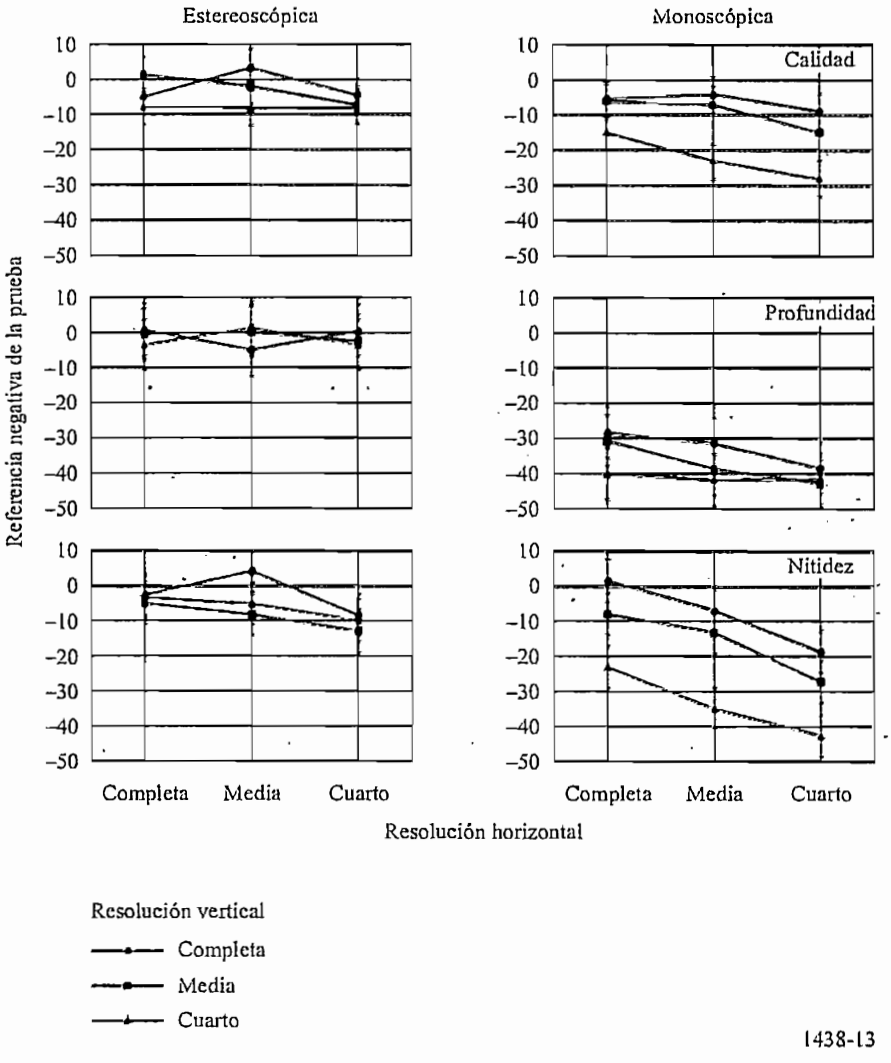
Las parejas de imágenes izquierda y derecha de una secuencia estereoscópica se visualizaron a 120 Hz utilizando un método secuencial en el tiempo, sobre una pantalla de visión directa de 29 pulgadas (74 cm) o sobre un retroproyector de 65 pulgadas (165 cm). Las imágenes de los ojos izquierdo y derecho fueron entrelazadas temporalmente y visualizadas en sincronía con la apertura y cierre de un par de cristales obturadores de cristal líquido Cristal Eyes fabricados por StereoGraphics. Los obturadores de cristal líquido tenían una transmitancia de aproximadamente el 30% y tiempos de

respuesta de 0,2 ms y 2,8 ms de cierre y apertura respectivamente. Ello significa que la cresta de luminancia hacia los ojos de los observadores era de 21 cd/m^2, más tenue de lo esperado, pero la mejor que puede conseguirse con la tecnología disponible de visualización secuencial en el tiempo. La distancia de observación era 4 *H*. Entre las secuencias A y B se presentaba un campo gris de 10 cd/m^2. Debe señalarse que cualquier método de visualización de imágenes estereoscópicas puede ser sustituido por el método secuencial en el tiempo sin que ello afecte al método DSCQS.

En la Fig. 13 se muestran los resultados de los experimentos. El eje Y indica la puntuación negativa de referencia de la prueba. Una puntuación cero indica que la secuencia de prueba fue puntuada igual que la secuencia de referencia estéreo no procesada. Una puntuación negativa significa que la secuencia de prueba fue puntuada más bajo que la secuencia de referencia.

FIGURA 13

Resultados del estudio ilustrativo utilizando el método DSCQS



Los efectos del filtrado paso bajo son evidentes en la pendiente y en el desplazamiento vertical de las líneas. Tal como se esperaba, en condiciones monoscópicas (véanse los diagramas de la derecha de la Fig. 13), el filtrado paso bajo tiene un gran efecto en la puntuación de la nitidez y la calidad de la imagen. Las bajas puntuaciones recibidas en lo que a profundidad se refiere, se debieron a que en las secuencias de prueba monoscópicas sólo existían indicaciones monoculares relativas a la profundidad. Asimismo, y tal como se esperaba, en las condiciones estéreo (véanse los

diagramas de la izquierda de la Fig. 13), las tres dimensiones (calidad, profundidad y nitidez) recibieron una puntuación mayor que en las condiciones monoscópicas. El filtrado paso bajo de un canal de una pareja estéreo tiene un efecto despreciable sobre la profundidad percibida y efectos menores sobre la nitidez percibida y la calidad en general. Evidentemente, la gran cantidad de información de frecuencia espacial de la imagen no filtrada del ojo izquierdo compensaba la falta de dicha información en la imagen del ojo derecho.

El estudio ilustrativo y otros trabajos sobre secuencias de imágenes estereoscópicas utilizando el método DSCQS [Stelmach y Tam, 1998] permiten concluir que este método es una herramienta valiosa y útil para el estudio de imágenes estereoscópicas. El método puede adaptarse a la medición de otros aspectos de las secuencias de imágenes estereoscópicas tales como presencia, potencia y naturalidad.

REFERENCIAS BIBLIOGRÁFICAS

BERTHOLD, A. [1997] The influence of blur on the perceived quality and sensation of depth of 2D and stereo images. *ATR Human Information Processing Research Laboratories Technical Report*, TR-H-232, Kyoto, Japón.

JULESZ, B. [1971] Foundations of Cyclopean Perception. *The University of Chicago Press*. Chicago, IL, Estados Unidos de América.

PASTOOR, S. [1991] 3-D- television: A survey of recent research results on subjective requirements. *Signal Processing: Image Communication*, 4(1), p. 21-32.

PASTOOR, S., WÖPKING, M., FOURNIER, J. Y ALPERT, T. [1995] Digital stereoscopic imaging & applications (DISTIMA): Human Factors Data. Deliverable ID: R2045/HHI/AT/DS/C/026/b1.

PERKINS, M. G. [1992] Data compression of stereopairs. *IEEE Trans. on Comm.*, 40(4), p. 684-696.

STELMACH, L. y TAM, W. J. [1998] Stereoscopic image coding: Effect of disparate image-quality in left- and right-eye views. *Signal Processing: Image Communication*, 14, p. 111-117.

INFORME UIT-R BT.2017

PERFIL MULTIVISIÓN MPEG-2 PARA TELEVISIÓN ESTEREOSCÓPICA

(1998)

1 Introducción al perfil multivisión (MVP, *multi-view profile*) MPEG-2

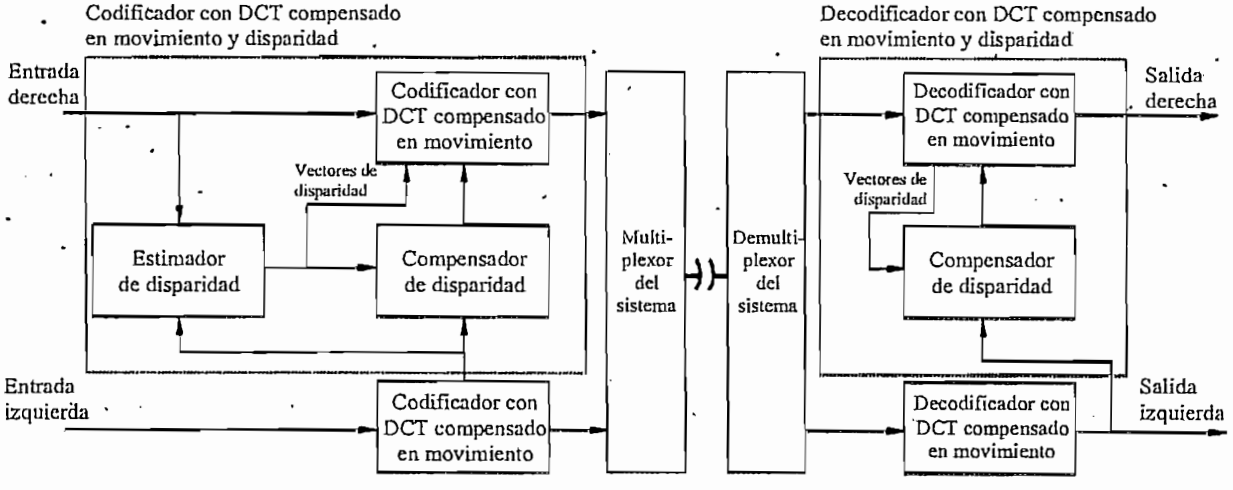
La ampliación de la norma de vídeo MPEG-2 (Recomendación UIT-T H.262 | ISO/CEI 13818-2: Tecnología de la información – Codificación genérica de imágenes en movimiento e información de audio asociada: Vídeo) en previsión de las aplicaciones multivisión (por ejemplo las utilizadas en el vídeo estereoscópico) ha sido elevada al rango de Norma Internacional final en la reunión ISO/CEI JTC 1/SC 29/GT 11 de septiembre de 1996 (Enmienda 3, GT 11 N1366). Se considera que el perfil multivisión (MVP) es idóneo para las aplicaciones que necesitan muchos puntos de visión en el contexto de la norma de vídeo MPEG-2. El MVP admite imágenes estereoscópicas como imágenes fuente para una amplia gama de resoluciones y calidades de imagen, que dependen de las necesidades de las aplicaciones de que se trate.

1.1 Esquema de codificación para el MVP

La Fig. 1 muestra un diagrama de bloques del modelo de códec de referencia para el MVP. Sus principales características son codificación monoscópica en su capa base a efectos de compatibilidad y predicción híbrida de movimientos y disparidad a fin de aumentar la eficacia de la compresión. Para codificar una capa de mejora se utilizan herramientas de escalonabilidad temporal.

FIGURA 1

Modelo de códec de referencia para el MVP



DCT: transformación discreta en coseno *(discrete cosine transform)*

Rap 2017-01

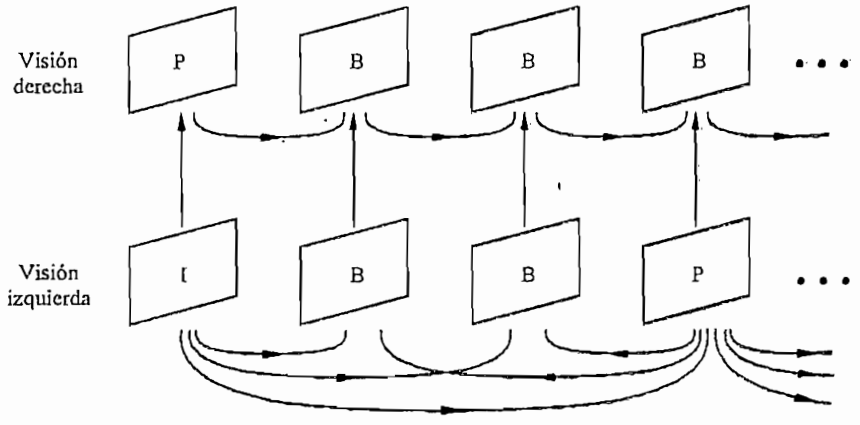
La Fig. 2 muestra una configuración de modos de predicción. A la capa base se le aplica una codificación monoscópica con las mismas herramientas que al perfil principal (MP, *main profile*), incluida la Norma ISO/CEI 11172-2. Se asigna una capa base de MVP a la visión izquierda y una capa de mejora a la visión derecha. La capa de mejora se codifica utilizando herramientas de escalonabilidad temporal y en la capa mejorada puede aplicarse la predicción híbrida de movimientos y disparidad. Se prevé una mayor compresión de la visión derecha del vídeo estereoscópico a causa del parecido entre la visión izquierda y la visión derecha.

El MVP, uno de los perfiles escalonables en términos de capas de múltiples puntos de visión, tiene las mismas características de compatibilidad que otros perfiles escalonables, entre ellas, la compatibilidad con el MP. Por ejemplo:

- decodificadores que se ajustan al MVP a un cierto nivel pueden decodificar trenes de bits que se ajustan al MP al nivel correspondiente (es decir, compatibilidad hacia adelante),
- decodificadores que se ajustan al MP a un cierto nivel pueden decodificar los trenes de bits de la capa base del MVP (es decir, compatibilidad hacia atrás).

FIGURA 2

Ejemplo de configuración de predicción con codificación $M = 3$ de la visión izquierda, imagen de trama de la visión derecha codificada mediante predicción de disparidad respecto a la visión izquierda y predicción de movimientos con respecto a sí misma



1.2 Valores de los parámetros del MVP

Los niveles del MVP son alto, alto-1440, principal y bajo. La escalonabilidad temporal comporta dos capas, una capa base y una capa de mejora. Ambas capas tienen la misma resolución espacial a la misma frecuencia de trama. Los Cuadros 1 a 4 indican los límites a los que se han de atener las velocidades de muestreo, las velocidades de los pels de luminancia, las velocidades binarias y los tamaños de memoria tampón del MVP.

CUADRO 1

Límites superiores de la densidad de muestreo

Nivel	Capa de resolución espacial		Perfil Multivisión
Alto	Mejorada (visión derecha)	Muestras/línea Líneas/trama Tramas/s	1 920 1152 60
	Inferior (visión izquierda)	Muestras/línea Líneas/trama Tramas/s	1 920 1152 60
Alto-1440	Mejorada (visión derecha)	Muestras/línea Líneas/trama Tramas/s	1 440 1152 60
	Inferior (visión izquierda)	Muestras/línea Líneas/trama Tramas/s	1 440 1152 60
Principal	Mejorada (visión derecha)	Muestras/línea Líneas/trama Tramas/s	720 576 30
	Inferior (visión izquierda)	Muestras/línea Líneas/trama Tramas/s	720 576 30
Bajo	Mejorada (visión derecha)	Muestras/línea Líneas/trama Tramas/s	352 288 30
	Inferior (visión izquierda)	Muestras/línea Líneas/trama Tramas/s	352 288 30

CUADRO 2

Límites superiores de la velocidad de muestreo de luminancia (muestras/s)

Nivel	Capa de resolución espacial	Perfil
		Multivisión
Alto	Mejorada (visión derecha)	62 668 800
	Inferior (visión izquierda)	62 668 800
Alto-1440	Mejorada (visión derecha)	47 001 600
	Inferior (visión izquierda)	47 001 600
Principal	Mejorada (visión derecha)	10 368 000
	Inferior (visión izquierda)	10 368 000
Bajo	Mejorada (visión derecha)	3 041 280
	Inferior (visión izquierda)	3 041 280

CUADRO 3

Límites superiores de las velocidades binarias (Mbit/s)

Nivel	Perfil
	Multivisión
Alto	130 ambas capas
	80 capa base
Alto-1440	100 ambas capas
	60 capa base
Principal	25 ambas capas
	15 capa base
Bajo	8 ambas capas
	4 capa base

CUADRO 4

Requisitos en cuanto a tamaño de memoria (bits)

Nivel	Capa	Perfil
		Multivisión
Alto	Mejorada	15 898 480
	Base	9 787 248
Alto-1440	Mejorada	12 222 464
	Base	7 340 032
Principal	Mejorada	3 047 424
	Base	1 835 008
Bajo	Mejorada	950 272
	Base	475 136

1.3 Ampliación de los parámetros de cámara

Se ha introducido en el MVP una ampliación para dar cabida a la información relativa a la cámara. La ampliación permite especificar la altura del dispositivo de imagen, la longitud focal, el número F, el ángulo vertical del campo de visión, la posición y dirección de la cámara y la dirección superior de la misma.

2 Pruebas de evaluación del MVP

Las pruebas de verificación del MVP se llevaron a cabo en tres emplazamientos de prueba diferentes situados en Japón, Alemania y Canadá. Los resultados de dichas pruebas se presentaron en la reunión del Grupo de Trabajo 11 celebrada en Chicago (GT 11 N1373) septiembre de 1996. Test and video subgroup. Results of MPEG-2 multi-view profile verification test. Los resultados de los distintos emplazamientos de pruebas son coherentes entre sí y ponen de manifiesto que en líneas generales, a las velocidades binarias utilizadas, los observadores opinaron que el esquema de codificación del perfil multivisión MPEG-2 no introducía ninguna perturbación.

2.1 Método de prueba

Se utilizó el método de escala de degradación con doble estímulo (variante II) de la Recomendación UIT-R BT.500. Para obtener evaluaciones más precisas se utilizó una escala continua en vez de la escala discreta recomendada por el UIT-R.

2.2 Condiciones de las pruebas

Se utilizaron las secuencias de prueba generadas durante el intercambio de trenes binarios. El Cuadro 5 resume las condiciones de las pruebas. En cada emplazamiento de prueba se utilizó un sistema de visualización diferente.

CUADRO 5

Resumen de las condiciones de las pruebas subjetivas

Secuencias	«Street organ (Organillo)», «Flower pot (Maceta)», «Trapeze (Trapecio)» (525/60) «Fun fair (Feria)» (625/50)
Algoritmos y velocidades binarias (visión izquierda/derecha)	MVP@ML: 6/3 Mbit/s, 9/4 Mbit/s Simulcast de perfil principal en el nivel principal (MP@ML): 4,5/4,5 Mbit/s, 6,5/6,5 Mbit/s Simulcast de MP@ML como anclaje inferior: 2,5/2,5 Mbit/s (para «Street organ», «Fun fair»), 1,5/1,5 Mbit/s (para «Flower pob», «Trapeze») Original/original como anclaje superior
Método de prueba	Método de escala de degradación con doble estímulo (variante II) descrito en la Recomendación UIT-R BT.500, con escala continua
Sistema de visualización estereoscópica (tamaño de la imagen, distancia de observación)	HHI: Sistema de visualización de doble espejo (19 cm × 14 cm, 5 *H*) CRC: Visualización secuencial en el tiempo y gafas con obturador LCD (40,6 cm × 30,5 cm, 4 *H*) NHK: Proyectores de televisión de alta definición (TVAD) con LCD y gafas polarizantes (82 cm × 57 cm, 5 *H*)
Observadores	HHI: 24 espectadores sin experiencia CRC: 18 espectadores sin experiencia NHK: 19 espectadores sin experiencia (se rechazó un observador en la selección realizada en base a la Recomendación UIT-R BT.500)

HHI: Heinrich-Hertz-Institut für Nachrichtentechnik (Alemania)

CRC: Communications Research Center (Canadá)

NHK: Nippon Hoso (Kyokai (Japón))

2.3 Resultados de las pruebas de evaluación subjetivas

Para cada condición de prueba se calcularon las notas medias y los intervalos de confianza del 95%. Los resultados de las pruebas de HHI, CRC y NHK se presentan en el Cuadro 6 y en la Fig. 3. HHI1 y HHI2 son los resultados obtenidos en HHI en dos partes distintas de la misma secuencia. HHI no pudo probar las secuencias en su totalidad porque no tenía suficiente memoria de pantalla.

CUADRO 6

Notas medias e intervalos de confianza del 95%

a) Secuencia: Street organ (Organillo)

	Fuente	MVP (9/4 Mbit/s)	MP × 2 (6,5/6,5 Mbit/s)	MVP (6/3 Mbit/s)	MP × 2 (4,5/4,5 Mbit/s)	Anclaje inferior
NHK	4,71 ±0,17	4,18 ±0,27	4,40 ±0,26	4,06 ±0,39	3,51 ±0,32	1,74 ±0,33
CRC	4,24 ±0,37	4,19 ±0,33	4,33 ±0,29	4,27 ±0,34	4,07 ±0,35	2,19 ±0,35
HHI1	4,89 ±0,12	4,55 ±0,21	4,58 ±0,22	4,23 ±0,26	3,63 ±0,35	1,30 ±0,19
HHI2	4,86 ±0,13	4,68 ±0,19	4,85 ±0,13	4,44 ±0,24	4,24 ±0,32	1,80 ±0,23

b) Secuencia: Flower pot (Maceta)

	Fuente	MVP (9/4 Mbit/s)	MP × 2 (6,5/6,5 Mbit/s)	MVP (6/3 Mbit/s)	MP × 2 (4,5/4,5 Mbit/s)	Anclaje inferior
NHK	4,79 ±0,16	4,03 ±0,44	4,28 ±0,25	4,07 ±0,33	4,13 ±0,37	2,28 ±0,32
CRC	4,53 ±0,14	4,57 ±0,20	4,45 ±0,22	4,40 ±0,20	4,40 ±0,21	2,70 ±0,34
HHI1	4,81 ±0,19	4,49 ±0,25	4,52 ±0,26	4,33 ±0,24	4,46 ±0,23	1,96 ±0,25
HHI2	4,83 ±0,14	4,48 ±0,21	4,33 ±0,22	4,08 ±0,26	4,16 ±0,25	1,69 ±0,24

c) Secuencia: Trapeze (Trapecio)

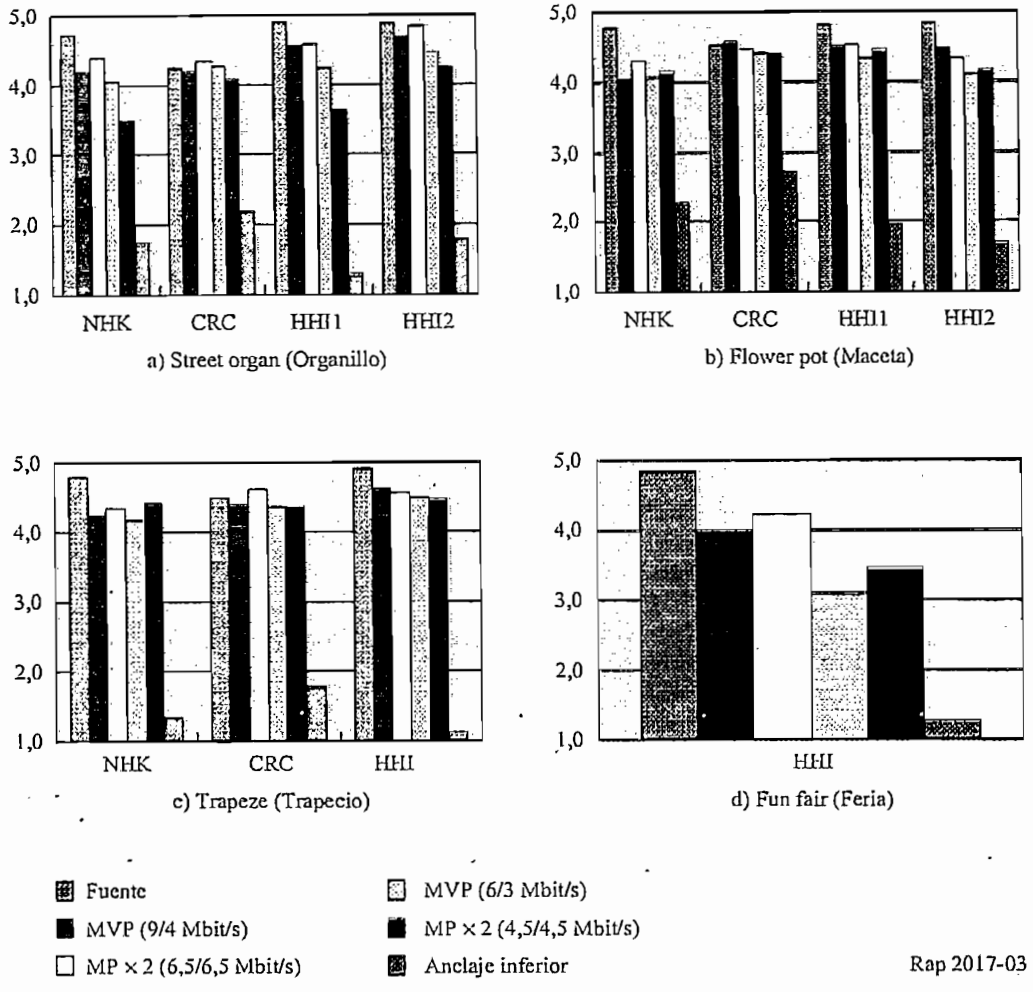
	Fuente	MVP (9/4 Mbit/s)	MP × 2 (6,5/6,5 Mbit/s)	MVP (6/3 Mbit/s)	MP × 2 (4,5/4,5 Mbit/s)	Anclaje inferior
NHK	4,77 ±0,13	4,24 ±0,25	4,34 ±0,38	4,16 ±0,24	4,41 ±0,23	1,33 ±0,18
CRC	4,48 ±0,22	4,38 ±0,24	4,62 ±0,14	4,37 ±0,23	4,36 ±0,24	1,78 ±0,31
HHI1	4,90 ±0,11	4,60 ±0,19	4,55 ±0,25	4,48 ±0,27	4,46 ±0,28	1,13 ±0,14

d) Secuencia: Fun fair (Feria)

	Fuente	MVP (9/4 Mbit/s)	MP × 2 (6,5/6,5 Mbit/s)	MVP (6/3 Mbit/s)	MP × 2 (4,5/4,5 Mbit/s)	Anclaje inferior
HHI1	4,83 ±0,14	3,96 ±0,32	4,23 ±0,29	3,10 ±0,35	3,46 ±0,27	1,27 ±0,20

FIGURA 3

Notas medias de la evaluación subjectiva



Vale la pena mencionar algunos aspectos de los resultados de estas pruebas:

- En cada una de las cuatro secuencias, la nota media de la secuencia MVP a la velocidad binaria de 9/4 Mbit/s no presenta una diferencia importante con respecto a la nota media del simulcast de los MP a la velocidad binaria de 6,5/6,5 Mbit/s. Asimismo, la nota media de la secuencia MVP a la velocidad binaria de 6/3 Mbit/s no se diferencia sensiblemente de la nota media del simulcast de los MP a la velocidad binaria de 4,5/4,5 Mbit/s, excepto el par de la secuencia «Street organ». Para «Street organ», la calidad del MVP es superior a la del simulcast de los MP. Estos resultados ponen de manifiesto que las diferencias en la evaluación subjetiva entre el MVP y simulcast de los MP son muy pequeñas a velocidades binarias superiores para imágenes de poco movimiento («Flower pot» y «Trapeze») y/o con diferencia de luminancia importante entre las visiones izquierda y derecha («Fun fair»).

- «Fun fair» es la escena con notas medias más dispares. En esta escena puede observarse un número mayor de movimientos (cambios en el contenido de imagen con respecto a la trama siguiente) que en las demás escenas. En «Fun fair» especialmente, en los objetos en movimiento cubren la mayor parte de la imagen.

3 Trabajos futuros sobre televisión estereoscópica

Los progresos realizados hasta la fecha han puesto de manifiesto que la televisión estereoscópica es técnicamente viable. El perfil multivisión MPEG recientemente aprobado ofrece una base para la codificación y compresión de las secuencias de vídeo estereoscópico. Las pruebas de evaluación de calidad llevadas a cabo también han evidenciado que, dentro de

los límites de los parámetros de prueba seleccionados, puede lograrse una calidad de imagen cuya percepción subjetiva sea satisfactoria. No obstante, quedan muchas cuestiones por resolver. Algunos de los temas en los que hay que profundizar son los siguientes:

3.1 Requisitos

- Sería conveniente que los futuros sistemas de televisión estereoscópica fuesen compatibles con los sistemas de televisión digital monoscópica que están apareciendo en la actualidad, y que la velocidad binaria adicional fuera lo más reducida posible.
- La calidad de la imagen principal monoscópica de una pantalla de televisión monoscópica debe ser lo más próxima posible a la de la imagen monoscópica que utilice toda la capacidad del canal.

3.2 Información necesaria tanto para la televisión digital con definición convencional (TVDC) como para la TVAD

- El grado posible de asimetría de la velocidad binaria asignada a las imágenes izquierda y derecha correspondientes a una secuencia de vídeo estereoscópico para reducir al mínimo la degradación de la calidad de imagen del nivel de base.
- La repercusión de la asimetría de la velocidad binaria asignada a las imágenes de visión izquierda y derecha sobre las perturbaciones debidas a la codificación y la compresión percibidas subjetivamente y la calidad global de la secuencia de vídeo estereoscópico.
- Los factores que pueden provocar fatiga en el espectador; y las medidas que pudieran reducir o suprimir dicha fatiga.
- La gama de velocidades binarias que se precisa para conseguir una calidad satisfactoria percibida subjetivamente tanto de la imagen estereoscópica como de la imagen monoscópica proporcionada por la imagen del nivel de base. Obtenida mediante pruebas de evaluación adicionales con numerosas secuencias de vídeo que representen una gran variedad de material de programación y una amplia gama de velocidades binarias.
- Métodos de pruebas adecuados para evaluar las imágenes estereoscópicas.
- Algoritmos de codificación con los que se consiga una compresión más eficaz de las señales de televisión estereoscópica.

Los estudios al respecto deben realizarse en coordinación con el GT 11B, el GMT 10-11Q y demás Grupos de Trabajo y organismos pertinentes.

ANEXO 6
PAPERS

SPIE Vol. 1457, Stereoscopic Displays and Applications II, ed. J O Merritt, S S Fisher (Aug 1991) Copyright SPIE

A COMPACT ZOOM LENS FOR STEREOSCOPIC TELEVISION.

P.M. Scheiwiller,
S.P. Murphy, A.A. Dumbreck.
AEA Technology, Decommissioning and Radwaste, Harwell Laboratory, DIDCOT, Oxon OX11 0RA, UK.

ABSTRACT

Previously we have emphasised the need for accurate picture matching and the proper convergence of left and right channels of a stereoscopic camera to ensure that the image is comfortable to view and does not cause fatigue. This usually procludes the use of standard zoom lenses in high quality 3D television cameras as the optical alignment frequently changes with focal length and two such lenses, when motorised, would have to be controlled with great accuracy to avoid viewer discomfort.

This paper describes the on-going development of a compact zoom lens specifically for stereoscopic television in nuclear environments. Custom designed optics in radiation tolerant glass provide a focal length of 12.5mm to 36mm with a package length of less than 110mm. A novel method of encoding the position of the lens elements allows for very compact motorisation and a precision mechanism designed to overcome backlash ensures the stability of picture matching.

The position of the lens elements is controlled by a 16 bit microcontroller and the control strategy allows focus and convergence to be maintained to a high degree of accuracy during zooming.

1.0 INTRODUCTION

Harwell stereoscopic (or 3D) television systems are increasingly being used in the nuclear industry and other hazardous environments for inspection, and in conjunction with dextrous manipulators and robots, to allow complex remote operations to be performed. Dramatic improvements in operator performance are often evident, sometimes to the extent of enabling otherwise impossible tasks to be carried out. However 3D television is not new and successful application and operator acceptance in such an exacting field is still relatively rare. Systems must not only be easy to use but comfortable to watch for protracted periods of time. They must also provide useful depth information rather than just give a 3D impression or a feeling of space and present a view which is realistic. The development, based on human factors considerations, and evaluation of our 3D TV systems to meet these requirements is described elsewhere[1]. Briefly, two TV cameras laterally displaced provide left and right eye views. Camera convergence is provided by axial offset of the lenses rather than the more usual method of toeing-in the cameras. Our stereo displays use two high quality TV monitors mounted at 90 degrees to each other. The pictures from these are combined by a semi-silvered mirror. Polarising filters in front of each monitor and the polarising glasses worn by the viewer ensure that each eye sees only the picture from the appropriate camera channel.

However of the cameras we have built so far none has employed a zoom lens. This paper describes the technical difficulties involved in making such a camera, and how by developing our own zoom lens these can be overcome.

2.0 REQUIREMENTS OF A ZOOM LENS FOR 3D TV.

The requirements of a zoom lens for 3D TV in nuclear environments will be considered, particularly for use with our camera designs, using axial offset for camera convergence. Convergence is thus assumed to be a function of the camera, whilst focus, zoom and iris are taken to be functions built into the lens. In general the requirement is for a compact unit with accurate picture matching between left and right channels and good optical properties, rather than a large zoom range at the expense of any of these.

 0-8194-0556-6/91/$4.00

2.1 Picture matching.

Whilst it is easy to generate pictures in which some stereoscopic effect can be seen, it is rather more difficult to build a 3D TV system which does not cause viewers discomfort and headaches when used for any length of time, and can be used as a practical tool. One of the main requisites is that the pictures seen by the left and right eyes should be the same except for the small horizontal disparities which give rise to the perception of depth. Any other visible differences present the viewers brain with conflicting information which cannot arise during normal vision. We believe that the following criteria[2] for picture matching should be met any zoom lens 3D camera throughout its zoom range in order to achieve acceptable results:-

Image Size:- The two images should correspond to within half of one percent in both vertical and horizontal directions.

Vertical Alignment:- An error of no more than two scanning lines between any two corresponding image points.

Horizontal Alignment:- An error of no more than $1/200^{th}$ of the screen width.

Rotational Alignment:- An error of no more than 0.25 degrees.

Left and right channels of the camera should be focused at the same distance to a high degree of accuracy; both pictures should appear equally sharp. Although this becomes a less stringent requirement where small apertures can be used and the depth of focus is large (ie bright lighting, short focal length) accurate positional control of lens elements is still required.

The brightness, contrast and colour of the two pictures should also be accurately matched, ideally so any differences are imperceptible.

2.2 Geometry of the stereoscopic image.

One of the lessons to be learned from many bad 3D movies is that attempts to reproduce excessive depth cause eyestrain and should be avoided. From practical experience we suggest a parallax limit (ie the maximum horizontal displacement between corresponding points in the left and right images) of $1/30^{th}$ of the screen width. This can be achieved by selecting a suitable inter-camera separation commensurate with the lens focal length and the distance of objects from the camera[3]. Although in practice this may be a compromise, the image should also be realistic; neither excessively compressed in depth, making objects look like cardboard cut-outs, nor exaggerated. In many applications the physical size of the lenses will set a lower limit on the interaxial separation.

For comfortable pictures it is usually desirable for the camera to be focused and converged at the same distance. This is for two reasons. Firstly, objects in the scene will have a parallax which increases with their distance from the plane where the camera is converged. If this is also where the camera is focused then the distracting effect of objects a long way behind or infront of the screen (ie with large parallaxes) will be minimised by putting them out of focus.

Secondly, in normal human vision there is a relationship between where the eyes are focused (accommodation) and where their visual axes converge (vergence). Cues for the eyes to converge also bring about a corresponding change in their focus. The relationship is reciprocal[4], so similarly, cues for accommodation bring about vergence of the eyes. The parallax between corresponding areas of left and right views in a stereoscopic display is a strong cue for vergence of the eyes, and a viewer will tend to converge at the point in space where the image appears to be situated. However the plane of the display screen is the optimum place for the viewers eyes to be focused and appropriate accommodation would normally bring about a corresponding vergence. In the extreme, such a conflict of cues may lead to problems, but the situation is largely avoided if the camera is focused and converged in the same

place (objects in sharpest focus appear at the plane of the screen) and maximum screen parallax is limited as above. This also simplifies operation of the camera, giving the operator a single control.

A practical zoom lens for 3D TV should therefore be compact enough to allow an appropriate camera separation, given the focal length, and interact with the camera so that the correct relationship between focus and convergence is maintained. Whereas for a fixed focal length lens the relationship is fairly simple and focusing can be achieved by moving the whole lens backwards and forwards, with a zoom lens focusing cannot readily be achieved in this way. The lens should also stay focused with a high degree of accuracy when the lens is zoomed, and similarly the camera should remain converged on the same plane. Where convergence is achieved by axial offset of the camera sensors (as in our cameras) the offset must change when the lens is zoomed as focal length is one of the factors determining the axial offset required to converge at a particular distance. A relatively sophisticated control system is therefore required to coordinate these variables and maintain the correct geometric properties of the stereoscopic image.

2.3 Other requirements.

As our 3D TV systems are designed primarily for use in hazardous nuclear environments the optics of a zoom lens should be realisable in radiation stable (ie non-browning) glass, and its construction make minimal use of materials susceptible to radiation damage (eg PVC). In addition the size of the final image formed should be large enough to allow convergence by our method of axial offset of either the lens or camera-sensors.

With fewer restrictions on signal bandwidth than in other applications, (for example, broadcast) we plan to uses zoom lens 3D with high resolution sensors (1000TV lines per picture height) and therefore, ideally, the lens optical performance should match this figure. Finally, the lens should ideally have a minimum focusing distance less then 1m, where 3D viewing is used in conjunction with dextrous manipulators.

3.0 DIFFICULTIES WITH CONVENTIONAL ZOOM LENSES

Although a camera could be built using two standard zoom lenses side-by side to provide left and right stereo views, there would be considerable difficulties with many commercially available lenses.

Firstly , the choice of lenses in radiation tolerant glass with an adequate optical performance is somewhat limited and many of those available are physically large. This would limit the minimum interaxial separation and may adversely affect the geometry of the stereoscopic image. Alternatively the lenses could be mounted at right angles looking into a semi-silvered mirror. Although this would allow any interaxial separations down to zero, the mirror would have to be large to accommodate a wide angle of view, and this configuration is intrinsically more bulky and less robust.

Secondly, problems would be encountered in achieving the accuracy in picture matching described above. To match the picture sizes the zoom controls of the two lenses would have to be coupled together either with a mechanical linkage, or for motorised lenses, by accurate servo control. The accuracy of the servo control may be achieved relatively easily, but a mechanical linkage is likely to be bulky. There is a further problem, that for many lenses the position of the lens elements is not a monotonic function of focal length. The movement is achieved by a peg-in-slot arrangement on a rotating barrel, but where the direction of movement reverses there may be backlash of the peg in the slot. At this point, no matter how accurate the servo control or how good the mechanical linkage, it may not be possible to achieve reliably the accuracy required in picture matching.

Problems with picture matching may also arise if the optical centre of the image moves, as is often found to be the case, as the lens is zoomed. Sometimes it will be possible to match a pair of lenses so that this movement is the same for both left and right views and the images remain in register as the focal length is changed. However if registration errors become larger than the limits set out above, the stereoscopic pictures will be uncomfortable to watch over at least part of the zoom and/or focussing range of the lens.

Finally, the additional requirement for a short minimum focus distance (less than 1m, ideally around 200mm) virtually rules out any commercially available lens which might otherwise be suitable.

4.0 TWO ANGLE OF VIEW STEREO CAMERAS

Previously we have built 3D cameras with more than one angle of view by switching between pairs of fixed focus lenses, circumventing the problems highlighted above.

The most sophisticated incarnation of this idea has been a radiation tolerant, high resolution black and white camera, two versions of which have been built. The first was designed for in-reactor inspection and repair, the second is to be used in decommissioning the prototype of a major type of nuclear facility in the UK (WAGR). Figure 1 shows a photograph of the WAGR camera. Two pairs of fixed focal length lens are mounted on a rotating turret (in principle not unlike early broadcast TV cameras) to provide wide and narrow angles of view. Camera tubes for left and right channels are mounted on cross-slides with position feedback to allow the camera to be converged by offsetting the sensors with respect to the axis of the lenses, and so that the two pairs of lenses can have different separations. To focus the camera the whole assembly of camera tubes, head amplifiers and cross slides is moved backwards and forwards on precision slides. An 8088 based microcomputer controls camera focus, convergence and the sequence of operations to change lenses, ensuring that the camera remains focused and converged in the same place after a lens change. The two pairs of lenses not only have different separations but the movement law for focus and convergence, depending on focal length, is very different between the two lens sets.

The computer also provides a readout on an LCD display of the camera status and the distance at which it is focused and converged. Thus by focusing the camera to make the left and right views of an object appear overlaid on the display (ie at the plane of convergence) the camera can also be used for simple range finding and measurement.

5.0 DEVELOPMENT OF A COMPACT ZOOM LENS FOR 3D TV.

The remainder of this paper describes the design and prototype development of a radiation tolerant 12.5mm to 36mm focal length zoom for 3D TV use, with a package length of less than 110 mm. The lens is intended to overcome the problems highlighted above.

5.1 Optical and mechanical design

Each channel of the lens comprises of three groups of elements, a front group which moves during both focusing and zooming, a middle group which moves only during zooming, and a fixed rear group. The lens is designed to be built out of Schott radiation stable (Cerium doped) glass and have a minimum focusing distance of 200mm. The diameter of the largest lens elements allows a minimum separation between the two channels of about 45mm and a maximum numerical aperture of f2.8. Calculations of the image aberrations indicate that the lens will be consistent with high resolution sensors (1000 TV lines per picture height).

The moveable groups of lens elements for left and right channels are mounted on a common mounting plate which slides backwards and forwards on precision guides. Left and right channel picture size, focus and alignment are therefore intrinsically matched. The groups of elements are moved by DC motors driving low backlash leadscrews via reduction gearing with position feedback. The iris for each channel is in the fixed rear group, simplifying the linkage for accurate tracking. This scheme is shown in figure 2.

Figure 3 shows the paths followed by the elements during zooming. Focusing is achieved by offsetting the front group by an amount which does not depend on zoom position. For the lens to remain in focus during zooming and the focal length set accurately enough to allow the camera to converge at the same place as the lens is focused, the lens elements must be controlled to within a positional tolerance of 0.05mm. Although the paths are monotonic the equation describing the curves is relatively complicated. In conjunction with the positional accuracy required this necessitates a sophisticated control system.

5.2 Control system design

The lens control system is based on an Intel 80196 16 bit microcontroller with a minimum of external components. External digital inputs connect directly to the device for focus, zoom and iris demand (ie increment or decrement position) and an analogue output is generated using an external digital to analogue convertor (DAC) for a reference to the camera convergence servo. For any combination of zoom and focus, the controller positions the two moveable lens groups as required and outputs a voltage to the camera representing the corresponding convergence position, so that the camera converges at the distance where the lens is focused.

Motors for the two lens groups are connect to two of the 80196 high speed output (HSO) lines via MOSFET drivers and opto-isolators, driven in pulse width modulation (PWM) mode at a frequency determined by the software. Position feedback to close the control loop is provided by encoders connected via some decoding logic to the 80196 high speed input (HSI) lines. Any transitions on these inputs generates a software interrupt which the processor recognises and services by updating internal 32bit position registers. With both drives moving at maximum velocity this could occur at a maximum rate of about once every 0.4 mS. By using a clock rate of 12Mhz the software overhead for the interrupts is small enough to leave plenty of processor time to implement the control algorithm[5]. The control algorithm is a common PID (proportional, integral and derivative) algorithm slightly modified to improve stability by limiting the maximum size of the integral term and resetting it to zero every time the position error changes sign.

The design of the position encoders is somewhat novel. Using an LED angled along the length of a tooth and a photo diode to detect the reflected light, the encoders count the teeth of the driven gear attached to the leadscrew which actuates a group of lens elements. Since the gear would be incorporated no matter what means of position feedback were used this provides a very compact solution. The sensors work on a part of the gear which is not used for driving and has a slightly modified profile, such that the mark:space ratio of the light reflected as the gear rotates is approximately 50:50. Two such sensors are used per gear, arranged with their outputs in quadrature, thus enabling a rotation of 1/4 of a gear tooth to be resolved. Using a gear with 35 teeth and a lead screw with a 1.25mm pitch a linear movement of 9um can be resolved.

To generate an index, indicating that the lens is in a calibrated position, disks are attached to the drive and driven gears. By chosing carefully the number of teeth on each gear it is possible to arrange for a pair of holes in the disks to line up only once in a large number of rotations, more than are required for the full travel of the lens elements. If the reference position, indicated by the alignment of the index holes occurs in the middle of the range of travel and is detected by a sensor connected to a further HSI line, then the calibration of the lens can be checked whilst in operation, every time the index position is passed.

6.0 CONCLUSIONS

The reasons for designing a zoom lens from scratch for 3D television to be used in the nuclear industry have been examined. In summary these are:-

-To obtain the required optical performance in radiation stable glass

-Produce the highest possible quality 3D pictures, avoiding the alignment problems which might otherwise be encountered.

-To obtain a small compact unit which integrates easily with our camera designs.

-Ease of setting up by virtue of a flexible control system.

A basic optical configuration, mechanical construction and microprocessor control scheme has been outlined along with other features of the prototype lens to achieve these objectives. This is expected to be operational within the next 9-12 months.

7.0 REFERENCES

. A.A. Dumbreck, C.W. Smith, S.P. Murphy, "The Development and Evaluation of a Stereoscopic Television System or Use in Nuclear Environments," American Nuclear Society, International Topical Meeting on Remote Systems and Robotics in Hostile Environments, Pasco, WA, USA 106-113, March 1987.

2. C.W. Smith, A.A. Dumbreck "3D TV: The Practical Requirements," Television, Journal of the Royal Television Society Vol 25, 9-15, 1988.

3. R. Spottiswoode, N.L. Spottiswoode, C.W. Smith. "Basic Principles of Three-Dimensional Film," Journal of the SMPTE, Vol 59, 249-286, October 1952.

4. H. Davson, Physiology of the Eye, 3rd Edition, p409-412, Churchill Livingstone, London, 1972.

5. T. Schafer, M. Chevalier "Distributed Motor Control Using the 80C196KB", Application Note AP428, Intel Corporation 1989.

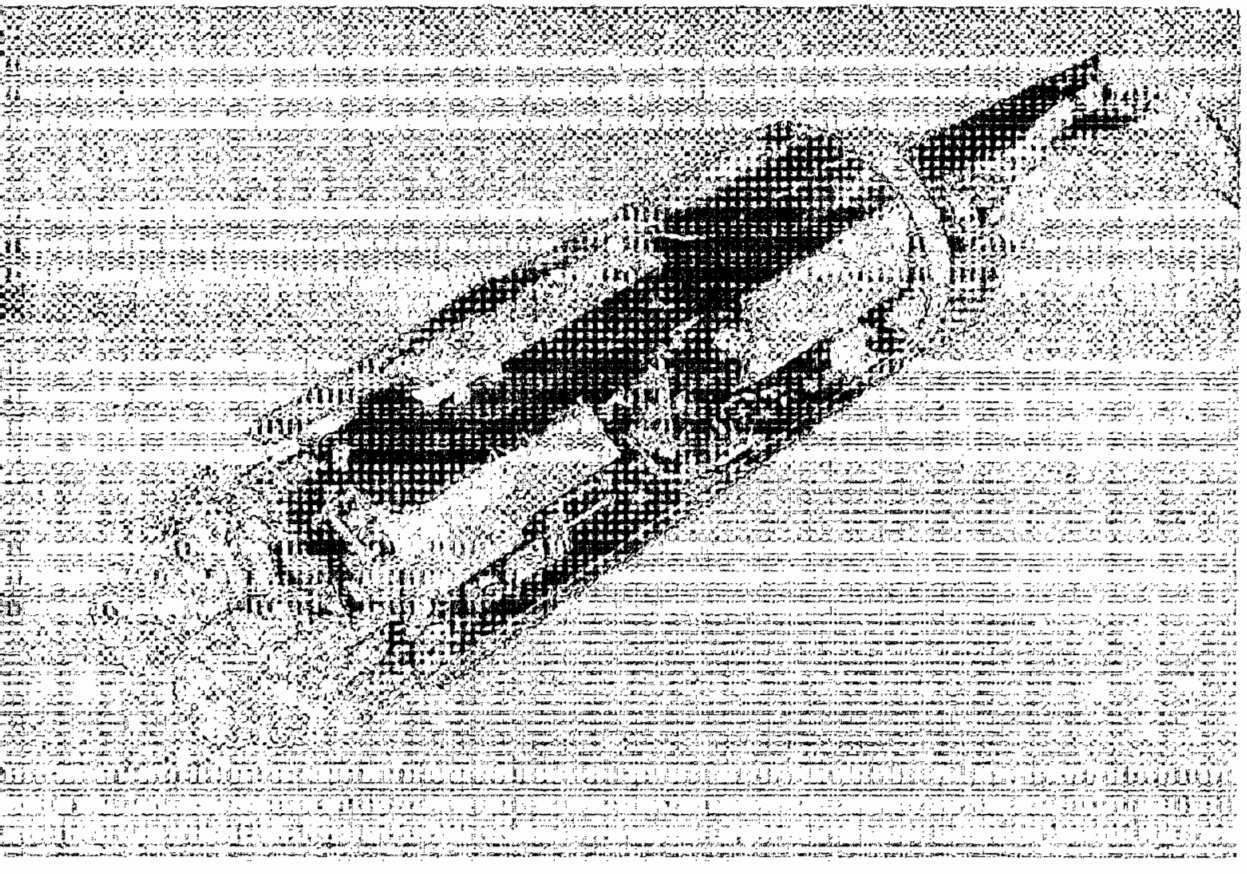


Figure 1
Two Angle of View Stereo Camera

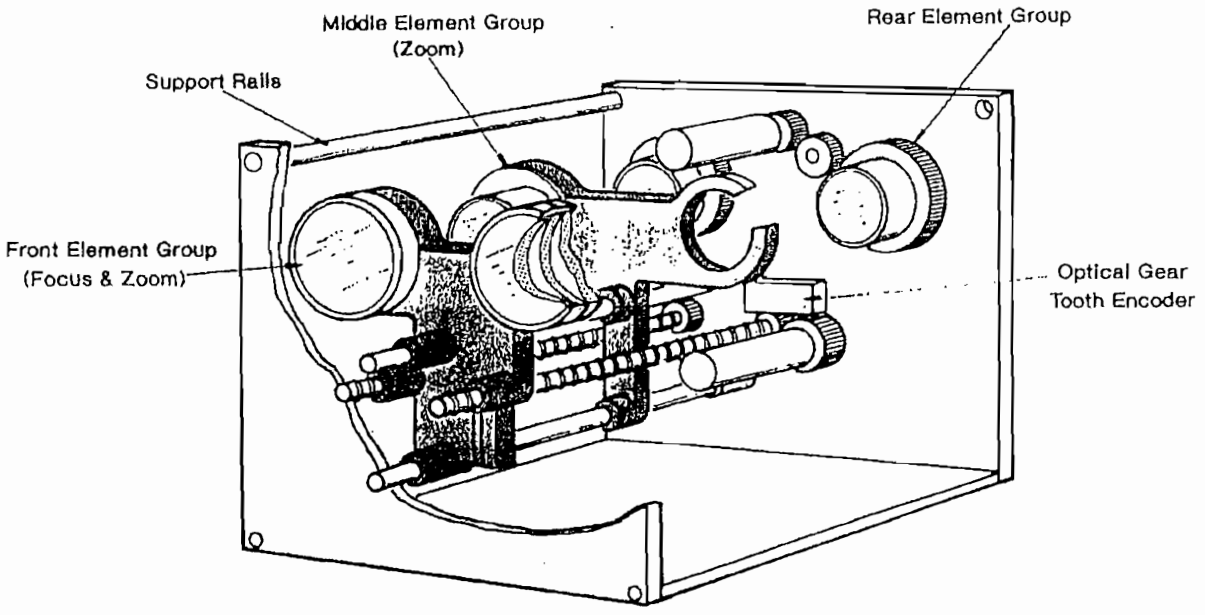


Figure 2
Cut-away of Zoom Lens

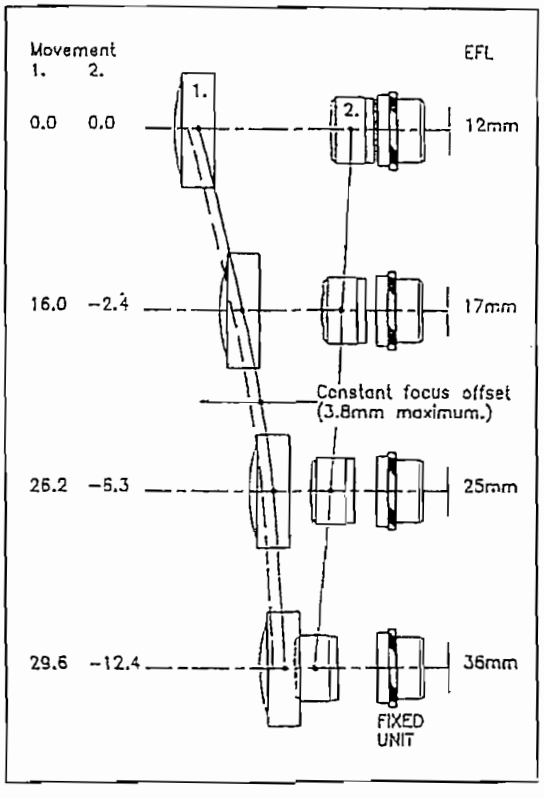


Figure 3
Path Followed by Lens Elements

SPIE Vol. 2177, Stereoscopic Displays and Virtual Reality Systems, ed. S S Fisher, J O Merritt, M T Bolas (Apr 1994) Copyright SPIE

Data Compression of an Autostereoscopic 3-D Image

T. Fujii *H. Harashima*

Department of Electrical Engineering, The University of Tokyo, Japan
7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan
TEL: +81-3812-2111 ext.6781, FAX: +81-3818-5706
Email: toshiaki@harashima.t.u-tokyo.ac.jp

ABSTRACT

This paper is concerned with the data compression and interpolation of multi-view image. In this paper, we propose a novel disparity compensation method based on geometric relationship. We first investigate the geometric relationship between a point in the object space and its projection onto view images. Then, we propose the diparity compensation scheme which utilize the geometric constraints between view images. This scheme is used to compress the multi-view image into the structure of the triangular patches and the texture data on the surface of patches. This scheme not only compresses the multi-view image but also synthesize the view images from any viewpoints in the viewing zone. Finally, we report the experiment, where three sets of multi-view image were used as original images and the amount of data was reduced to 1/20 with SNR 34 dB.

1. INTRODUCTION

Three dimensional television (3-D TV) will constitute the next stage after the arrival of High Definition Television. In order to achieve the 3-D image communication and broadcasting, the development of the 3-D image coding technique is important.

What is needed for 3-D image coding? First, the compression of 3-D image data is required before transmission and storing, because a 3-D display needs a tremendous amount of data to generate the stereoscopic visual effect. Secondly, it is desirable to be able to reconstruct the intermediate image to simplify the input system ,that is, to reduce the number of cameras. It is also desirable that the 3-D image data could be reduced to a common data form, which could be easily expandable to any data type, which is required by the type of display and the specification of the display.

Here, we assume the multi-camera and multi-viewpoint 3-D TV system as the standard 3-D TV system in the near future. Therefore, our final objective is the efficient coding of multi-view image sets and reconstruction of intermediate images. The problem to be discussed here is:

- how to compress the multi-view image data, and
- how to generate an intermediate image between view images.

Two approaches have been studied on the 3-D image coding: the waveform coding and the structure extraction coding.

First, we focus on the previous works based on the waveform coding. To date most of the work on 3-D image coding is intended for stereo image[2,3], and most of them employs the concept of the **disparity compensation**. Many algorithms have been developed in the past decade, where either motion estimation in a sequence of successive images or disparity estimation in stereo pairs is treated.

0-8194-1472-7/94/$6.00

On the other hand, only few studies have so far been made at multi-view image coding. The conventional 2-D video coding technique(e.g. H.261, MPEG) can be applied for multi-view image coding because the multi-view image is equivalent to moving image sequence. But we can not conclude that such an interframe coding scheme is optimal for multi-view image coding, since it aims at the coding of moving image sequence, not multi-view images. The first problem is that the interframe coding scheme utilize only the correlation between consecutive two frames. Although all of the view images are spatially related, the correlation among them is not considered. The second problem is that the reconstruction of the intermediate images is impossible.

Another approach on 3-D image coding employs the geometric relationship between multiple views and compresses the 3-D images using the structural properties (e.g. x-y-z coordinates and image brightness) of the 3-D object [6]. Although this scheme can easily reconstruct the intermediate view images, it is hampered by the difficulty of extracting the 3-D information, especially when the scene objects have complex shapes. Our final objective is the efficient coding of multi-view image sets and reconstruction of intermediate images, not the 3-D scene analysis. Therefore, this scheme should be evaluated by the coding efficiency (i.e. bit rate and signal-to-noise ratio (SNR)), but no works before introduced this evaluation.

In order to generate the intermediate image, we need the 3-D information of the scene. Conversely, the structure extraction coding should be evaluated by the SNR criterion, because our goal is not the 3-D scene analysis but the efficient coding.

In this paper, we propose a novel disparity compensation scheme based on the geometric relationship between view images, which not only compress the multi-view image with high SNR but reconstruct the intermediate images. In Section 2, we formulate the geometric relationship between view images. In Section 3 and 4, we review the disparity compensation and then propose a new disparity compensation of multi-view image utilizing geometric relationship. Section 5 explains the encoding/decoding algorithm. Experimental results on real images are given in Section 6. In Section 7, we introduce a segmentation to cope with the occlusion.

2. GEOMETRIC RELATIONSHIP BETWEEN VIEW IMAGES

A multi-view image consists of images taken by cameras which look at the same scene from the slightly different view angle (Fig. 1).

Figure 2 shows the configuration for obtaining the multi-view image. The object space is denoted by (x, y, z) and the image data from $n-th$ viewpoint is denoted by $I_n(m, l)$. multi-view image are taken at a distance of F from (0,0,0) on every c cm intervals. The film is parallel to $x-y$ plane, and the origin $(0, 0, 0)$ is assumed to be projected onto the origins of each view image $I_n(0, 0)$.

Here, we introduce two spaces shown in figure 2. First, we introduce a *multi-view image space* denoted by (n, m, l). This space is obtained by piling up the view images $I_n(m, l)$ according to n. Secondly, we introduce *normalized object space* denoted by (X, Y, Z). This space is obtained by normalizing the coordinates (x, y, z), where the coordinates X, Y are obtained by normalizing x, y with m, l , respectively, and the coordinate Z is obtained by normalizing by z with the displacement of the pixel in adjacent view images. In this space, any points on the $X-Y$ plane are assumed to be projected onto every view images $I_n(m, l)$, where $X = m$, $Y = l$. The normalized depth Z represents the amount of displacement between adjacent view images.

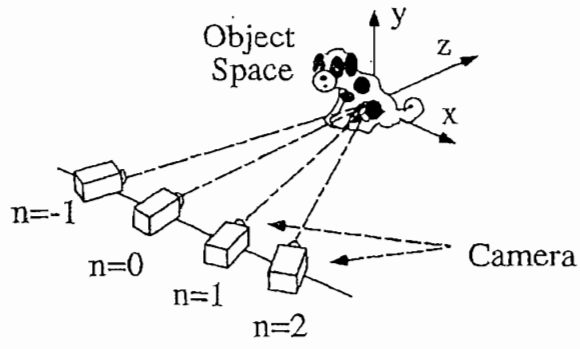


Figure 1. configuration for taking a multi-view image set

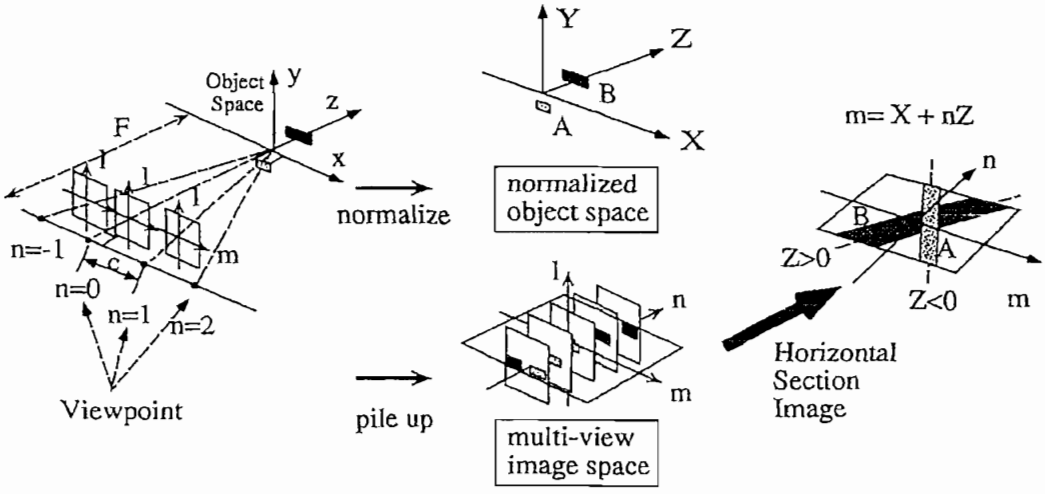


Figure 2. multi-view image space and normalized object space

Figure 2 shows the relationship between the *normalized object space* and the *multi-view image space*. We assume that the brightness of a point (X, Y, Z) be recorded on a pixel $I_n(m, l)$. The relationship between (X, Y, Z) and (n, m, l) is given by:

$$\begin{aligned} m &= X + nZ \\ l &= Y \end{aligned} \tag{1}$$

According to Eq. (1), the brightness data of a point (X, Y, Z) is recorded along the line in *multi-view image space*, (n, m, l).

To illustrate this, we show the horizontal section of *multi-view image space* in Fig. 2. In the figure, the objects A and B are recorded along the straight line in *multi-view image space*. The points having the same depth Z are aligned with the slope Z on the plane of $m - n$. Therefore, the brightness data on the plane $m - n$ construct a texture pattern which have many stripes with various slopes, being intersected and occluded mutually.

We can see that the *multi-view image space* (n, m, l) has the special features. First, the *multi-view image space* has the much correlation because one object is recorded in many views according to Eq. (1). Secondly, the *multi-view image space* (n, m, l) contain the structural information and the Z coordinates of points in the space can be determined by a texture analysis of this pattern.

The data compression problem of multi-view image set is how to compress the data in the *multi-view image space* (n, m, l), and the interpolation problem is how to synthesize the view image whose viewpoint n is not an integer.

3. BLOCK-BASED DISPARITY COMPENSATION OF MULTI-VIEW IMAGES

Presently, the motion compensation is widely used for 2-D video coding. This concept can be applied to stereo image coding. Figure 3 explains the concept of the disparity compensation. One image (e.g. left image) is subdivided into blocks and coded with the conventional coding method (e.g. discrete cosine transform) and transmitted separately. The other image (e.g. right image) is disparity compensated, in which correspondence is computed by the block matching method and the disparity and prediction error are encoded and transmitted.

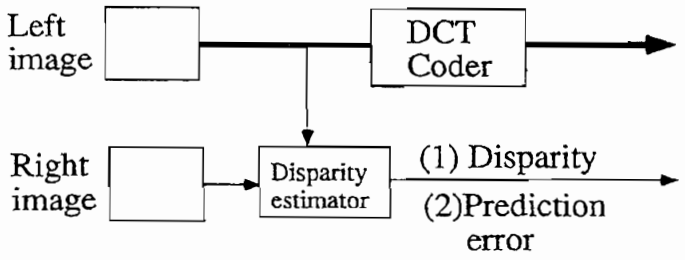


Figure 3. disparity compensation of stereo pair image

The concept of block-based disparity compensation can be extended for multi-view image using the geometric relationship between view images. We have seen that the *multi-view image space* (n, m, l) has the much correlation in n-axis direction. We propose a disparity compensation of multi-view image utilizing this correlation. This scheme not only compresses the multi-view image but also can generate the intermediate image, because geometric constraints are taken into account. Figure 4 explains the disparity compensation of multi-view image.

The coding/decoding process is as follows:

- Subdivide the central view image (i.e. $n = 0$)..
- Find the optimal slope (= depth Z) of each block. The optimal slope is determined in terms of the intensity variance along the line of Eq. (1). This slope determines the depth Z of the block.
- Determine the texture data by averaging texture data of the the corresponding blocks in all view images.
- Decoded images are obtained by projecting the blocks with texture using computer graphics (CG) procedure.

In this approach, multi-view image can be compressed into roughly one view image, and we can also reconstruct the intermediate image, if the disparity is estimated properly.

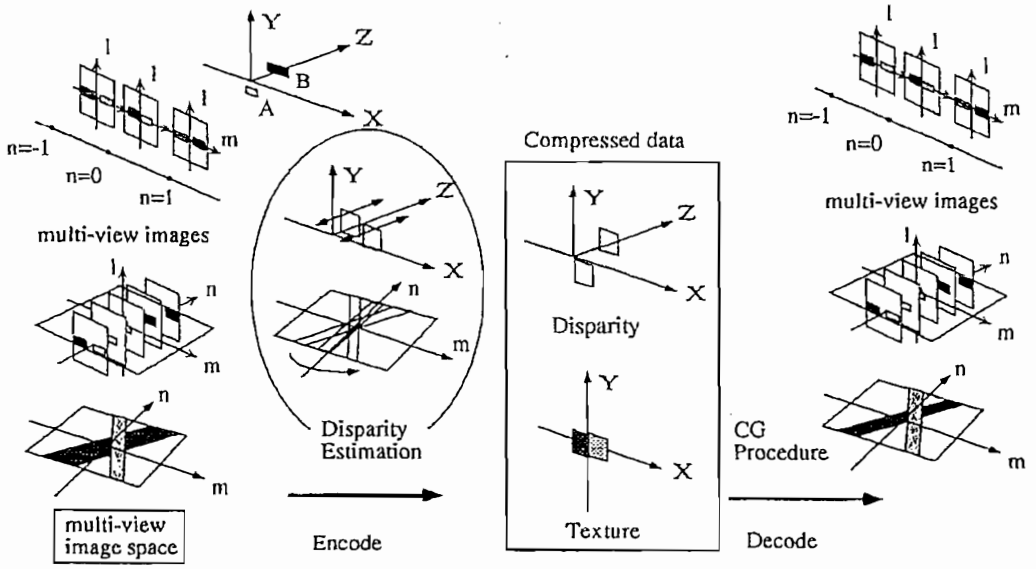


Figure 4. block-based disparity compensation of multi-view image

However, the problem arises: in the block-based disparity compensation, the size of the corresponding block is constant through all of the view images. The constant block size fails to account for the size variation of projections viewed from different angles, and furthermore, there may exist the artifacts in the decoded view image except the central view.

4. DISPARITY COMPENSATION BASED ON AFFINE TRANSFORMATION

In order to avoid the problem mentioned above, we propose a novel disparity compensation method which utilize triangular patches and affine transformation. In the 2-D video coding, some motion compensation methods that utilize triangular patches and affine transformation has been proposed (e.g. [4]).

Figure 5 shows the main idea of encoding/decoding process. First, the multi-view image are piled up to form the multi-view image space. Then, the multi-view image space is analyzed and compressed into the structure and texture data of triangular patches. The compressed data consist of the coordinates of grid nodes and texture data. The decoding process is done by projecting the triangular patches with texture onto the multi-view image space according to the specified viewpoint using computer graphics procedure. In this step, the intermediate view image can be generated, if necessary.

We should notice that the size of the projection of triangular patches can vary according to the viewpoints and that no artifact can be seen in decoded view image.

The point of this scheme is: (1) how to determine the shape of triangular patches, that is, how to determine the optimal position of grid nodes, and (2) how to determine the optimal texture data. We will consider this point next.

4.1. Variance space

The question which we must consider is how to determine the optimal positions of the grid nodes. To consider this question we introduce Avg-space and Var space.

Note that this analysis can be applied only to the point which is recorded in all views. Therefore, the shape function is determined so that any surfaces of triangles can be seen from any viewpoints in the viewing zone. This means that the difference of the depth Z between adjacent grid nodes is restricted to a certain value which is determined by the viewing angle.

5. ALGORITHM

In this section, we explain the coding algorithm. The encoding algorithm is as follows :

1. Calculate the Var-space
2. Cover $X-Y$ plane ($Z=0$) by triangular patches
3. Determine the optimal depth that the sum of $Var(X,Y,Z)$ value along the surfaces has minimum value
4. Map the averaged texture data (Avg value) on patches as the texture data

Note that the optimal positions of grid nodes are determined under the condition that all the surfaces of triangles can be seen from any viewpoint in the viewing zone.

The decoding is done by synthesizing the predicted image according to the specified viewpoint by projecting the texture data on the triangular patches using computer graphics procedure.

6. EXPERIMENTAL RESULTS

In the experiments, we used three multi-view image sets: "Toy Dog", "Garden-Flower 1", and "Garden-Flower 2".

In the "Toy Dog" experiment, view images were taken photographically. View data was obtained with a camera movable on a rail in the lateral direction. To conform with the simplified treatment, keystone and lens distortions of the raw data have been corrected and mutually registered. All the views were separately smoothed by a median filter. The number of views is 19 ($n=-9,-8,\ldots,-1,0,1,\ldots,8,9$) and the size of each image is 256×256 pixels.

In the "Garden-Flower" experiment, we used the ISO test sequence "Garden-Flower". Each image was geometrically transformed so that the epipolar line corresponds to the horizontal line. The number of views is 21 ($n=-10,-9,\ldots,-1,0,1,\ldots,9,10$) and the size of each image is 352×240 pixels. The decisive difference of two sequences is that "Garden-Flower 2" include a large tree in front of the flower garden. Figure 7 shows the original images (central views) of three multi-view image sets.

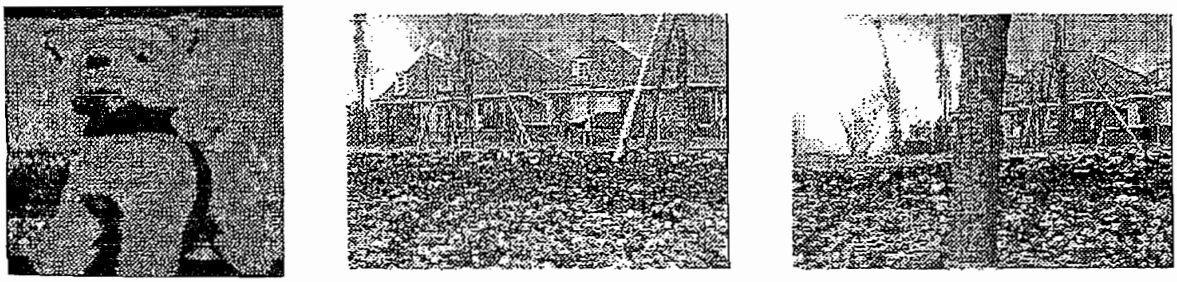


Figure 7. Original view images (center views: n=0). Toy Dog(left), Garden-Flower 1(middle), Garden-Flower 2(right).

We implemented the coding algorithm, "disparity compensation based on triangular patches and affine transformation". The algorithm was implemented in C with Silicon Graphics Iris Indigo Elan. The encoding process is done in a few minutes. Figure 8-10 shows the structures of the compressed data (shapes of the triangular patches and that with texture), and the decoded images. We obtained the averaged SNR 34 dB in the "Toy Dog" ,20 dB in the "Garden-Flower 1", and 17 dB in the "Garden-Flower 2" experiments, respectively. The compressed data contains patch size(1 byte), Z coordinates of every grid nodes(approximately 300 bytes), and the texture data (roughly the amount of one view image). The texture data is not compressed in this experiment. Therefore, original multi-view image data could be compressed to 1/(the number of multi-view image). This scheme not only compresses the multi-view image, but also reconstruct the intermediate image, as we have mentioned.

Through these experiments, we can conclude that this scheme is highly efficient when the scene is not so complex, that is, there exists no large occlusion in the scene. The "Toy Dog" and "Garden-Flower 1" images are suitable to this scheme.

On the other hand, the coding efficiency decreases very much when the occlusion occurs (e.g. the "Garden-Flower 2" experiment). From the viewpoint of structure recovery of 3-D object, this scheme can be viewed as the approximation of object space by a single polyhedron. In other words, this scheme compresses the multi-view image into one polyhedron and texture data on the surface. This is the reason why the prediction error increases in the occluded regions.

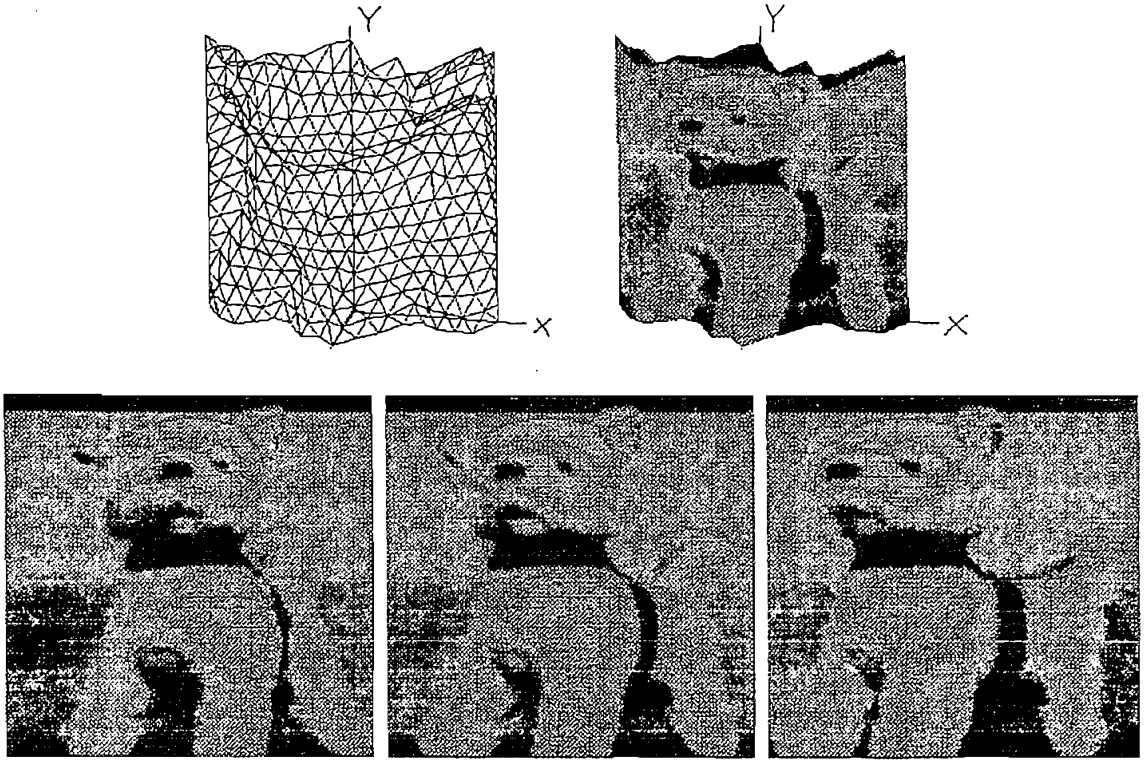


Figure 8. results of "Toy Dog", structure of the compressed data(above left), with texture(above right), and decoded images (n=-9, 0, 9)

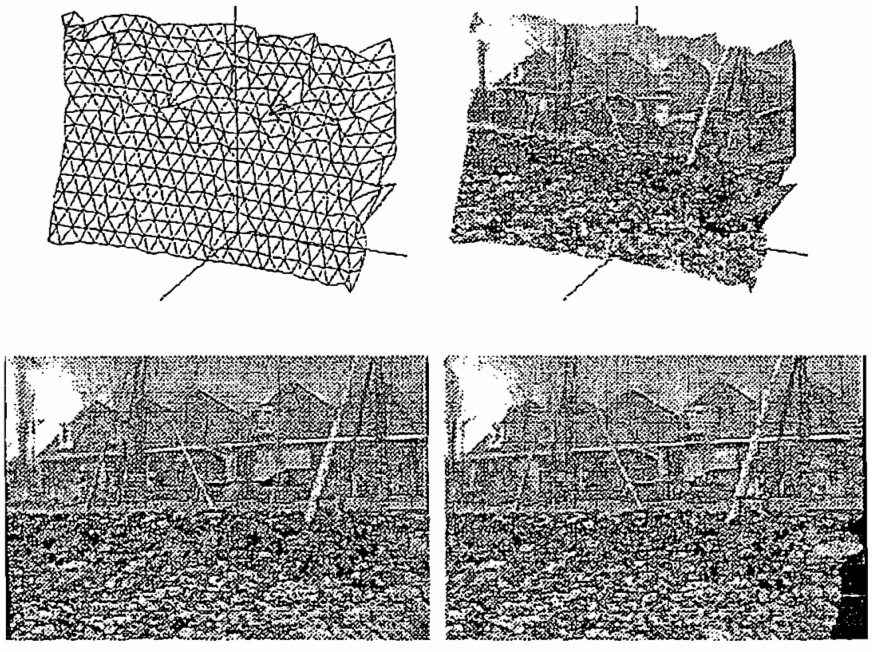


Figure 9. results of "Garden-Flower 1",structure of the compressed data, with texture, and decoded images (n=0, 10)

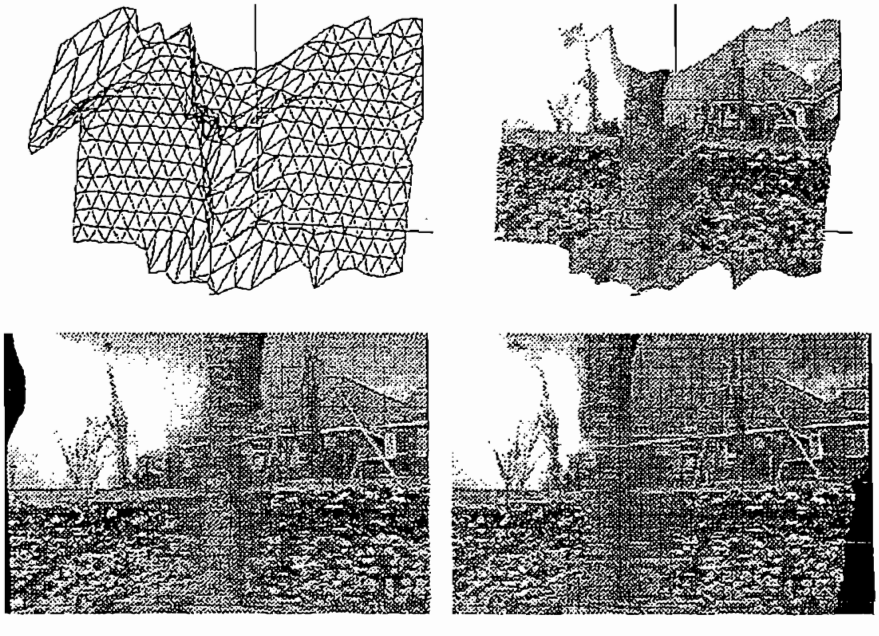


Figure 10. results of "Garden-Flower 2",structure of the compressed data, with texture, and decoded images (n=0, 10)

7. SEGMENTATION OF THE VAR–SPACE

In order to cope with the occlusion, we need to introduce the segmentation on the Var–space. The coding algorithm is as follows:

1. Calculate the Var space
2. Extract the regions. All the point in the region must have the Var value below the threshold and the size of the region must not be less than $4 \times 4 \times 4$ pixels.
3. Cover the extracted region by triangular patches.
4. Encode the structure of the triangular patches and the texture data.
5. Peel off the region from original view images.

This procedure is repeated until all the regions are peeled off.

We implemented this algorithm to the "Garden-Flower 2" image sequence. The computation time is about 10 minutes. Figure 11 shows the extracted region and the structure of the compressed data: shapes of the triangular patches and that with texture. In this experiment, the object space was divided into two regions : the region around the tree and the background.

Figure 11. extracted region: tree and background

8. CONCLUSION

We proposed a new 3-D image data compression scheme based on geometric relationship between view images. This method is viewed as hybrid of both disparity compensation and the structure estimation coding, and therefore, includes the advantages of both coding scheme: high SNR of the decoded image and the interpolation of view image. We also proposed the segmentation of Var–space to cope with the occlusion. The further development of the present method is promising for the data compression adaptable to many types of display for 3-D images in motion.

ACKNOWLEDGMENTS

I wish to express my gratitude to Prof. Hamasaki of Toa University for for his helpful suggestions on the present work.

REFERENCES

[1] M.E. Lukacs: "Predictive Coding of Multi-Viewpoint Image Sets", *ICASSP '86*, pp. 521-524 (1986).

[2] W.A.Schupp, Y.Yasuda: "Efficient coding of 3-D moving pictures with adaptive motion/disparity compensation", Journal of three dimensional images, Vol.3, No.1, pp.47-52(1989).

[3] M.G. Perkins: "Data Compression of Stereopairs", *IEEE Trans. Commun.*, Vol. 40, No. 4, pp. 684-696 (Apr. 1992).

[4] Y. Nakaya and H. Harashima: "An iterative motion estimation method using triangular patches for motion compensation", Proc. SPIE Visual Communications and Image Processing '91: Visual Communications, vol. 1605, pp.546-557, Nov. 1991.

[5] J.Hamasaki, M.Fukazawa, R.Ishima: "Sampling errors and data compression of multi-view lens-plate 3D images", SPIE Vol. 1319, pp.350-351, 1990, Germany.

[6] T. Fujii, J. Hamasaki, and M. Pusch: "Data compression of an autostereoscopic 3D image", The international symposium on three dimensional image technology and arts, Seiken symposium (Tokyo, February 1992).

[7] W.Richards: "Structure from stereo and motion", J. Opt. Soc. Amer. A., Vol. 2, No. 2, February 1985, pp.343-349.

[8] R. Skerjanc and J. Liu: "A three cameras approach for calculatin disparity and synthesizing intermediate pictures:, Signal Processing: Image Communication, Vol. 4, No. 1, Nov. 1991, pp. 55-64.

[9] R. Thoma and M. Bierling: "Motion compensating interpolation considering covered and uncovered background", Signal Processing: Image Communication, Vol. 1, No. 2, October 1989, pp. 191-212.

[10] T. Fujii, J. Hamasaki, and H. Harashima, "Data Compression for an Autostereoscopic 3-D Image", *PCS '93*, 13.21 (Lausanne, March 1993).

[11] Jin Liu, Robert Skerjanc, "Stereo and motion correspondence in a sequence of stereo images", Signal Processing: Image Communication, Vol. 5, No. 4,pp. 305-318 (Oct. 1993).

[12] T. Fujii, H. Harashima, "3-D Image Coding Based on Affine Transform", *ICASSP '94*, 81.8 (Adelaide, April), to appear.

SPIE Vol. 2177, Stereoscopic Displays and Virtual Reality Systems, ed. S S Fisher, J O Merritt, M T Bolas (Apr 1994) Copyright SPIE

Compression of stereo image pairs and streams

M. W. Siegel[1]
Priyan Gunatilake[2]
Sriram Sethuraman[2]
A. G. Jordan[1,2]

[1]Robotics Institute, School of Computer Science
[2]Department of Electrical and Computer Engineering
Carnegie Mellon University
5000 Forbes Ave., Pittsburgh, PA, 15213

ABSTRACT

We exploit the correlations between 3D-stereoscopic left-right image pairs to achieve high compression factors for image frame storage and image stream transmission. In particular, in image stream transmission, we can find extremely high correlations between left-right frames offset in time such that perspective-induced disparity between viewpoints and motion-induced parallax from a single viewpoint are nearly identical; we coin the term "WorldLine correlation" for this condition. We test these ideas in two implementations, (1) straightforward computing of blockwise cross- correlations, and (2) multiresolution hierarchical matching using a wavelet- based compression method. We find that good 3D-stereoscopic imagery can be had for only a few percent more storage space or transmission bandwidth than is required for the corresponding flat imagery.

1. INTRODUCTION

The successful development of compression schemes for motion video that exploit the high correlation between temporally adjacent frames, e.g., MPEG, suggests that we might analogously exploit the high correlation between spatially or angularly adjacent still frames, i.e., left-right 3D-stereoscopic image pairs. If left-right pairs are selected from 3D-stereoscopic motion streams at different times, such that perspective-induced disparity left-right and motion-induced disparity earlier-later produce about the same visual effect, then extremely high correlation will exist between the members of these pairs. This effect, for which we coin the term "WorldLine correlation", can be exploited to achieve extremely high compression factors for stereo video streams.

Our experiments demonstrate that a reasonable synthesis of one image of a left-right stereo image pair can be estimated from the other uncompressed or conventionally compressed image augmented by a small set of numbers that describe the local cross-correlations in terms of a disparity map. When the set is as small (in bits) as 1 to 2% of the conventionally compressed image the stereoscopically viewed pair consisting of one original and one synthesized image produces convincing stereo imagery. Occlusions, for which this approach of course fails, can be handled efficiently by encoding and transmitting error maps (residuals) of regions where a local statistical operator indicates that an occlusion is probable.

Two cross-correlation mapping schemes independently developed by two of us (P.G. and S.S.) have been coded and tested, extensively on still image pairs and more recently on some motion video streams. Both methods yield comparable compression factors and visual fidelity; which can be coded more efficiently, and whether either can be coded efficiently enough to make it practical for real time use, is under study.

0-8194-1472-7/94/$6.00

The method developed by P.G. is based on straightforward computing of blockwise cross-correlations; heuristics that direct the search substantially improve efficiency at the price of occasionally finding a local maximum rather than the global maximum.

The method developed by S.S. is based on multiresolution hierarchical matching using wavelets; efficiency is achieved by doing the search for the best match down a tree of progressively higher resolution images, starting from a low resolution highly subsampled image.

In the following sections we discuss the need and opportunity for compression of 3D-stereoscopic imagery, discuss the correlations that can be exploited to achieve compression, describe and refine the approach, summarize the content and performance of the two implementations we have prototyped to date, and outline several topics we have targeted for ongoing research.

This paper is intended as a high level introduction to our thoughts about and our progress toward compression for 3D-stereoscopy. The specific references that we cite in the text and the general references that we also include in the bibliography point to background literature, as well as to three recent papers [5, 6, 7] in which we document the low level details of our recent work.

2. NEED AND OPPORTUNITY

The scenario we imagine is that binocular 3D-stereoscopy is grafted onto "flat" (monoscopic) display infrastructures; we regard the alternative scenario, that 3D-stereoscopy is built into the foundations of the infrastructure, as being somewhat farfetched in light of the cost and effectiveness of the current generation of 3D display devices and systems.

Displays become rapidly more expensive as their spatial resolution and temporal frame rate increases. Thus in any application the display is usually chosen to meet but not to exceed substantially the application's requirements. In flat applications each eye sees, at no cost to the other eye, the full spatial and temporal bandwidth that the display delivers. When a 3D-stereoscopic application is grafted onto a flat infrastructure the display's capabilities must be divided between the two eyes. The price may be extracted in either essentially the spatial domain, e.g., by assigning the odd lines to the left eye and the even lines to the right eye, or in essentially the temporal domain, e.g., by assigning alternate frames to the left and right eye. The distinction is in part semantic, since the "spatial" method of this example is often implemented in practice via sequential fields in an interlaced display system. The fundamental issue is that when 3D-stereoscopy is implemented on a single display each eye gets in some sense only half the display. A user contemplating using 3D-stereoscopy must thus acquire a display (and the underlying system to support it) with twice the pixel-per-second capability of the minimal display needed for the flat application; the alternatives require choosing between a flickering image or a reduced spatial resolution image.

As indicated, lower level capacities of the system's components must also be doubled. In particular, all the information captured by two cameras (each equivalent to the original camera) must be stored or transmitted or both. Doubling these capacities may be more difficult than doubling the capability of the display, inasmuch as (except at the very high end) the capability of the display can be increased by simply paying more. The most difficult system component to increase is probably the bandwidth of the transmission system, which is often subject to powerful regulatory as well as technical

constraints. Nevertheless, the bandwidth must apparently be doubled to transmit 3D-stereoscopic image streams at the same spatial resolution and temporal update frequency as either flat image stream.

In fact, because the two views comprising a 3D-stereoscopic image pair are nearly identical, i.e., the information content of both together is only a little more than the information content of one alone, it is possible to find representations of image pairs and streams that take up little more storage space and transmission bandwidth than the space or bandwidth that is required by either alone. The rest of this paper is devoted to an overview of how this can be done, some details of our early implementations, and a discussion of possibilities for the future.

2.1. Background

We remind the reader that image compression methods fall into two broad categories, "lossless" and "lossy". Lossless compression exploits the existence of redundant or repeated information, storing the image in less space by symbolically rather than explicitly repeating information, and by related methods such as assigning the shortest codes to the most probable occurrences. Lossy compression exploits characteristics of the human visual system by discarding image content that is known to have little or no impact on human perception of the image.

Our approach to compression of 3D-stereoscopic imagery has two components, related to there being two perspective views in a 3D-stereoscopic pair. One component may be either lossless or slightly lossy, as in conventional compression of flat imagery; the other component is by itself a very lossy (or "deep") method of compression. The intimate connection between the two views makes it possible to synthesize a perceptually acceptable image from a compression so deep that, by itself, it would be incomprehensible.

The left and right views that comprise a 3D-stereoscopic image pair or motion stream pair are obviously very similar. There are various ways of saying this: they are often described as "highly redundant", in that most of the information contained in either is repeated in the other, or as "highly correlated" in that either is for the most part easily predicted from the other by application of some external information about the relationship (the relative perspective) between them. We can thus synthesize a reasonable approximation to either view given the other view and a little additional information that describes the relationship between the two views. A useful form for the additional information is a disparity map: a two dimensional vector field that encodes how to displace blocks of pixels in one view to approximate the other view.

Fortunately a "reasonable approximation" is enough: perfection is not required. This is the case because of two psychophysical effects, one well known, the other less so.

It is well known that one good eye and one bad eye together are better than the good eye alone, i.e., the information they provide in a sense adds rather than averages. The resulting perception is sharper than the perception provided by the better eye alone. Thus presenting one eye with the original view intended for it, and presenting the other eye with a synthetic view (which might be imperfect in sharpness and perhaps even missing some small features), the perception of both together is better than the perception of the original view alone.

A related perceptual effect that we have observed informally has been documented in several controlled experiments: a binocular 3D-stereoscopic image pair with one sharp member and one blurred member successfully stimulate appropriate depth perception.

Thus we expect that if one member of a 3D-stereoscopic image pair is losslessly or nearly losslessly compressed and the other is (by some appropriate method) deeply compressed, the pair of decompressed (higher resolution) and synthesized (lower resolution) views will together be perceived comfortably and accurately.

In the following section we describe several approaches to compression, ultimately focusing on the method we are now developing along two complementary implementation paths.

2.2. Correlations

We identify four kinds of correlations or redundancies that can be exploited to compress 3D-stereoscopic imagery. The first two make no specific reference to 3D-stereoscopy; they are conventional image compression methods that might (inefficiently!) be applied to two 3D-stereoscopic views independently. The third kind applies to still image pairs, or to temporally corresponding members of a motion stream pair. The fourth kind, which is really a combination of the second and third kinds, applies to motion stream pairs.

- ***Spatial correlation***: Within a single frame, large areas with little variation in intensity and color permit efficient encoding based on internal predictability, i.e., the fact that any given pixel is most likely to be identical or nearly identical to its neighbors. This is the basis for most conventional still image compression methods.
- ***Temporal correlation***: Between frames in a motion sequence, large areas in rigid-body motion permit efficient coding based on frame-to-frame predictability. The approach is fundamentally to transmit an occasional frame, and interpolation coefficients that permit the receiver to synthesize reasonable approximations to the intermediate frames. MPEG is an example.
- ***Perspective correlation***: Between frames in a binocular 3D-stereoscopic image pair, large areas differing only by small horizontal offsets permit efficient coding based on disparity predictability. If one imagines the two perspective views as being gathered not simultaneously but rather sequentially by moving the camera from one viewpoint to the second, then perspective correlation and temporal correlation are to first order equivalent.
- ***WorldLine correlation***: We borrow the term "worldline" from the Theory of Special Relativity, where the worldline is a central concept that refers to the path of an object in 4-dimensional space-time. Observers moving relative to each other, i.e., observers having different perspectives on space-time, perceive a worldline segment as having different spatial and temporal components, but they all agree on the length of the segment. Analogously in 3D-stereoscopic image streams, when vertical and axial velocities are small and horizontal motion suitably compensates perspective, time-offset frames in the left and right image streams can be nearly identical. WorldLine correlation is the combination of temporal correlation and perspective correlation; the most interesting manifestation of WorldLine correlation is the potential near-identity of appropriately time-offset frames in the left and right image streams respectively.* The concept is useful for situations in which the camera is fixed and parts of the scene are in motion, the scene is fixed and the camera is in motion, and both the camera and parts of the scene are in motion.

WorldLine correlation is depicted pictorially in Figure 1.

*Thinking in a suitable generalized fourier domain, simultaneous pairs from different perspectives and pairs from one perspective at different times are characterized by nearly identical amplitude spectra but substantially (although systematically) different phase spectra.

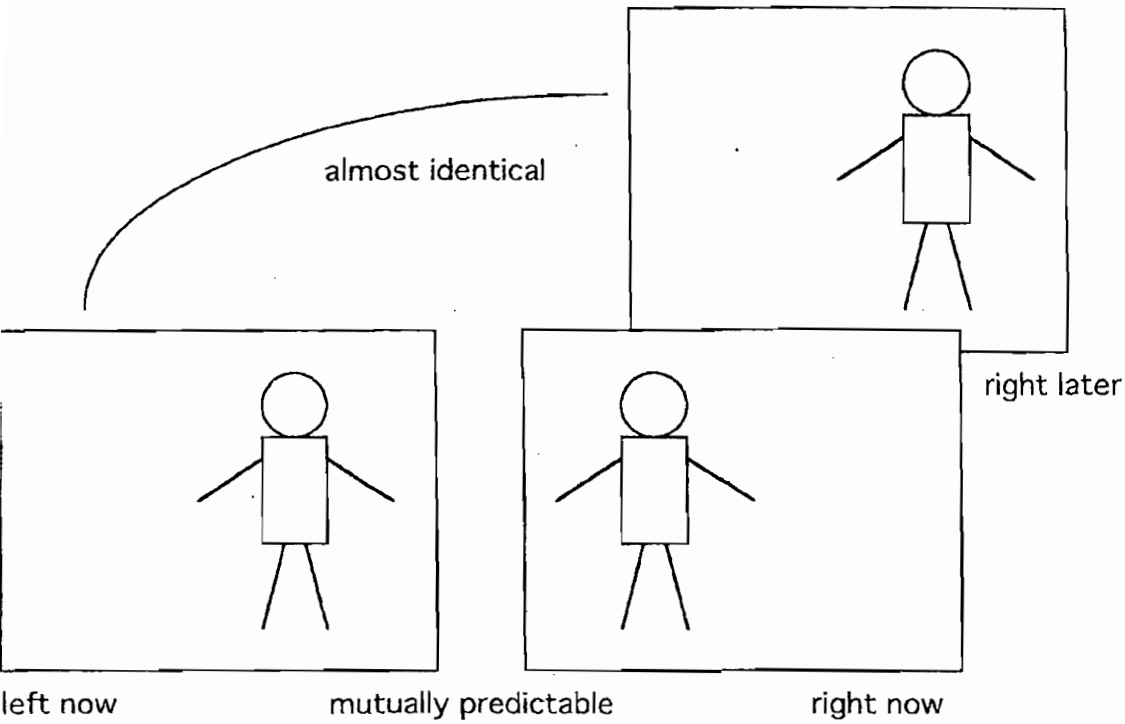


Figure 1: Pictorial depiction of WorldLine correlation.

3. APPROACH

3.1. Basic Approach

Our basic approach to compression of 3D-stereoscopic imagery is based on the observation that disparity, the relative offset between corresponding points in an image pair, varies only slowly over most of the image field. Given the validity of this assumption, either member of an image pair can be synthesized (or "predicted") given the other member and a low-resolution map of the relative disparity between the two members of the pair. It is the possibility that the disparity map can be low resolution, combined with the fact that the disparities vary slowly and can be represented by small numbers (few bits) that permits deep compression.

As a numerical example, suppose that over most of the image field the disparity does not change significantly over eight pixels. Then a disparity map can be represented by a field with 1/64 the number of entries as the image itself. Each disparity is a vector with two components, horizontal and vertical, so the net compression has an upper bound of 1/32, about 3%. In fact further significant advantages can be obtained by recognizing that the disparity components can be encoded with fewer bits than the original intensities, e.g., perhaps three bits for the vertical disparities (four pixels up or down) and perhaps five bits for the horizontal disparities (sixteen pixels left or right). Removal of redundancy in this map, e.g., run length encoding, leads to even further gains.

Our basic approach to coding 3D-stereoscopic image pairs, or corresponding pairs of a 3D-stereoscopic image stream, is easily seen from the following outline:

- Generate:
  - Code either image conventionally
  - Code the disparity map
- Store/Move:
  - Transmit the coded components
- Use:
  - Decode the conventionally coded image
  - Decode the disparity map
  - Synthesize the other image
  - Display 3D-stereoscopically

3.2. Problem with the Basic Approach

The basic approach has a basic fault: it cannot cope with occlusions, i.e., features that can be seen from only one of the two perspectives. This follows simply from the fact that the synthesized view is just a "rubber sheet" map of the conventionally compressed view. Thus features that are occluded in the conventionally compressed view (visible only in the view that is subsequently deeply compressed) cannot be synthesized. Similarly, features that are visible in the conventionally compressed view but are occluded in the subsequently deeply compressed view do not fit comfortably into this scheme.

The human visual perception system has an effective way to deal with occlusions: we have a detailed understanding of the image semantics, from which we effortlessly and unconsciously draw inferences that fill in the missing information. If this capability could be duplicated in a computer algorithm it would be essentially the solution to the general image understanding problem; its pursuit, let alone its solution, is beyond the scope of the present work.

Fortunately a pragmatic alternative exists: we can code and transmit the residuals (a map of the pixel-by-pixel differences between the original and its prediction from the disparity map). The differences are usually small, permiting it to be coded efficiently by conventional methods. In fact we can achieve a particularly efficient implementation in either of two equivalent ways. Both approaches work by coding and transmitting the residuals only in limited regions. In one approach the residuals are preserved only where they exceed a predetermined threshold. In the other approach a local statistical operator is used to identify regions in the image where occlusions are probable, and the residuals are computed, coded, and transmitted only for these regions.

3.3. Resulting Hybrid Method

The result is a hybrid algorithm whose flow should be clear from the preceding discussion, but which we will outline explicitly for completeness:

- Generate:
  - Code one image conventionally
  - Code the disparity map
  - Code the residuals of the prediction
- Store/Move:
  - Transmit the coded components
- *Use:*
  - Decode the conventionally coded image
  - Decode the disparity map
  - Synthesize the other image
  - Decode the residuals
  - Add the residuals to the prediction
  - Display 3D-stereographically

We are also conducting several subsidiary experiments aimed at understanding how the detailed coding scheme can be optimized for the human perceptual system. For example, it seems plausible that rapidly alternating which eye sees the conventionally compressed view and which eye sees the deeply compressed view may be more comfortable than fixing this choice. We are testing this and comparable hypotheses.

4. IMPLEMENTATION AND RESULTS

We have implemented two methods and are experimenting with them in parallel.

The first method, implemented by P.G., uses straightforward blockwise cross-correlation. This is the obvious candidate for initial experiments because it is easy to code and because we have a strong intuitive understanding of its parameters. It is thus straightforward to experiment with and understand the results of varying the parameters. In this implementation simple heuristics efficiently direct the matching search, decreasing the run time of the algorithm; however, as expected, avoiding exhaustive search makes the method somewhat prone to finding erroneous local matches.

The second method, implemented by S.S., uses a wavelet-based multiresolution hierarchical matching approach. The high spatial frequency content of the image is preserved at half the initial resolution; despite its high resolution, it can be coded efficiently because pixel values differ significantly from zero only in the immediate vicinity of the edges in the original image. The low spatial frequency content of the image is preserved in reduced resolution images. High and low frequency sub-images are computed down several hierarchical levels. The disparity map is built from the bottom up in a coarse-to-fine updated search; it is thus robust against finding incorrect local matches. It is computationally efficient, essentially because compression and disparity map building make use of the same intermediate results. Its hierarchical structure permits graceful degradation with lower-capability displays or noisy channels.

To date we have demonstrated in both implementations:

- Acceptable binocular perception with 1 to 2% of the total bandwidth allocated to disparity coding, and
- Excellent binocular perception with 10 to 20% of the total bandwidth allocated to disparity and residual coding.

For example, Figure 2 shows an original right and left 3D-stereoscopic image pair, and Figure 3 shows the right image after

conventional compression and decompression and the left image synthesized from the left member of Figure 2 and the disparity map computed (by the simple block matching method) from the left and right members of Figure 2.

Figure 2: Original Left and Right Views

We expect that in our ongoing work compression depth and synthesis fidelity will both increase substantially.

Topics that we need to address in the context of compression of 3D-stereoscopic imagery include:

- Optimizing implementation of the WorldLine approach.
- Optimizing the left-right alternation sequence of conventionally coded and synthesized views.
- Addressing asymmetric resource issues (consequences of the fact that we can afford more hardware at the coding side than at the decoding side)
- Addressing delay penalties (which are relatively unimportant for unidirectional broadcast, but which are a serious problem for real-time two-way communication and teleoperation)
- Implementing formal performance evaluation using appropriate statistical measures of compression effectiveness.
- Implementing psychophysical performance evaluation using appropriate human factors experimental methods and measures.

Topics we intend to pursue later with a view toward long-term payoffs include:

Figure 3: Synthesized Left and Decompressed Right Views

- Using three cameras: compute predictors for left and right views given the middle view, transmit the middle view and the predictors, synthesize 3D-stereoscopic views at the receiver. This approach has several practical advantages including compatibility with flat display systems and ease of adapting the 3D-stereoscopic rendering to the preferences and visual abilities of the viewer.
- Object based methods: apply the methods of machine vision and automated image understanding to augment deeply compressed imagery with semantic information that is used at the receiver to synthesize apparently losslessly transmitted imagery; it should be obvious that this is an extremely ambitious goal.

5. CONCLUSIONS AND PLANS

Because they are highly redundant, binocular 3D-stereoscopic image streams can be encoded in little or no more s .ce (transmitted in little or no more bandwidth) than either component stream.

Single step and hierarchical encoding methods produce psychophysically pleasing imagery.

Future research will address in the short term fine-tuning the architectures and algorithms and understanding their fundamental mathematical and psychophysical efficiencies, and in the long term issues such as multiple camera schemes and object based compression methods.

6. ACKNOWLEDGEMENTS

The ideas discussed in this paper were refined in the course of many discussions with (alphabetically) Tom Ault, Victor Grinberg, Alan Guisewite, Joe Mattis, Jeff McVeigh, Steve Roth, and Scott Safier. This work was funded by ARPA High Definition Systems Grant MDA972-92-J-1010.

7. REFERENCES AND BIBLIOGRAPHY

[1] I. Dinstein, J. Tselgov, et al.
Compression of Stereo Images and the Evaluation of Its Effects on 3-D Perception.
In *SPIE Applications of Digital Image Processing*, pages 522-530. Polytechnic University, Electrical Engineering Dept. and Ben-Gurion University, Behavioral Sciences Dept., Brooklyn, NY and Beer Sheva, Israel, 1989.

[2] I. Dinstein, J. Tselgov, et al.
On Stereo Image Coding.
In *Ninth International Conference on Pattern Recognition.* IEEE Computer Society, Beer Sheva, Israel, 1988.

[3] Michael G. Perkins.
Data Compression of Stereopairs.
In *IEEE Transactions on Communications, Vol. 40, No. 4*, pages 684-696. Apr, 1992.

[4] Oliver Rioul and Martin Vetterli.
Wavelets and Signal Processing.
IEEE SP Magazine :16-38, Oct, 1991.

[5] Priyan Gunatilake, A. G. Jordan, and M. W. Siegel.
Compression Technique for 3-D Stereo Video Streams.
In Metin Akgun (editors), *International Workshop on HDTV '93 (Ottawa)*, pages TBD. IEEE, SMPTE, EURASIP, ITE, EiC, Elsevier Science Publishers, Ottawa, Ontario, Canada K2H8S2, October 26-28, 1993.
Accepted.

[6] Sriram Sethuraman, A. G. Jordan, M. W. Siegel.
Multiresolution based hierarchical disparity estimation for stereo image pair compression.
In A N Akansu (editor), *Applications of SubBands and Wavelets*, pages TBD. IEEE, IEEE, NJIT ECE Dept, University Heights, NJ 07102, March, 1994.
Accepted.

[7] Sriram Sethuraman, M. W. Siegel, and A. G. Jordan.
A multiresolution framework for stereoscopic image sequence compression.
In J. Woods et al (editors), *Proceedings of the 1994 International Conference on Image Processing (Austin TX)*, pages tbd. IEEE/ICIP'94, IEEE, IEEE, November, 1994.
Submitted.

[8] R. Skerjanc and J. Liu.
A three camera approach for calculating disparity and synthesizing intermediate pictures.
In *Signal Processing: Image Communication 4*, pages 55-64. Elsevier, Heinrich-Hertz Institute, Berlin, GERMANY, 1991.

[9] A. Tamtaoui and C. Labit.
Schemas de compression de sequence d'images stereoscopiques par compensation de mouvement et disparite.
In *Journees de la Television en Relief*, pages . Elsevier, CCETT, Rennes, FRANCE, 1990.

[10] A. Tamtaoui and C. Labit.
Constrained disparity and motion estimators for 3DTV image sequence coding.
In *Signal Processing: Image Communication 4*, pages 45-54. Elsevier, IRISA/INRIA, Rennes Cedex, FRANCE, 1991.

[11] A. Tamtaoui and C. Labit.
Coherent disparity and motion compensation in 3DTV image sequence coding schemes.
In *ICASSP '91*, pages . Elsevier, IRISA/INRIA, Rennes Cedex, FRANCE, 1991.

[12] K. Metin Uz, Martin Vetterli, and Didier J. LeGall.
Interpolative Multiresolution Coding of Advanced Television with Compatible Subchannels.
In *IEEE Transactions on Circuits and Systems for Video Technology, Vol. 1, No. 1*, pages 86-99. Mar, 1991.

[13] Hiroyuki Yamaguchi, et al.
Stereoscopic Images Disparity for Predictive Coding.
In *Proceedings ICASSP 1989*, pages 1976-1979. Osaka, JAPAN, 1989.

SPIE Vol. 3295, Stereoscopic Displays and Virtual Reality Systems V, ed. M T Bolas, S S Fisher, J O Merritt (Apr 1998) Copyright SPIE

Depth controlled 3D-TV image coding

Armando Chiari[a], Bruno Ciciani[b], Milton Romero[b], Riccardo Rossi[a]

[a]Fondazione Ugo Bordoni, Rome, Italy, [b]Università di Roma "La Sapienza", Rome, Italy

ABSTRACT

Conventional 3D-TV codecs processing one down-compatible (either left, or right) channel may optionally include the extraction of the disparity field associated with the stereo-pairs to support the coding of the complementary channel. A two-fold improvement over such approaches is proposed in this paper by exploiting the three-dimensional features retained in the stereo-pairs to reduce the redundancies in both channels, and according to their visual sensitiveness. Through an a-priori disparity field analysis, our coding scheme separates a region of interest from the foreground/background in the volume space reproduced in order to code them selectively based on their visual relevance. Such a region of interest is here identified as the one which is focused by the shooting device. By suitably scaling the DCT coefficients in such a way that precision is reduced for the image blocks lying on less relevant areas, our approach aims at reducing the signal energy in the background/foreground patterns, while retaining finer details on the more relevant image portions. From an implementation point of view, it is worth noticing that the system proposed keeps its surplus processing power on the encoder side only. Simulation results show such improvements as a better image quality for a given transmission bit rate, or a graceful quality degradation of the reconstructed images with decreasing data-rates.

Keywords: Stereoscopic television coding, 3D-TV image coding, Disparity map estimation, Video bit-rate control, Image graceful degradation.

1. INTRODUCTION

Recently growing efforts have been spent on coding schemes for the compression of stereoscopic video signals in multimedia environments[1], including communications channels or storage devices. In the so called "backwards compatible" codecs a conventional coding technique is applied to one channel (either left, or right), taking the function of a down-compatible (i.e. monoscopic TV); the disparity field associated with the stereo-pairs is optionally extracted to support the coding of the complementary channel.

A two-fold improvement over such approaches is proposed in this paper by exploiting the three-dimensional features retained in the stereo-pairs to reduce the redundancies: 1) in both channels, 2) according to their visual sensitiveness. The basic idea is to separate a region of interest from the foreground/background in the volume space reproduced in order to code them selectively based on their visual relevance. This aims at reducing the signal energy in the background/foreground patterns, while retaining finer details on the more relevant image portions. Such a region of interest is here identified as the one which is focused by the shooting device; this assumption is in accordance with the known 3DTV production grammar rule[2] associating the observer capability of stereo fusion with focused image areas.

In order to achieve such goals, in this work the structure of a 3DTV coder has been developed, which basically exploits the spatial correlation between the left and right channels, and includes the concept of down-compatibility; moreover the new feature is supported of a selective coding capability, in that different quality levels can be reproduced within each single frame: this is obtained by scaling the DCT coefficients in such a way that precision is incremented / reduced for the image blocks lying on more / less relevant areas respectively. A segmentation of the scene into depth slices of different visual interest is guided by an estimation of the disparity map for each stereo-pair. In this work the extraction of the

Further author information -
A.C., R.R.: Email: {chiari,riccardo}@fub.it Phone: +39 (6) 5480 2136 Fax: +39 (6) 5480 4401
B.C., M.M.: Email: {ciciani,miltonr}@dis.uniroma1.it Phone: +39 (6) 4991 8325 Fax: +39 (6) 8530 0849

Part of IS&T/SPIE's Stereoscopic Displays and Applications IX • San Jose, California, USA • January 1998
SPIE Vol. 3295 • 0277-786X/98/$10.00

disparity map is based on a quad-tree algorithm[3] to both improve the disparity accuracy and decrease the computational complexity with respect to the full block-matching algorithm.

To fine tune the production of the video code according to a specified channel data rate, a novel algorithm for the control of the transmission buffer has been developed, which is also computationally efficient (binary search).

The description of the basic components of our codec is outlined in the following sections: in section 2 the psycho-visual criterion supporting the selective coding, as well as the system principles of operations are introduced; in section 3 the architecture of the encoder is presented, and a description is given of its main components from an algorithmic point of view: a quad-tree based disparity evaluation module is discussed, which is oriented to a real-time semi-systolic[4] structure, a novel algorithm for the control of the stereoscopic video output data rate is also introduced; the performance evaluation of the single functional modules is discussed in their respective sections, whereas the overall system performances are reported in section 4, according to our simulation results; in section 5 some future research related items are proposed; finally in section 6 we outline some conclusions.

2. PRINCIPLES OF OPERATION

A basic property of optical systems, which are employed in video cameras, is their capability of focusing a limited volume space; objects within such a space yield a sharp representation, whereas some blur affects other image portions. It is also a motion picture and television programmer production established rule, to selectively focus the relevant object in a scene; in this case objects out of focus, often on the background and/or foreground, are meant as less meaningful, and observers are naturally let to concentrate on the focused portion of the scene only. The above considerations suggest the possibility to code images after such a psycho-visual criterion: to this purpose a separation of the image contents into foreground, medium distance, background objects is performed, by evaluating the stereo-pairs depth field. Comparing the estimated distance of objects to the camera focus plane position results in the identification of the depth layer of interest.

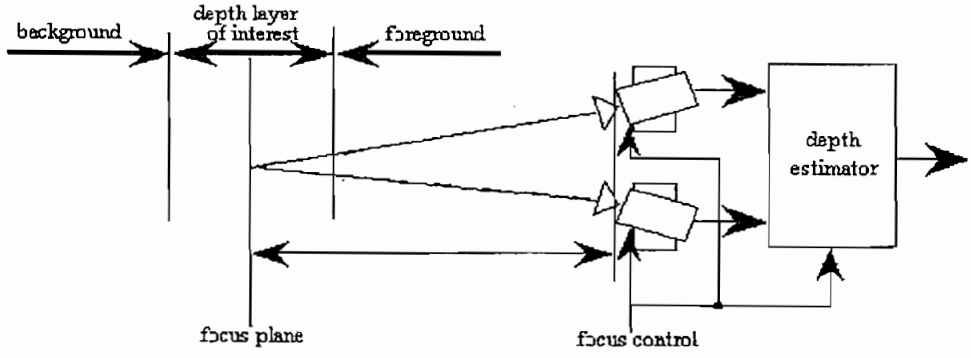


Fig. 1 - Shooting system.

In fig. 1 the principle of operations of the system is described: a stereoscopic video camera is interconnected to a digital application specific processor, taking the function of estimating the depth of the corresponding pixels in the stereo pairs. In order to provide a reference to compare with, a signal coding the camera focus depth is provided to the processor. It should be noted that the matching algorithm selected for the depth analysis may act on luminance signals only, rather than on the full color component signals, which results in a dramatic reduction of the processor hardware complexity.

3. CODEC ARCHITECTURE

3.1 ENCODER

In the functional block diagram shown in fig. 2, a separate coding is assumed for the left channel only, while the right channel is disparity-compensated with reference to the left one (and motion compensated, as well). This arrangement

allows to exploit the cross-correlation between the stereo-pairs, thus yielding a better performance over two separate coding chains. Furthermore, one extra module has been embedded to carry out an estimation of the stereo pair disparity, aiming at a scene segmentation, rather than a bit-rate optimization; the reason is clearly that here the buffer state operates the coder activity according to the distance estimated for the point originating the current pixel to the focus plane of the camera. In fact, the feedback loop connection allows to regulate the output data rate, according to the channel availability, by controlling the scaling of the DCT coefficients in such a way that precision is reduced for the image blocks lying on less relevant areas. As a result, variable precision bits assignment aims at reducing the signal energy in the background patterns, while retaining finer details on focused image portions.

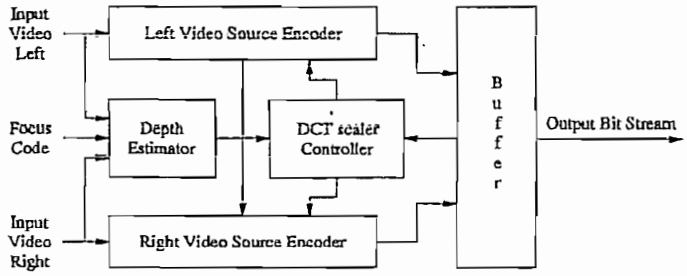


Fig. 2 - Principle of the coding scheme for 3DTV signals.

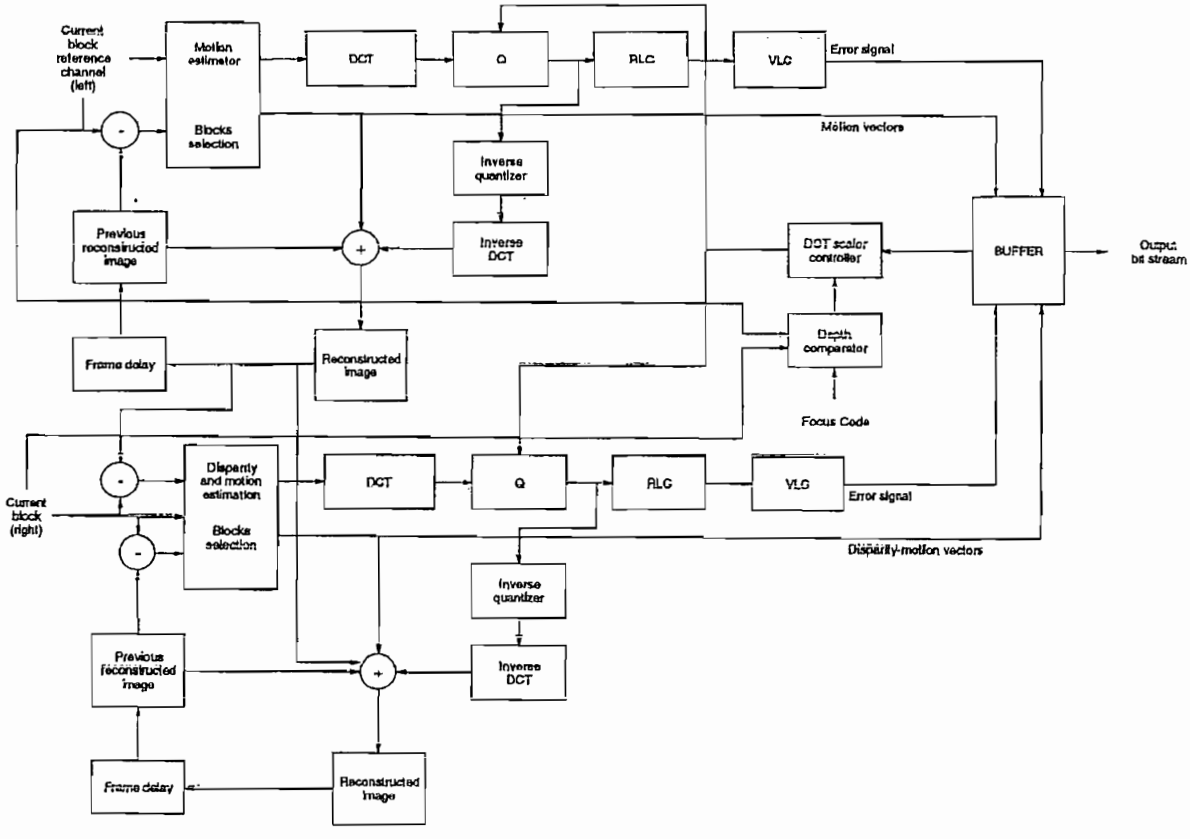


Fig. 3 - Coding scheme for 3DTV signals.

A somewhat more detailed scheme of the encoder is reported in fig. 3; it basically acts as an MPEG-like encoder[5], in that frames can be coded in modes I, P. A description is given in the following sections of the algorithms underlying the blocks responsible for the disparity evaluation and the buffer control mechanism.

3.2 DECODER

The structure of the decoder matches the encoder functions with a very little overhead to track image segmentation. This is a valuable feature of the system, as it involves additional processing power mostly on the coder, to support image segmentation.

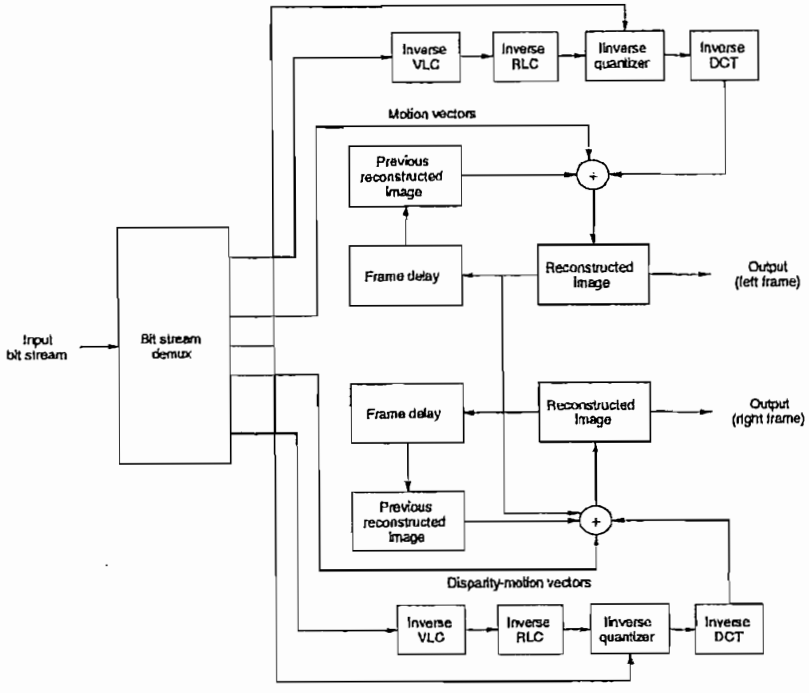


Fig. 4 - Decoding scheme for 3DTV signals.

3.3 DISPARITY EVALUATION

In this work the Quad-Tree algorithm has been specialized to the evaluation of the "sparse" disparity map of 3D-TV images, involving an estimation of a disparity vector for each block of 8x8 pixels of the original images. As well known, the concept of disparity arises from the two side-images of a stereo-pair being representations of the same scene captured from slightly different view-points, so that a parallax is generated for each real point projected onto the two image planes. The horizontal displacement between the two projected pixels on the two images is here referred to as the disparity. The association of a disparity value with each pixel in one image of the stereo-pair defines a disparity field, or map, which of course takes the same dimensions as the images themselves of the stereo-pair. A disparity map can be graphically represented as an artificial intensity image where the disparity is represented by the gray level. With this position, objects off the pick-up devices (located on the scene background) appear darker than nearer objects devices (located on the scene foreground), which will appear brighter instead[3,6].

As well known, a Quad-Tree approach to the motion/disparity estimation of stereo image pairs acts at different image resolution levels, according to a hierarchical model[3,6]. At each resolution level the algorithm processes a suitably decimated (sub-sampled) version of the original stereo-pairs in order to perform a local disparity estimation, also taking into account the estimation processed at the lower resolution level. An estimation of the best candidate as a correspondent

pixel in the left image to match a given pixel in the right image can be obtained by a block-matching process, which is a well established component of the standard coding schemes for digital video signals.

In this work the reliability of the disparity estimation has been further enhanced by applying a bi-directional consistency check constraint to a combination of luminance values and a set of image features (cornerness, edgeness, edges and its direction)[7,8].

For the sake of a verification of the effectiveness of our algorithm, several stereo-pairs have been selected as test vectors from a data-base available to the scientific community for results interchange. For demonstration purposes in fig. 5 a still stereo-pair from the stereo-sequence "Train" (courtesy of C.C.E.T.T. - F) is reported. Images resolution is 720 pels x 576 lines/pel, according to the European standard digital video scan format[9,10].

Fig. 6a represents the sparse disparity map produced by our algorithm. Accordingly, a segmentation of the original images into three different depth layers is possible by a suitable selection of three disparity ranges: figures 6b, 6c and 6d show the right views of the foreground, the region of interest and the background respectively of the original image reported in fig. 5b.

3.4 BUFFER CONTROL

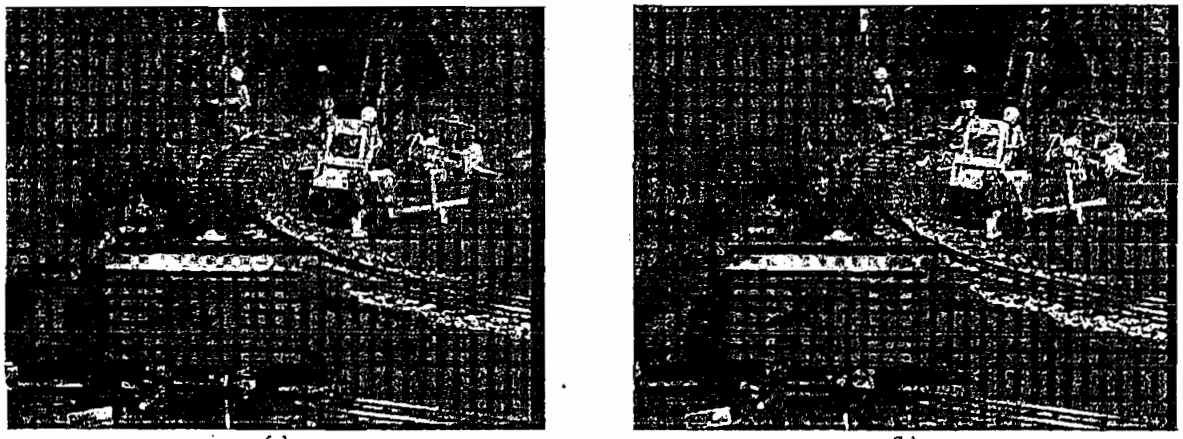
The transmission buffer is controlled by an algorithm which has been designed to extend the features performed by known computational-efficient 2D video codecs[11] to accommodate the new additional specifications set-up for our codec: (1) taking into account a double video channel, (2) performing a selective coding of the images. The basic structure of our algorithm is described in fig. 7, where a stereoscopic GOP (N stereo-frames) is supposed to be coded according to the following scheme: the first stereo-frame is coded by an intraframe mode, the following stereo-frames are compensated both for motion and disparity (cmp. fig. 2, 3).

Quantities $\underline{Q_I}$ and $\underline{Q_P}$ describe sets of the three quantization coefficients used to selectively code the stereo-frames of type I and P respectively; the amount of bits produced by the various types of frames (I, P, Left, Right) are labeled according to their positions within the GOP (B_{Il} B_{ir} B_{pl} B_{pr}). B_s is the specified amount (target) of bits to be produced by the whole stereoscopic GOP. Tolerated errors are indicated as target_err for the percentage of B_s and psnr_err for the signal-to-noise ratio of the reconstructed images versus the original ones. The computing complexity of our algorithm is characterized by the evaluation of $\underline{Q_I}$ and $\underline{Q_P}$, that are efficiently found according to a binary search. The first two stereo-frames are coded jointly, to allow for an accurate estimation of the amount of bits produced by I and P-type frames of the whole GOP. Bit estimation is dynamically adjusted at the time every following stereo-frame is singularly coded.

The buffer state corresponding to the bit production of a single stereo-GOP is described in fig. 8 for various bit-rates. An average error about 1% has been measured for the residual occupancy of the buffer at the end of the stereo-GOP.

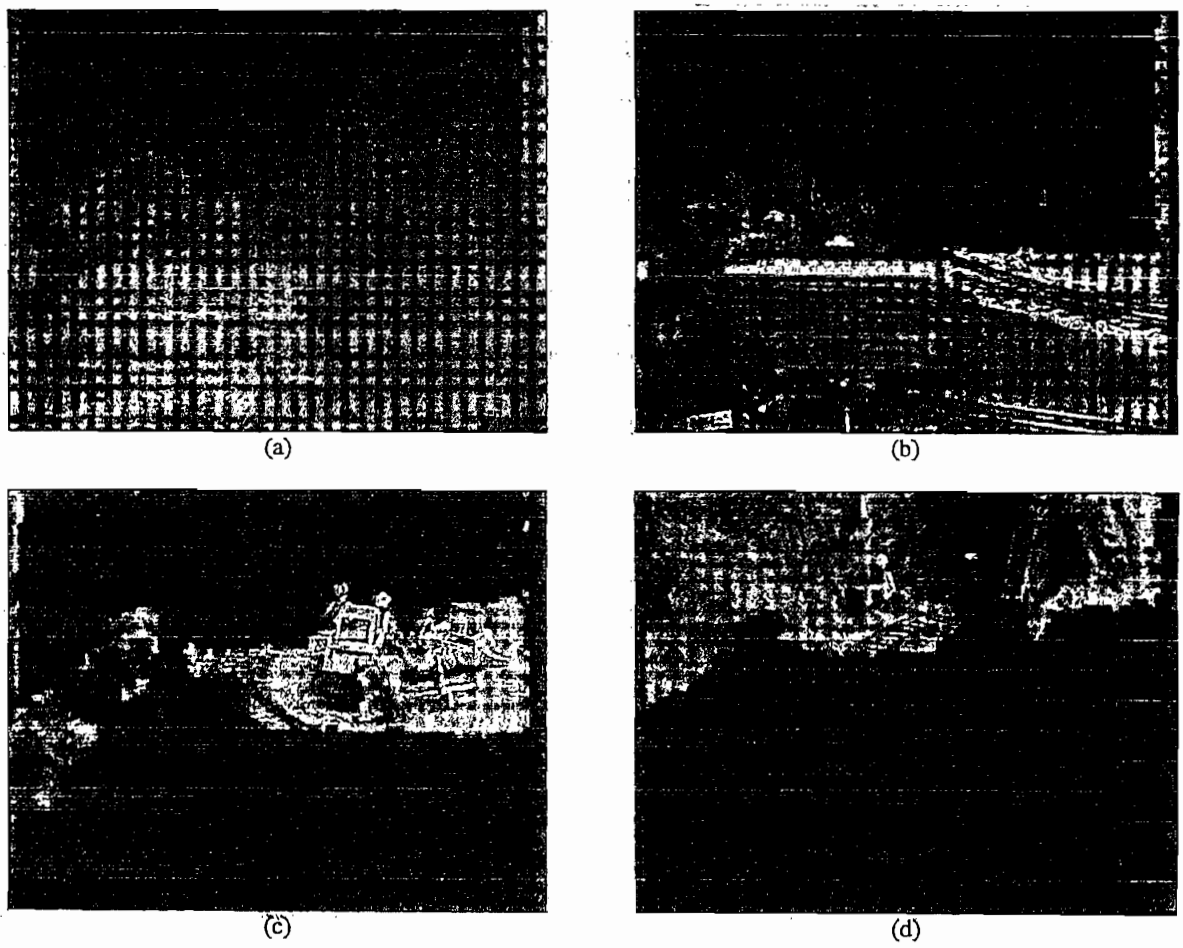
4. PERFORMANCE EVALUATION

To compare the behaviour of the buffer control in the two cases of uniform Q and selective Q's, the bit-rate 10.5 Mbits/sec has been selected; namely in fig. 9a the Q quantity needed for a uniform quality (quantization) within each single stereo-frame is plotted versus the frames sequence within a stereo-GOP; the plots are overlapped of the three Q values associated with the three depth layers as in fig. 6b (level 0: foreground), 6c (level 1: region of interest), 6d (level 2: background). For ease of graphical representation, the plots refer to the right views; left views exhibit the same behaviour. As expected from a theoretical standpoint, experimental results confirm that the region of interest is coded with smaller Q values; this outcome in turn accounts for higher PSNR values, or an image superior quality, as can be seen in fig. 9b. Of course, in order to keep constant the bit-rate in the two cases (uniform Q, selective Q's) such gain in quality is expected to be compensated by a controlled quality degradation in the background/foreground portions; also this expectation is confirmed by the simulation results plotted in fig. 9.



(a) (b)

Fig. 5 - Original stereo-pair "Train"; (a): left image, (b): right image.



(a) (b)

(c) (d)

Fig. 6 - (a): Disparity Map for the stereo-pair of fig. 5;
(b): foreground estimation; (c): central depth slice estimation; (d): background estimation.

```
init Q_i
while (abs [B_I1 + B_Ir + (B_pl + B_pr) (N-1) - B_s] > target_err B_s) do
begin
        evaluate Q_i
        init Q_p
        while (abs [PSNR_I1 - PSNR_pl ] > psnr_err) do
        begin
                evaluate Q_p
        enddo
enddo
I_code (frame I)
P_code (frame P_1)

for k=2 to (N-1) do
begin
        update B_s
        init Q_p
        while (abs [(B_pl + B_pr) (N-k) - B_s] > target_err B_s) do
        begin
                evaluate Q_p
        enddo
        P_code (frame P_k)
enddo
```

Fig. 7 - Buffer control algorithm.

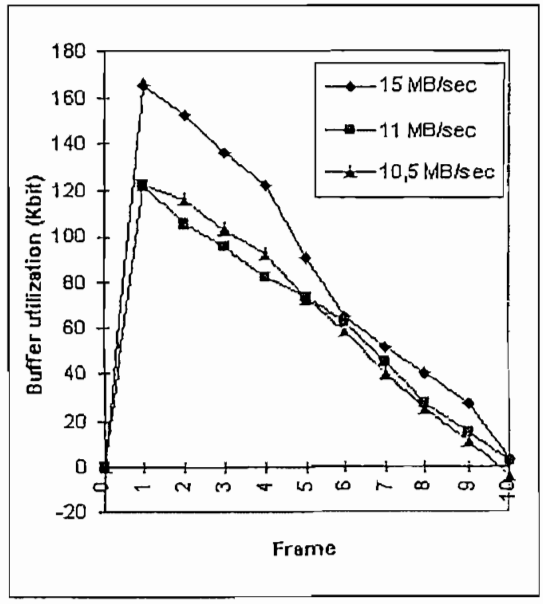
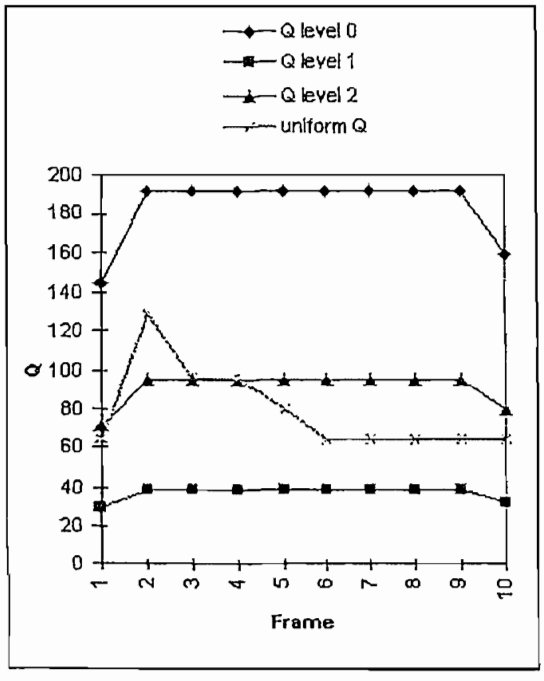
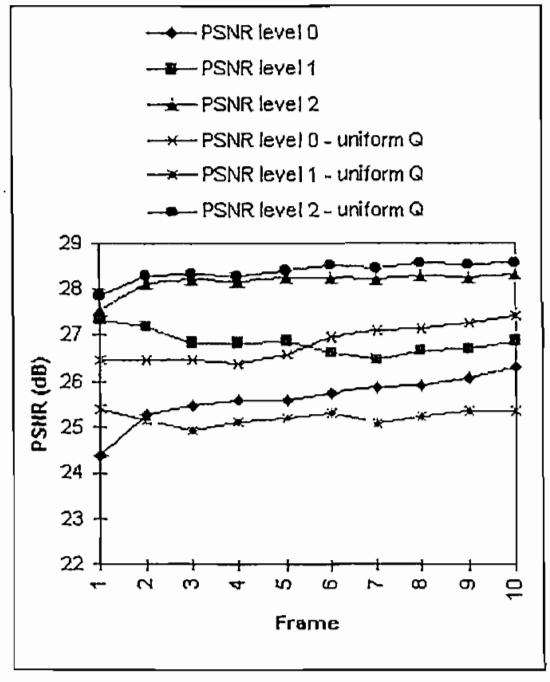


Fig. 8 - Buffer utilization in a GOP (Group-Of-Pictures).

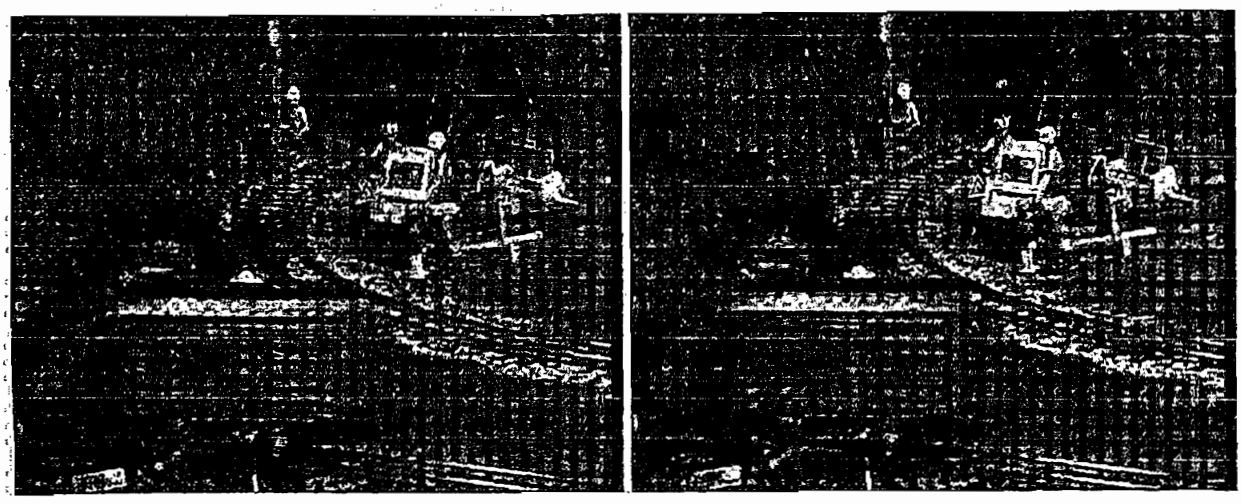


(a)



(b)

Fig. 9 - Scaling coefficients (a) and PSNR values (b) for uniform coding and selective coding. (10.5 Mbit/sec)



(a) (b)

Fig. 10 - Right image reconstructed after: (a) uniform coding, (b) selective coding.

Results show improvements over conventional coding schemes in the capability to obtain for the visually relevant image contents a better quality for a given transmission bit rate.

Such numerical improvements are confirmed by a visual inspection of the right images reconstructed by the two methods, as can be seen in fig. 10; a better image quality can be subjectively appreciated in the region of interest, while higher quality degradation are confined in the background and especially in the foreground; it should be noticed that the unbalance between the visual quality of the foreground/background portions reflects our a-priori choice of their respective Q values, as shown in fig. 9a; this result is also consistent with objective evaluations of the signal to noise ratio parameter reported in fig. 9b.

It is worth noticing that the coding method described can be applied to a true stereoscopic video signal, as well as to a single compatible view thereof, which results in a downwards application of 3DTV to compatible TV[12,13].

5. FUTURE DEVELOPMENTS

Due to the modular architecture of the encoder, the performances of the overall system may be increased by focusing on local improvements of the single components. Among these, one of the most fundamental in our codec is the disparity evaluation module, clearly because it is responsible for the critical task of the image segmentation. Therefore, future research is expected to enhance the accuracy of the disparity estimation by including a stereo-motion consistency analysis.

Also, improvements on the buffer control performances can be expected, as regards both the bit estimation error and the computing complexity, by suitably combining our basic binary search algorithm with a Q prediction law.

An alternate approach for selective coding we are going to investigate is based on low-pass filtering the image areas which are less relevant. This leads to a different image quality alteration, as smoothness is resulted rather than block effects. Such an approach is based on a pre-filter bank acting on the input image pair, whose original blocks may be pre-processed according to their visual relevance in the sense proposed here.

The basic *principle of the* separation into depth slices of different visual relevance may support the definition of priority levels in video packet transport by ATM networks. Image coders for ATM networks take into account the non-zero probability of traffic congestion, resulting in a quantity of video packets being discarded. In order to retain image quality from a visual perception standpoint, the proposed depth-based visual criterion may assist in a selective data reduction, thus allowing for a controlled, or graceful, image quality degradation.