

**Instituto Tecnológico y de Estudios Superiores de Monterrey**

**Campus Monterrey**

**División de Electrónica, Computación, Información y Comunicaciones**

**Programa de Graduados**



**Conversión Autónoma de Video Monoscópico a Estereoscópico  
2D-3D**

**TESIS**

Presentada como requisito parcial para obtener el grado académico de

**Maestro en Ciencias en Ingeniería Electrónica**

**Especialidad en Sistemas Electrónicos**

**Ing. Carlos Alejandro Covarrubias Valadez**

Diciembre de 2001



© Carlos Alejandro Covarrubias Valadez, 2001



# Instituto Tecnológico y de Estudios Superiores de Monterrey

**Campus Monterrey**

**División de Electrónica, Computación, Información y Comunicaciones**

**Programa de Graduados**

Los miembros del comité de tesis recomendamos que la presente tesis de Carlos Alejandro Covarrubias Valadez sea aceptada como requisito parcial para obtener el grado académico de Maestro en Ciencias en:

**Ingeniería Electrónica**

**Especialidad en Sistemas Electrónicos**

**Comité de tesis:**

---

Ramón M. Rodríguez Dagnino, Ph.D.

Asesor de la tesis

---

Ing. Ricardo Guzmán Navarro

Sinodal

---

Gregorio Rivera Medellín, M.A., M.Ed.

Sinodal

---

David Garza Salazar, Ph.D.

Director de los Programas de Graduados en  
Electrónica, Computación, Información y  
Comunicaciones

Diciembre de 2001



A Dios, a mis padres y a mis abuelos.



## Agradecimientos

**Mamá**, gracias por todo tu cariño, tu paciencia y por las enseñanzas que me has dado.

**Papá**, gracias por tus consejos siempre certeros y por ser mi mejor amigo.

A **Miris** por su cariño y apoyo durante todo este tiempo, así como por las largas charlas terapéuticas.

A mis socios **Sergio Aguirre** y **Cesar Centeno** por todos los buenos momentos que hemos pasado con la 3DTV y por los que están por suceder.

A mis viejos amigos **Oziel González** y **Ricardo Gómez** por su sincera amistad y por evitar que perdiera el *ritmo* de la tesis.

A mi asesor **Dr. Ramón M. Rodríguez Dagnino** por su orientación y guía que hicieron posible concretar esta investigación.

A mi maestro y sinodal **Ing. Ricardo Guzmán Navarro** por haberme inculcado el gusto por cuestiones de televisión y por haberme enseñado a llegar más allá de los límites ingenieriles.

A mi sinodal **Gregorio Rivera Medellín M.A. M.Ed.** por su dedicación en la revisión de esta tesis y sus valiosas aportaciones.

A mis maestros del Departamento de Ingeniería Eléctrica.

A **Toñita** por sus consejos, su cariño y su dedicación durante más de 20 años, gracias por ser una segunda mamá.

A mi abuelito **Rodolfo** y a mis tíos **Pepe** y **Mireya** por su cariño y apoyo en todo momento.

A mi buen amigo **Julio Esparza** por su interés en esta investigación.

A mi familia boliviana a quienes debo el 50% de mis raíces.

CARLOS ALEJANDRO COVARRUBIAS VALADEZ

*Instituto Tecnológico y de Estudios Superiores de Monterrey*  
*Diciembre 2001*



# Conversión Autónoma de Video Monoscópico a Estereoscópico

## 2D–3D

Carlos Alejandro Covarrubias Valadez, M.C.  
Instituto Tecnológico y de Estudios Superiores de Monterrey, 2001

Asesor de la tesis: Ramón M. Rodríguez Dagnino, Ph.D.

En los últimos años las imágenes estereoscópicas (3D) han despertado la atención y el interés del mundo en un variado rango de aplicaciones, dado que una imagen 3D brinda mayor información que su contraparte en 2D.

Actualmente se pueden generar imágenes 3D mediante dos métodos: gráficas computacionales y grabación de escenarios reales; a pesar de que el primer camino se ha visto muy beneficiado por el auge tecnológico en el último lustro, aún resulta mayor la calidad de una grabación en escenarios reales.

El problema de la adquisición de tomas reales reside en la necesidad de utilizar cámaras estereoscópicas cuyo precio es muy elevado, o en emplear dos cámaras convencionales, hecho que duplica el costo de la producción. Así mismo se requiere de técnicos con un alto grado de pericia para poder manejar de forma adecuada la posición, orientación, distancias, alineamientos y *zooms* de ambas cámaras, pues la calidad del efecto estereoscópico depende totalmente de la manipulación de las variables anteriores.

El propósito de esta tesis es desarrollar un convertidor 2D-3D para eliminar las barreras mencionadas obteniendo una imagen 3D siguiendo la técnica de grabación monocular actual.



# Índice General

Índice de Tablas	v
Índice de Figuras	vii
<b>Capítulo 1 Introducción</b>	<b>1</b>
1.1 Antecedentes	1
1.2 Alcance de la tesis y motivación	2
1.3 Organización de la tesis	2
<b>Capítulo 2 Conceptos básicos</b>	<b>5</b>
2.1 Breve perspectiva histórica de la estereoscopia	5
2.2 Teoría básica de estereoscopia	6
2.3 Variables clave para la percepción humana de la estereoscopia	7
2.3.1 Tipos de paralaje	8
2.4 Parámetros críticos en un sistema estereoscópico	8
2.4.1 Máximo paralaje binocular aceptable para un observador	9
2.4.2 Límites de <i>cercanía</i> y <i>lejanía</i> en la región estereoscópica útil	11
2.4.3 Resolución de profundidad	12
2.4.4 Distorsión de despliegue	12
<b>Capítulo 3 Metodología propuesta</b>	<b>13</b>
3.1 Fundamento básico	13
3.1.1 Experimento canónico de Ross	13
3.2 Etapas de la metodología	15
3.3 Hardware y software utilizado para la investigación	15
<b>Capítulo 4 Preprocesamiento</b>	<b>19</b>
4.1 Formato UYVY	19
4.1.1 Dimensiones de la secuencia de video	20
4.2 Extracción e interpolación de luminancia	21
<b>Capítulo 5 Detección de Movimiento</b>	<b>23</b>
5.1 Métodos basados en la ecuación de flujo óptico	23
5.2 Métodos recursivos de <i>pels</i>	24
5.3 Métodos basados en bloques	24
5.3.1 Algoritmos	26
5.3.1.1 Full search (FS)	26

5.3.1.2	Three step search (3SS)	26
5.3.1.3	Four step search (4SS)	27
5.3.1.4	New three step search (N3SS)	28
5.3.1.5	Block-based gradient descent (BBGDS)	31
5.3.1.6	Hybrid search (HS)	31
5.3.1.7	Coherent search (CS)	33
5.3.1.8	Partial distortion full search (PDFS)	34
5.3.1.9	Unrestricted center-biased diamond search (UCBDS)	36
5.3.2	Criteria para medición de distorsión entre bloques	38
5.3.2.1	MSE	38
5.3.2.2	MAE	38
5.3.3	Desempeño de la detección de movimiento	38
5.3.4	Desempeño para conversión estereoscópica	39
5.3.4.1	Evaluaciones diseñadas	40
<b>Capítulo 6 Resultados de los análisis de movimiento</b>		<b>43</b>
6.1	Resultados de las evaluaciones propuestas	43
6.1.1	Análisis visual de desempeño global	43
6.1.2	Análisis analítico de desempeño global	44
6.1.3	Análisis analítico de desempeño relativo	44
6.2	Clasificación final de desempeños de detección	48
<b>Capítulo 7 Generación de la imagen estéreo-artificial</b>		<b>51</b>
7.1	Relación entre movimiento y la conversión estéreo-artificial	51
7.1.1	Cámara fija	51
7.1.1.1	Dirección del movimiento detectado	51
7.1.2	Cámara en movimiento	53
7.2	Determinación del retraso entre cuadros	56
7.3	Continuidad de los parámetros de conversión	57
7.3.1	Histéresis en la dirección del retraso	57
7.3.2	Histéresis en la magnitud del retraso	57
7.4	Caso especial: Escenas sin movimiento	60
<b>Capítulo 8 Métodos de despliegue estereoscópico</b>		<b>61</b>
8.1	Clasificación de métodos para despliegue estereoscópico	61
8.2	Técnicas para despliegue estereoscópico	61
8.2.1	Holográfica	62
8.2.2	Volumétrica	63
8.2.3	Basada en paralaje	64
8.2.4	Lenticular	65
8.2.5	De separación visual	65
8.2.5.1	Separación pasiva	66
8.2.5.2	Separación activa	66
8.2.5.3	Separación activo-pasiva	66
8.2.5.4	Separación por colores	68
8.3	Implementación del método de separación por colores	68

<b>Capítulo 9</b>	<b>Análisis de la conversión estéreo-artificial</b>	<b>73</b>
9.1	Análisis de los parámetros críticos en estereoscopia . . . . .	73
9.1.1	Máximo paralaje obtenido . . . . .	73
9.1.2	Límites de cercanía y lejanía . . . . .	73
9.1.3	Resolución de profundidad y distorsión de despliegue . . . . .	74
9.2	Estéreo-artificial <i>vs</i> estéreo-natural . . . . .	74
9.3	Secuencia de laparoscopia . . . . .	75
<b>Capítulo 10</b>	<b>Conclusiones</b>	<b>77</b>
10.1	Extensión de resultados y futuras investigaciones . . . . .	77
<b>Apéndice A</b>	<b>Secuencias de video</b>	<b>79</b>
A.1	Secuencias estereoscópicas . . . . .	79
A.2	Secuencias monoscópicas . . . . .	79
A.2.1	Secuencias monoscópicas sobre medicina . . . . .	81
A.2.2	Otras secuencias monoscópicas utilizadas . . . . .	81
<b>Apéndice B</b>	<b>Glosario</b>	<b>83</b>
	<b>Bibliografía</b>	<b>85</b>
	<b>Vita</b>	<b>87</b>



## Índice de Tablas

3.1	Características de la tarjeta SunVideo Plus. . . . .	17
4.1	Periodos de muestreo UYVY. . . . .	20
4.2	Rango de valores de las componentes YUV. . . . .	20
5.1	Valores de $s_p$ y $t_p$ para cada distorsión $p$ . . . . .	36
6.1	Tiempos de ejecución. . . . .	43
6.2	Sensibilidades con mejores desempeños visuales. . . . .	44
6.3	Sensibilidades con mejores desempeños analíticos. . . . .	49
6.4	Porcentajes de movimiento detectado. . . . .	49
6.5	Clasificación de acuerdo a desempeño. . . . .	49
7.1	Tipos de paralajes obtenidos para secuencias con cámara fija. . . . .	54
7.2	Tipos de paralajes obtenidos para secuencias con cámara en movimiento. . . . .	54
7.3	Reglas para la creación de paralajes en secuencias sin movimiento . . . . .	60
A.1	Características las cámaras utilizadas. . . . .	79
A.2	Secuencias estereoscópicas. . . . .	80
A.3	Secuencias monoscópicas. . . . .	80
A.4	Secuencias laparoscopia. . . . .	81
A.5	Otras secuencias monoscópicas de prueba. . . . .	81



## Índice de Figuras

2.1	Estereoscopio . . . . .	5
2.2	Visión estereoscópica. . . . .	6
2.3	Suma simple de dos imágenes. . . . .	7
2.4	Imagen monoscópica en perspectiva ( $2D+$ ). . . . .	7
2.5	Tipos de paralaje. . . . .	9
2.6	Relación entre profundidad y paralaje. . . . .	10
2.7	Ángulo del área de Panum. . . . .	10
2.8	Región estereoscópica útil. . . . .	11
3.1	Proceso visual para objetos en movimiento. . . . .	14
3.2	Diagrama de flujo del algoritmo. . . . .	16
4.1	Formato UYVY. . . . .	20
4.2	Extracción e interpolación de luminancia. . . . .	21
5.1	Full search. . . . .	26
5.2	Patrón de búsqueda de <i>full search</i> . . . . .	27
5.3	Three step search. . . . .	28
5.4	Four step search. . . . .	29
5.5	New three step search. . . . .	30
5.6	Block based gradient search. . . . .	31
5.7	Orden utilizado para la búsqueda por HS. . . . .	32
5.8	Relación entre $D(0,0)$ y $M$ . . . . .	32
5.9	Bloques sin patrones distintivos (planos). . . . .	33
5.10	Bloques con patrones repetitivos. . . . .	34
5.11	Bloques utilizados para el cálculo de la predicción coherente. . . . .	34
5.12	Distorsiones parciales. . . . .	35
5.13	Ubicación de cada distorsión parcial. . . . .	36
5.14	Unrestricted center-biased diamond search. . . . .	37
5.15	Reglas para contabilizar porcentajes de detección de movimiento. . . . .	41
6.1	Desempeño visual de <i>full search</i> a diferentes sensibilidades (1). . . . .	45
6.2	Desempeño visual de <i>full search</i> a diferentes sensibilidades (2). . . . .	46
6.3	Desempeño visual de <i>full search</i> a diferentes sensibilidades (3). . . . .	47
6.4	Histogramas a diferentes sensibilidades para FS. . . . .	48
6.5	Porcentajes de movimiento detectado. . . . .	50

7.1	Movimiento de izquierda a derecha con paralaje negativo. . . . .	52
7.2	Movimiento de derecha a izquierda con paralaje negativo. . . . .	53
7.3	Movimiento de izquierda a derecha en el fondo con paralaje positivo. . . . .	55
7.4	Espacios visuales. . . . .	56
7.5	Histéresis en dirección. . . . .	58
7.6	Histéresis en magnitud. . . . .	59
8.1	Holografía. . . . .	62
8.2	Espejo varifocal. . . . .	63
8.3	Barrera de paralaje. . . . .	64
8.4	Sistema lenticular. . . . .	65
8.5	Polarización. . . . .	66
8.6	Separación pasiva. . . . .	67
8.7	Separación activa. . . . .	67
8.8	Separación activo-pasiva. . . . .	68
8.9	Proceso de “tintado”. . . . .	71

## Capítulo 1

# Introducción

En los últimos años las imágenes estereoscópicas (3D) han despertado la atención y el interés del mundo en un variado rango de aplicaciones como son videojuegos, sistemas de multimedia, televisión estereoscópica (3dtv), realidad virtual, etc. Esto se debe principalmente a que una imagen 3D brinda mayor información que su contraparte en 2D, así como una sensación más placentera a la vista.

Actualmente se pueden generar imágenes 3D mediante dos métodos: gráficas computacionales y grabación de escenarios reales; a pesar de que el primer camino se ha visto muy beneficiado por el auge tecnológico en el último lustro, aún resulta mayor la calidad de una grabación en escenarios reales.

El problema de la adquisición de tomas reales reside en la necesidad de utilizar cámaras estereoscópicas cuyo precio es muy elevado, o en emplear dos cámaras convencionales, hecho que duplica el costo de la producción. Así mismo se requiere de técnicos con un alto grado de pericia para poder manejar de forma adecuada la posición, orientación, distancias, alineamientos y *zooms* de ambas cámaras pues la calidad del efecto estereoscópico depende totalmente de la manipulación de las variables anteriores [6].

Al desarrollar un convertidor 2D-3D, se eliminan las barreras mencionadas pues se obtendría una imagen 3D siguiendo la técnica de grabación monocular actual; así mismo se abre la posibilidad de utilizar la extensa videoteca creada hasta el momento por numerosas casas productoras y televisoras a nivel internacional.

## 1.1 Antecedentes

El área de investigación de convertidores 2D-3D no ha sido considerada formalmente en Estados Unidos, ante la existencia de muchos escépticos que proponen que es imposible reconstruir un segundo punto de vista a partir de una imagen monocular. En Japón se tiene 3 años de investigación tratando de demostrar lo contrario, fundamentados en el gran desarrollo tecnológico existente.

La muestra más notable de este esfuerzo fue cristalizada por 4 ingenieros de la compañía Sanyo<sup>1</sup> quienes propusieron un sistema llamado *3DR* basado en técnicas de visión computarizada.

Dicho sistema posee varias limitantes, ya que este algoritmo *3DR* [14] permite la reconstrucción de imágenes 3D, sin embargo su calidad no es aceptable para un espectador común ya que sacrifica tiempo de procesamiento con tal de obtener un producto de “bajo costo” y en tiempo real, por lo que se truncó gran parte del trabajo de cálculo para la obtención del segundo punto de vista.

En México el área de investigación y desarrollo sobre estereoscopía ha permanecido estéril por muchos años, y hasta la fecha no se posee conocimiento de alguna tesis, artículo o similar generado sobre este campo.

---

<sup>1</sup>Sanyo es una marca registrada de Sanyo Electric Co.

## 1.2 Alcance de la tesis y motivación

Esta investigación es la continuación del desarrollo de un sistema de televisión estereoscópica, realizado durante la carrera profesional del autor<sup>2</sup>. Dentro de las principales motivaciones de esta tesis se encuentra, como ya se mencionó, simplificar de sobremanera los obstáculos encontrados durante la grabación de material estereoscópico utilizando ya sea dos cámaras profesionales o una cámara especial con dos lentes, al permitir continuar empleando la producción monocular que existe actualmente; así mismo se pretende ampliar enormemente la videoteca de material estereoscópico pudiendo convertir cualquier video monocular existente que cuente con características de movimientos de complejidad limitada y en dirección horizontal.

El alcance de esta investigación contempla el desarrollo e investigación de un convertidor 2D-3D utilizando los principios de detección de movimiento y conversión estéreo-artificial a partir de imágenes con retraso.

El formato de video utilizado se limita a NTSC, el cuál es entrelazado con una frecuencia de despliegue de 60 campos por segundo.

Este trabajo de tesis no tiene contemplado como punto prioritario el generar un algoritmo convertidor en tiempo real, por lo que queda abierta la posibilidad de en un futuro optimizar e implementar en hardware dedicado el convertidor.

Esta técnica de conversión se puede emplear en un amplio rango de aplicaciones que van desde sistemas de entretenimiento hasta sistemas de inspección visual en la industria, sin embargo para esta investigación solo se estudiarán sus resultados sobre secuencias de video en medicina, específicamente sobre una laparoscopia láser.

## 1.3 Organización de la tesis

En el capítulo 2 se plantean conceptos básicos de estereoscopia así como una breve perspectiva de su evolución a lo largo de los dos últimos siglos.

En el capítulo 3 se propone la metodología a seguir para la conversión estéreo-artificial, así como su fundamento básico, información que será analizada y explorada en los subsecuentes capítulos.

En el capítulo 4 se analizan las características básicas del formato de video digital utilizado durante la investigación y se explica el preprocesamiento que es necesario aplicarle a fin de adecuar la señal de video para su conversión.

Durante el capítulo 5 se describen diferentes métodos para la detección de movimiento, aspecto que constituye una de las partes fundamentales del convertidor, así mismo se analiza su desempeño orientado a la conversión estereoscópica.

En el capítulo 6 se muestran los resultados de las pruebas propuestas en el capítulo anterior y se determina cuál método de detección constituye la mejor herramienta para el proceso de conversión.

El capítulo 7 constituye la parte medular de la investigación, pues en él se explica a detalle la relación entre los parámetros de movimiento y la estereoscopia, así como otras consideraciones sobre los diferentes tipos de movimientos en secuencias monoculares.

Una vez creada la secuencia de video estereoscópica es necesario desplegarla para observar su desempeño, para ello en el capítulo 8 se describen diferentes técnicas para despliegue estereoscópico y se muestra a detalle la implementación del método de separación por colores (azul y rojo).

---

<sup>2</sup>En la actualidad las patentes de este sistema pertenecen a la empresa Innouva Technologies de México.

A lo largo del capítulo 9 se analizan las secuencias estéreo-artificiales utilizando los conceptos básicos establecidos en los primeros capítulos y comparándolas contra las mismas secuencias pero grabadas originalmente en formato estereoscópico. Así mismo se comentan los resultados de la conversión de videos de laparoscopías.

En el capítulo 10 se concluyen los resultados más destacados de este trabajo y se ofrece una perspectiva del futuro de la estereoscopía, mencionado las posibles investigaciones de extensión de resultados que pueden derivarse de ésta.

Por último se incluyen dos apéndices para complementar esta tesis, mencionando aspectos básicos sobre las secuencias de video utilizadas, así como un glosario con la terminología utilizada más importante sobre estereoscopía.



## Capítulo 2

### Conceptos básicos

#### 2.1 Breve perspectiva histórica de la estereoscopia

El concepto de estereoscopia se remonta principios del siglo XVIII, ya que poco después de la aparición de la fotografía, la cual comenzó a ser popular por su habilidad para capturar y reproducir imágenes de todo tipo, se inventó la estereoscopia tratando de obtener mayor realismo en dichas imágenes.

Los estereoscopios fueron los primeros dispositivos que permitieron ver fotografías que aparentaban presentar la profundidad de los objetos, estos instrumentos fueron desarrollados por Wheatstone [12] en 1838, convirtiéndose en un concepto muy atractivo para la mayoría de las personas.

A través de su invento Wheatstone explicó que existe un único sentido de profundidad llamado *estereopsis* el cual es producido por la disparidad retinal.

Investigadores contemporáneos se preguntaban porque al utilizar dicho dispositivo no se percibía una doble imagen; esto fue solucionado por Wheatstone al explicar la fusión mental de las dos imágenes retinales en una sola con estereopsis.

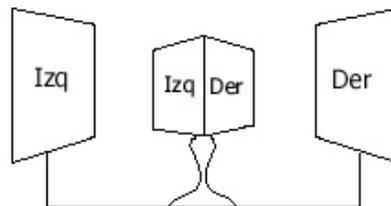


Figura 2.1: Estereoscopio: La nariz del observador se sitúa donde se unen los espejos izquierdo y derecho, por lo que cada ojo recibe una imagen diferente.

Las investigaciones sobre el tema continuaron hasta que en 1895 se proyectó la primera película estereoscópica en una función privada en París, pero no fue hasta la década de 1950 que dichas proyecciones fueron comercializadas ofreciendo lo que llamaron una *nueva dimensión* en el cine. Sin embargo la pobre calidad del efecto estereoscópico, que por momentos resultaba exagerado siendo más irreal que sorprendente, y los extraños lentes azul-rojo que producían mareos y dolores de cabeza, contribuyeron al fracaso de dicha tecnología.

En la década de 1980 comenzó el auge de los videojuegos, sin embargo no fue hasta después de 1990 que la tecnología se desarrolló lo suficiente creando juegos de alto realismo con gran interactividad

con el usuario, sonido sofisticado y gráficas complejas en  $2D+$ .

Fue entonces cuando Sega<sup>1</sup> diseñó unos lentes especiales de cristal líquido para ser utilizados junto con su consola y permitir el desarrollo de juegos en 3D. Sin embargo esta tecnología fracasó ya que a pesar de mostrar un buen efecto estereoscópico, se presentaba un parpadeo en las imágenes debido a características de operación de la televisión, lo cual resultaba un poco molesto y el público no adquirió el producto como se pensaba, por lo que se retiró del mercado.

Durante la misma década de los 90's se presentó un verdadero auge por el desarrollo de sistemas estereoscópicos tanto para televisión como para computadora y otros tantos dispositivos independientes. Los parques de diversiones de Estados Unidos de América combinaron varios de los desarrollos realizados junto con su experiencia en grandes producciones, creando espectáculos filmados en 3D, los cuales a pesar de no haber llegado nunca a *Hollywood* constituyen uno de los mejores atractivos de dichos parques hasta la fecha.

Así mismo actualmente existen varias compañías dedicadas a la generación de juegos en 3D para computadora, limitándose principalmente a títulos de acción y aventura.

## 2.2 Teoría básica de estereoscopía

Una imagen estereoscópica es aquella que presenta a cada ojo del observador un punto de vista con diferente perspectiva, simulando la manera en que una persona ve el mundo real.

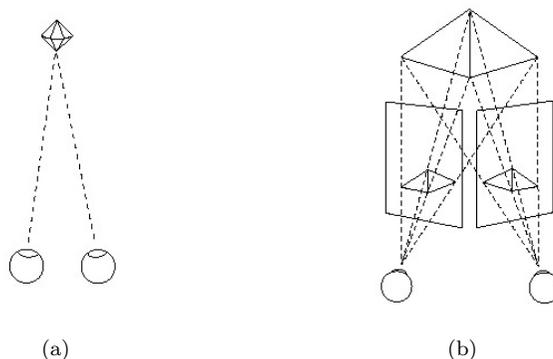


Figura 2.2: Visión estereoscópica.

A partir de estas dos imágenes (figura 2.2(b)) los ojos y el cerebro humano sintetizan una imagen con profundidad estereoscópica, es decir, el observador percibe una sola imagen y no una doble, ya que ambas perspectivas son fusionadas en una sola, la cual contiene mayor información que la simple suma de sus componentes (figura 2.3).

Actualmente existe una confusión lingüística entre los términos *estereoscópico (3D)* y *tridimensional*, esto se debe a que la industria de la gráficas computacionales utilizó durante mucho tiempo, el vocablo tridimensional para referirse a imágenes de aspecto “real” que pueden ser apreciadas con un solo ojo, mientras que la mayoría de la gente utiliza dicha terminología para referirse a lo que se conoce técnicamente como estereoscopía.

<sup>1</sup>Sega es una marca registrada de Sega Corp.

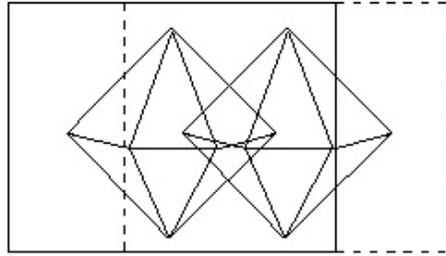


Figura 2.3: Suma simple de dos imágenes.

Para evitar dicha confusión se ha acordado designar a las imágenes monoscópicas en perspectiva como  $2D+$  o *imágenes de dos dimensiones y medio*, y a las imágenes estereoscópicas como  $3D$  o simplemente *estereoscópicas* tratando de evitar la utilización de la palabra *tridimensional*.

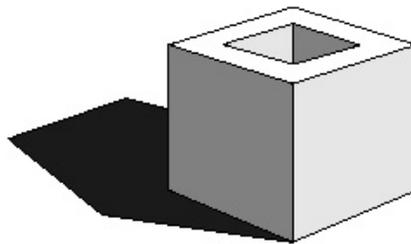


Figura 2.4: Imagen monoscópica en perspectiva ( $2D+$ ).

## 2.3 Variables clave para la percepción humana de la estereoscopia

Existen diversos aspectos que son esenciales para la percepción de profundidad por los seres humanos. De acuerdo con T. Okoshi [20] un observador puede utilizar hasta diez variables para captar imágenes con profundidad, las cuales se pueden agrupar de la siguiente manera:

- Variables fisiológicas:
  - Paralaje binocular.
  - Paralaje monocular de movimiento.
  - Enfoque.
  - Convergencia ocular.
- Variables psicológicas:

- Oclusión o interposición.
- Sombras e iluminación.
- Perspectiva lineal.
- Gradiente de textura.
- Tamaños relativos de objetos.
- Perspectiva aérea.

Las variables psicológicas pueden ser inducidas en una imagen bidimensional (p.ej. una fotografía) mediante la experiencia e imaginación del observador.

Sin embargo las variables de carácter fisiológico deben existir como tales para ser percibidas, siendo el paralaje binocular el aspecto más importante, el cual no se puede presentar si los ojos del observador no están correctamente enfocados y convergidos en un objeto [7].

### 2.3.1 Tipos de paralaje

Los valores de paralaje son la distancia entre los puntos homólogos entre las perspectivas izquierda y derecha, y constituyen la parte análoga a la disparidad retinal. Es decir el paralaje y la disparidad son entidades similares con una correspondencia proporcional, la única diferencia radica en que el paralaje se mide en la pantalla de despliegue mientras que la disparidad en la retina.

Al observar una imagen estereoscópica se pueden encontrar 4 tipos básicos de paralaje:

- Nulo: Se presenta cuando los puntos homólogos ocupan la misma posición en ambas imágenes. Cuando los ojos del observador ven imágenes con paralaje nulo, convergen sobre el plano de la pantalla (figura 2.5(a)).
- Positivo: Se presenta cuando el valor de paralaje es mayor a cero pero menor que la distancia interocular. Al observar una imagen estereoscópica bajo estas condiciones la persona percibe los objetos con una profundidad más allá del plano de la pantalla por lo que se dice que dichos objetos se encuentran en el *espacio del monitor* (figura 2.5(b)).
- Divergente: Es una variante de la condición de paralaje positivo y sucede cuando el valor de paralaje excede la distancia interocular. No se presenta bajo circunstancias normales en el mundo real por lo que no tiene un sentido lógico en imágenes estereoscópicas, ya que lo único que produce es un esfuerzo muscular para fusionar dichas imágenes, generando molestias (figura 2.5(c)).
- Negativo: Se presenta cuando los ojos del observador deben *cruzarse*. Bajo esta condición el observador percibe que los objetos se encuentran entre el plano del monitor y él, por lo que se dice que dichos objetos se encuentran en el *espacio del observador* (figura 2.5(d)).

En la figura 2.6 se puede observar la relación entre la profundidad de los objetos y los tipos de paralaje.

## 2.4 Parámetros críticos en un sistema estereoscópico

Los siguientes conceptos son críticos en el desarrollo de un sistema estereoscópico:

- Máximo paralaje binocular aceptable por un observador.

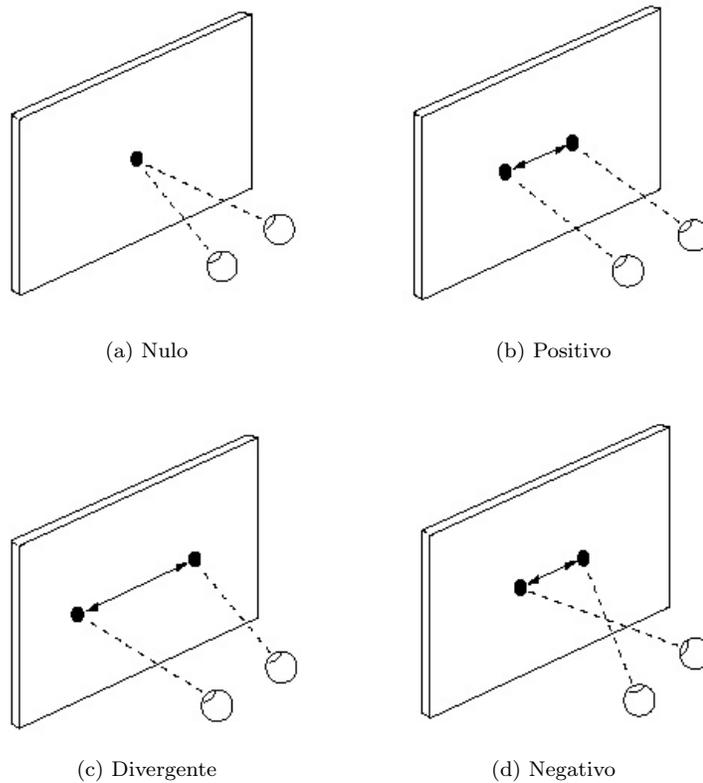


Figura 2.5: Tipos de paralaje.

- Límites de *cercanía* y *lejanía* de lo que se denomina región estereoscópica útil.
- Resolución de profundidad o mínimo cambio detectable en profundidad.
- Distorsión de despliegue producida por amplificación de profundidad.

### 2.4.1 Máximo paralaje binocular aceptable para un observador

Los valores de paralaje binocular de las imágenes deben de encontrarse dentro de un rango específico para ser aceptados por un observador.

Si el paralaje comienza a exceder el máximo permisible, entonces el observador sentirá molestias oculares al tratar de fusionar los dos puntos de vista. Si dicho paralaje se aumenta aún más, el observador terminará percibiendo una doble imagen plana al no poder fusionar las imágenes correspondientes.

Los ojos humanos logran fusionar cómodamente imágenes con disparidad cuando los puntos de la imagen en la retina de un ojo tienen puntos correspondientes en el otro ojo dentro la región restrictiva conocida como *área fusionable de Panum*.

Las dimensiones del área de Panum no son constantes y son significativamente mayores en la periferia de la retina (hasta 6 veces más que en el centro). Por lo tanto un gran paralaje es tolerado si la imagen desplegada cubre una gran parte del campo de visión del observador, sin embargo si

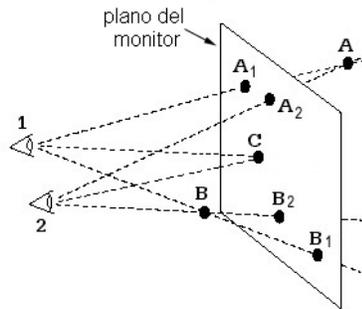


Figura 2.6: Relación entre profundidad y paralaje.

se considera el caso de televisiones de dimensiones regulares (de 12 a 16 pulgadas) como medios de despliegue la condición anterior no se cumple.

Por lo tanto el tamaño del área de Panum en el centro de la retina limita la máxima disparidad tolerable. Un valor de  $0.3^\circ$  de arco [7] es considerado razonable para condiciones donde el observador se encuentra a una distancia aproximada de 0.5m a 1.5m de un monitor de tamaño regular, sin embargo valores hasta de  $1.5^\circ$  son aceptables [12].

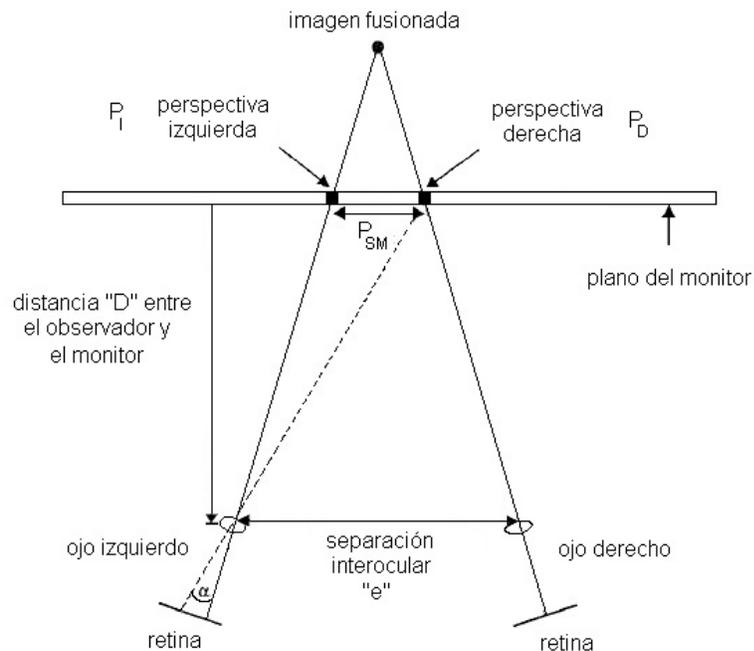


Figura 2.7: Relación entre el máximo paralaje permisible y el ángulo del área de Panum.

Los puntos  $p_I$  y  $p_D$  en la figura 2.7 son los límites del máximo paralaje permisible  $P_{SM}$ .

Asumiendo que la distancia entre el observador y el monitor es mucho más grande que la distancia interocular se puede obtener la siguiente expresión para relacionar el máximo paralaje con el ángulo

del área de Panum.

$$P_{SM} = d \tan \alpha \quad (2.1)$$

donde:

- $P_{SM}$  Máximo paralaje permisible en la pantalla.
- $d$  Distancia entre el observador y la pantalla
- $\alpha$  Ángulo del área fusionable de Panum.

Para las condiciones anteriormente mencionadas los valores de máximo paralaje son del orden de  $\pm 4$  a  $\pm 10$  mm.

### 2.4.2 Límites de *cercanía* y *lejanía* en la región estereoscópica útil

La región estereoscópica útil es formada por la intersección de los campos de vista de cada ojo. Cualquier objeto ubicado en esta zona genera una proyección correspondiente izquierda y una derecha a menos que una o ambas proyecciones tengan puntos ocultos por otros objetos.

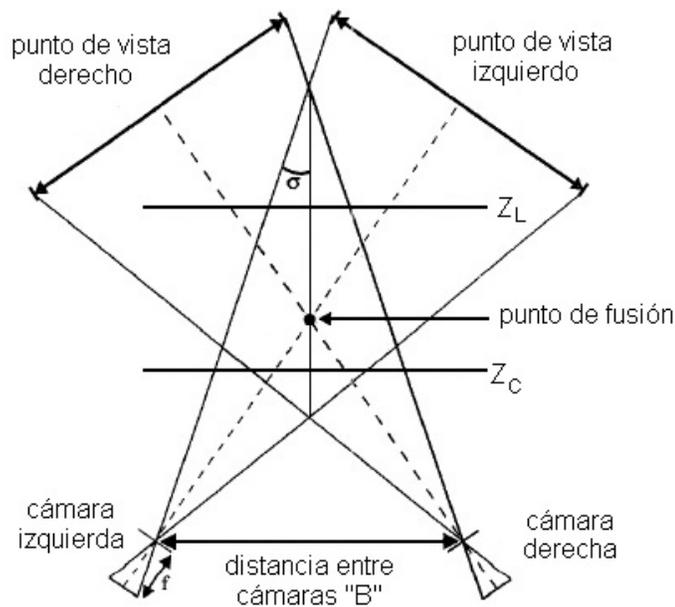


Figura 2.8: Región estereoscópica útil.

Para que el observador perciba *cómodamente* las imágenes estereoscópicas es necesario que el máximo paralaje en la pantalla limite la región estereoscópica para facilitar la fusión de las imágenes por el cerebro humano. Por lo tanto la región estereoscópica es truncada tanto en su extremo más cercano al observador como en el más lejano.

En la figura 2.8 se observan los cortes indicados por la terminología  $Z_C$  y  $Z_L$ .

Mediante los valores máximos de paralaje se puede establecer la posición de  $Z_C$  y  $Z_L$ , a través de la siguiente ecuación.

$$Z_{C:L} = \frac{t_c f}{2f \tan \sigma \pm D_M} \quad (2.2)$$

donde:

$t_c$  Distancia entre cámaras.

$f$  Longitud focal de los lentes de las cámaras.

$\sigma$  Ángulo de convergencia.

$D_M$  Máxima disparidad<sup>2</sup> permisible.

### 2.4.3 Resolución de profundidad

La resolución de profundidad se refiere al mínimo incremento en profundidad que puede ser detectado por la vista. Este parámetro constituye una medida de la efectividad de un sistema estereoscópico y se puede obtener reordenando la ecuación 2.2 y derivando  $D$  con respecto a  $Z$

$$\delta Z = \frac{D^2 \delta D}{t_c f} \quad (2.3)$$

donde:

$\delta D$  Razón de cambio en disparidad.

$\delta Z$  Razón de cambio en la profundidad detectable.

### 2.4.4 Distorsión de despliegue

La distorsión de despliegue es una amplificación o atenuación de la profundidad real de una escena por lo que los objetos pueden parecer estar situados a una profundidad diferente de la original. Estas distorsiones se deben a desalineamientos entre los pares de imágenes y a factores propios del funcionamiento de un sistema de reproducción de video.

Esta distorsión puede ser cuantificada mediante la siguiente ecuación:

$$\mu = \left[ \frac{t_c d}{e Z} \right] \quad (2.4)$$

donde:

$e$  Distancia interocular.

---

<sup>2</sup>Todo paralaje en un monitor tiene su contraparte en la retina humana, la cual se denomina disparidad retinal.

## Capítulo 3

### Metodología propuesta

De acuerdo a lo estipulado en los capítulos anteriores una imagen estereoscópica consiste de un par de imágenes con disparidad. Cuando se cuenta con un video monoscópico el punto clave reside en generar imágenes con el paralaje apropiado, basándose en varios aspectos de la secuencia bidimensional de video.

En general, las partes de una escena proyectada en una imagen 2D se mueven cuando el punto de vista cambia; esto es lo que se conoce como paralaje de movimiento (*motion parallax*). Esta característica por si misma permite el acceso a varios puntos de vista de una imagen en diferentes instantes por lo que mediante un cálculo adecuado de la dirección y magnitud del movimiento de la escena, es posible determinar la diferencia en tiempo entre dos cuadros para generar una imagen estereoscópica a partir de una secuencia monocular.

La conversión de una secuencia de video monoscópico estándar en uno estereoscópico se puede catalogar como un proceso de *interpolación espacio-temporal* que sintetiza una imagen estereoscópica dividiendo una señal convencional en dos canales (uno para cada ojo) con un retraso temporal entre ambos.

Dado que esta técnica implica una conversión 2D-3D que no obtiene exactamente la misma secuencia de video que si se grabara utilizando dos puntos de vista (dos cámaras), se le designará como **estéreo-artificial** para diferenciarla a lo largo de esta tesis.

### 3.1 Fundamento básico

La visión humana tiene componentes espaciales y temporales, tal y como lo describió Newton en la “persistencia de la visión” en su tratado *Opticks*.

La persistencia de la percepción de profundidad estereoscópica fue descubierta por Helmholtz en el siglo XIX, quien observó que una imagen binocular con disparidad produce una sensación de profundidad aún y cuando solo es presentada durante un instante en una secuencia a un observador. Sin embargo el parteaguas para la tecnología estéreo-artificial ocurrió en 1974, cuando el psicólogo John Ross [3, 22, 23] publicó los resultados acerca de la creación de estereoscopia por retraso binocular.

#### 3.1.1 Experimento canónico de Ross

Este experimento establece que un retraso binocular es un estímulo suficiente para obtener estereoscopia. Dicho experimento se fundamenta en la consideración de los procesos visuales al seguir objetos en movimiento (figura 3.1)

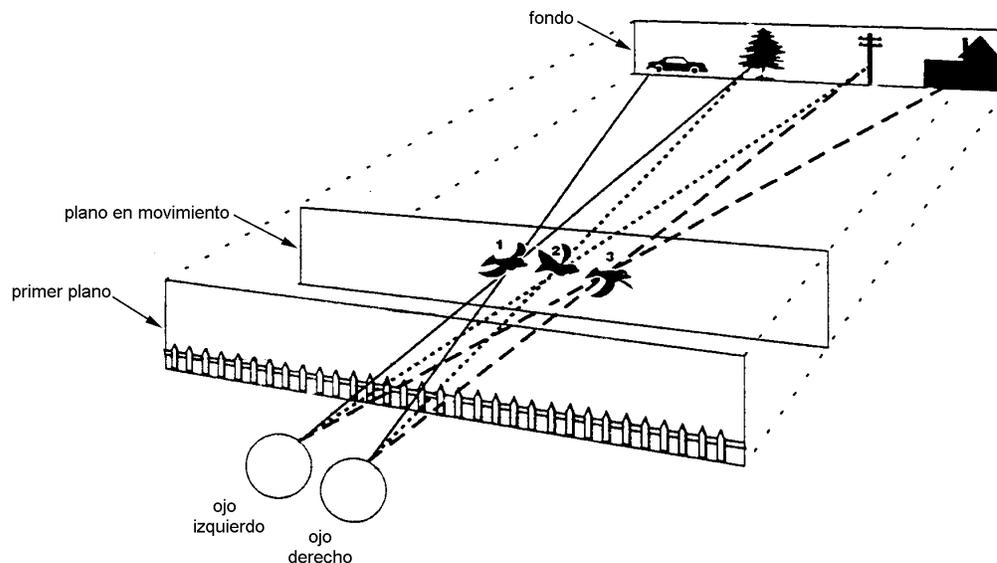


Figura 3.1: Proceso visual para objetos en movimiento.

Cuando los ojos se fijan en un objeto con movimiento horizontal, partes de los campos visuales de fondo y de primer plano son vistas en diferentes instantes por cada ojo, dependiendo de la magnitud y dirección del movimiento.

La teoría clásica establece que la estereoscopia es producida por disparidades binoculares instantáneas en los ojos del observador. Al seguir un objeto en movimiento un ojo “adelanta” al otro viendo diferentes porciones del campo visual en cada instante. En otras palabras cada ojo ve el mismo campo visual pero en tiempos diferentes. Por lo tanto para mantener la estereoscopia en tales circunstancias Ross dedujo que la visión debía poseer un factor temporal como lo es la memoria de corto término [22].

Para simular las condiciones del proceso visual mencionado, Ross adaptó el estereograma de puntos aleatorios (random-dot stereogram, RDS) de Julesz. Estos experimentos utilizando RDS examinan la percepción de la profundidad estereoscópica eliminando potencialmente factores monoculares y variables geométricas claves.

El experimento de Ross utiliza dos osciloscopios (uno para cada ojo) para desplegar los estereogramas. Un cuadrado es definido en el centro de la pantalla por una región de puntos presentados a ambos ojos de manera simultánea. El fondo que rodea a esta región está conformada por puntos de ruido dinámico; cada ojo visualiza el mismo ruido pero uno se encuentra retrasado con respecto al otro.

Ross encontró que lo anterior generaba una profundidad diferencial, ya que el cuadrado central parecía estar adelante del fondo, cuando el retraso interocular excedía los 50 mseg. Este efecto de profundidad persistía aún con retrasos mayores a 1 seg.

Para este experimento y otros posteriores [3, 22, 23] se concluyó que la visión humana maneja disparidades virtuales (las creadas por retrasos) de la misma forma que las disparidades reales (las creadas por desplazamientos espaciales). Así mismo se demostró que el umbral para la creación de disparidades virtuales es de 160 microsegundos, lo que se traduce en 2.4" de disparidad virtual "espacial".

## 3.2 Etapas de la metodología

El algoritmo para la construcción de secuencias estéreo-artificiales se basa en los principios básicos de disparidad virtual y se apoya en herramientas básicas de detección de movimiento que serán explicadas en el capítulo 5.

El algoritmo propuesto se puede descomponer en las siguientes etapas:

1. Digitalización y preprocesamiento de una secuencia de video monoscópico.
2. Análisis de detección de movimiento 2D.
  - Determinación del sentido del movimiento.
  - Determinación de la velocidad del movimiento.
3. Determinación de parámetros de retraso interocular y desplazamientos espaciales.
4. Generación de la imagen estéreo-artificial.
5. Procesamiento de la imagen estéreo-artificial para su despliegue mediante lentes anaglíficos (azul y rojo).

Cada una de estas etapas será descrita a detalle según la organización estipulada en la sección 1.3 y los resultados globales sobre la secuencia estéreo-artificial serán descritos en el capítulo 10.

Así mismo en la figura 3.2 se puede observar un diagrama de flujo que muestra la interacción entre las diversas etapas del algoritmo, así como algunas *etapas* de prueba sobre el desempeño de algunos bloques, específicamente sobre el detector de movimiento.

## 3.3 Hardware y software utilizado para la investigación

La digitalización y el procesamiento del video se realizó utilizando una estación de trabajo SUN Ultra 60. La tarjeta empleada para la captura de video fue SunVideo Plus. Dicha tarjeta tiene entradas de video en formato NTSC (National Television System Committee) o PAL (Phase Alternation Line), y el origen de la señal pueden ser video compuesto o S-Video. Sin embargo no cuenta con salida de video.

En la tabla 3.1 se resumen las características principales de la tarjeta SunVideo Plus en cuanto a video.

Los códigos para el procesamiento fueron programados en C ANSI por lo que pueden ser fácilmente transportados a otra plataforma como Windows.

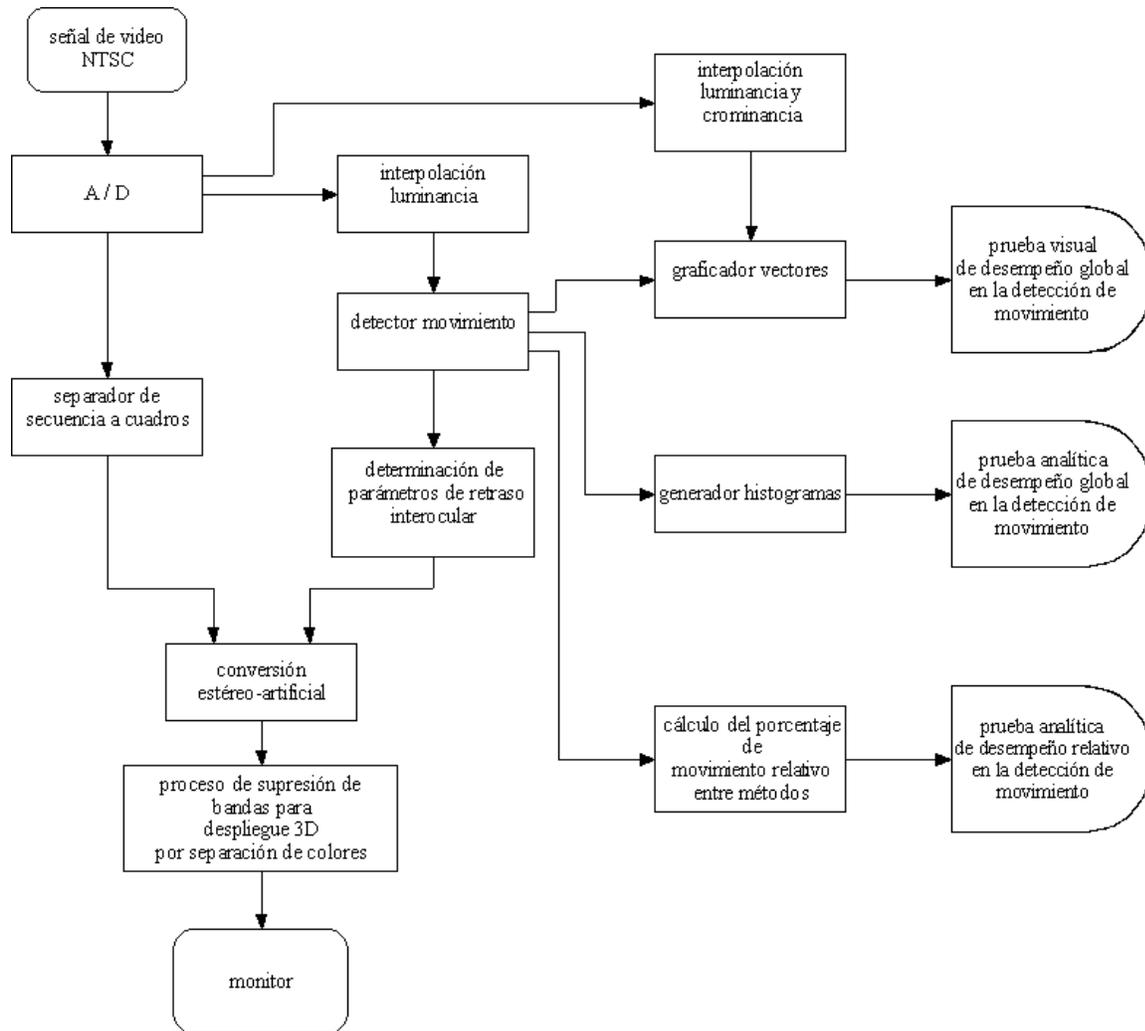


Figura 3.2: Diagrama de flujo del algoritmo.

Característica	Descripción
Entradas/salidas de video	Dos entradas de video compuesto, una de S-Video
Formatos de entrada	NTSC o PAL
Muestreo de video	YUV (4:2:2)
Compresión de video	CellB, H.261, JPEG, MPEG1 (CIF, I-Frame) - sólo NTSC
Formatos de video sin comprimir	YUV, RGB de 8 bits, RGB de 24 bits, Y de 8 bits
Plataformas de hardware	Sistemas Sun con bus PCI
Plataformas de software	Solaris 2.5.1 y versiones posteriores
Compatibilidad de SunVideo	Compatibilidad binaria con todas las aplicaciones de la tarjeta SunVideo basadas en XIL
Aplicaciones de demostración	Grabación y reproducción de video Videoconferencia a través de TCP/IP

Tabla 3.1: Características de la tarjeta SunVideo Plus.



## Capítulo 4

### Preprocesamiento

Como se mencionó en la sección 1.2, el formato de video seleccionado para esta investigación es NTSC, el cual constituye el estándar para televisión en México.

Para poder realizar un procesamiento estéreo-artificial es necesario tener la capacidad de acceder la señal de video en diferentes instantes, es decir en distintos cuadros, por lo que el video debe ser almacenado en forma digital para lograr tiempos de acceso muy cortos mejorando el proceso.

Existen numerosos formatos para digitalizar video, muchos de ellos utilizan técnicas de compresión ya que por su naturaleza las imágenes secuenciales tienden a poseer redundancia temporal y espacial. Sin embargo en esta investigación es indispensable poder acceder en forma individual cada uno de los *pixeles* que constituyen un cuadro de video, por lo que el formato utilizado debe ser sin compresión para no tener que resolver reglas de dependencias entre los valores de *pixeles* ya sea entre cuadros o entre *pixeles* vecinos en un cuadro.

Debido a ello se decidió utilizar el formato UYVY, el cual se describirá en la siguiente sección.

#### 4.1 Formato UYVY

El formato UYVY [2] constituye la versión de muestreo 4:2:2 del espacio de color YCbCr, el cual a su vez es una modificación del espacio de color YUV para equipos de cómputo.

El espacio de color YUV es el utilizado por el estándar PAL para televisión en Europa, mientras que el YIQ es utilizado en por el NTSC en Norteamérica y Japón.

La componente Y de YUV es la luminancia, es decir contiene la información necesaria para mostrar una imagen en blanco y negro en un televisión convencional.

La información de color se encuentra en las componentes U y V. La componente U (Cb) es la diferencia de crominancia azul (*Chrominance Blue*) mientras que V (Cr) es la diferencia crominancia roja (*Chrominance Red*). Por crominancia se entiende la cantidad de brillantez que es percibida por el ojo humano.

El formato UYVY es del tipo *empaquetado* ya que las muestras de Y, U y V son colocadas juntas en *macropixeles* (figura 4.1), es decir todas las componentes se encuentran almacenadas en un mismo arreglo.

El formato UYVY (YUV 4:2:2) tiene un periodo de muestreo horizontal de 2 tanto para la componente U como para la V, indicando que sus muestras son tomadas cada dos *pixeles* a lo largo de una línea. En cambio su periodo de muestreo vertical es unitario, por lo que el muestreo horizontal se aplica a todas las líneas. En otras palabras el intervalo de muestreo horizontal describe que tan frecuentemente a través de una línea una componente es muestreada mientras que el vertical indica cada cuántas líneas las muestras son tomadas.

macropixel = 2 pixeles de imagen

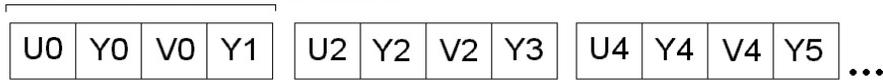


Figura 4.1: Formato UYVY.

A diferencia de las componentes de diferencia de crominancia la componente de luminancia tiene una muestra para cada *pixel* ya que el ojo humano es mucho más sensible a detalles de luminosidad que a colores.

	Horizontal	Vertical
Y Periodo de muestreo	1	1
V Periodo de muestreo	2	1
U Periodo de muestreo	2	1

Tabla 4.1: Periodos de muestreo UYVY.

Los rangos de valores para cada componente se muestran en la tabla 4.2, como se puede apreciar se requieren 8 bits para codificar cada muestra de cada componente.

	Rango Nominal	Blanco	Amarillo	Cyan	Verde	Magenta	Rojo	Azul	Negro
Y	16 a 235	180	168	145	133	63	51	28	16
U (Cb)	16 a 240	128	44	147	63	193	109	212	128
V (Cr)	16 a 240	128	136	44	52	204	212	120	128

Tabla 4.2: Rango de valores de las componentes YUV.

### 4.1.1 Dimensiones de la secuencia de video

Una imagen analógica de televisión en el sistema NTSC se compone aproximadamente de 475 líneas útiles o visibles por lo que normalmente se propone que el número de *pixeles* para la digitalización sea de 640 x 480 de tal forma que no se pierda resolución y se conserve la proporción 4:3 entre ancho y alto de la imagen.

Originalmente se propuso estas dimensiones para la implementación de los algoritmos sin embargo la tarjeta de digitalización utilizada (Sun Video Plus) no es capaz de obtener 30 cuadros por segundo a 640 x 480 *pixeles* por lo que al reproducir el video se observaban “brincos” en los objetos con movimiento.

Debido a lo anterior fue necesario utilizar un tamaño inferior al recomendable, es decir 320 x 240 *pixeles*, de tal suerte que sacrificando resolución se pudiese obtener una secuencia de video 4:3 a 30 cuadros por segundo.

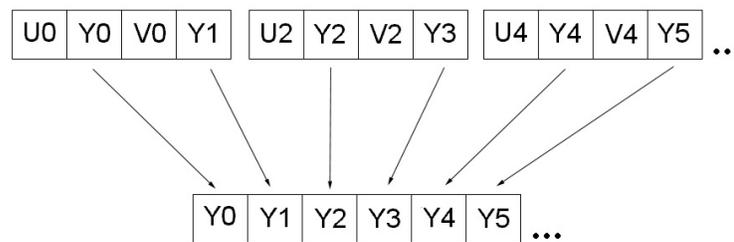
En la siguiente sección se describirá el preprocesamiento que es necesario dar a la secuencia de video antes de aplicarle el detector de movimiento.

## 4.2 Extracción e interpolación de luminancia

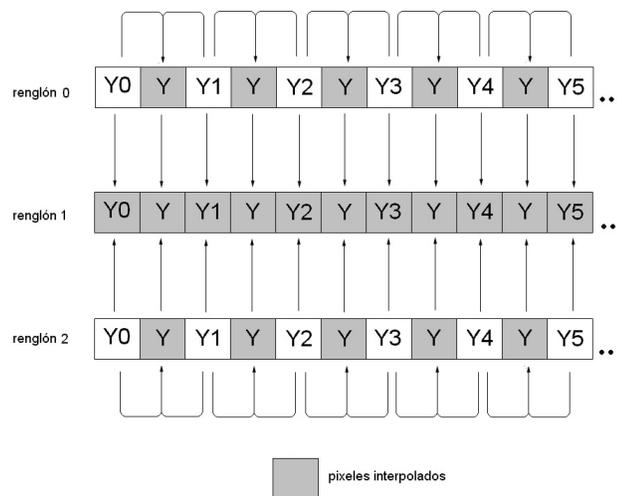
Según las investigaciones de Netravali [18] se observa que para la estimación de los parámetros de movimiento basta con analizar la componente de luminancia. Dado que la información de luminancia y color se encuentra empaquetada para el formato UYVY, es necesario extraer la componente de luminancia y colocarla en archivos independientes.

Así mismo se aplica una interpolación a dicha componente para obtener una imagen de 640 x 480 *pixeles* de tal suerte que el desempeño del detector de movimiento no sea perjudicado por la decisión de utilizar un tamaño de imagen menor al recomendado.

En la figura 4.2 se ilustra la extracción e interpolación de la componente de luminancia.



(a) extracción



(b) interpolación

Figura 4.2: Extracción e interpolación de luminancia.

En el siguiente capítulo se aborda el tema del detector de movimiento, en el cual se describe brevemente su teoría básica así como las pruebas de desempeño realizadas a varios algoritmos.



## Capítulo 5

### Detección de Movimiento

De acuerdo al capítulo 3, el proceso de detección de movimiento constituye una de las partes medulares de la conversión estéreo-artificial, ya que la magnitud y dirección del movimiento encontrados serán las variables clave para crear las disparidades virtuales necesarias.

Mediante los algoritmos de detección de movimiento se puede determinar los desplazamientos de objetos entre un cuadro de referencia y el actual.

Dicha detección se realiza encontrando el conjunto de vectores que indican los movimientos en la escena, basándose en las variaciones espacio-temporales las cuales son conocidas como flujo óptico.

Durante la última década se han desarrollado muchos algoritmos de estimación de movimiento, los cuales se pueden dividir en tres grandes categorías:

- Métodos basados en la ecuación de flujo óptico.
- Métodos recursivos de *pels* (bloque de *pixeles*).
- Métodos basados en bloques (*block matching*).

Todos estos métodos asumen que la iluminación es uniforme a lo largo de una trayectoria de movimiento y desprecian la oclusión de objetos (partes de objetos que son ocultos por otros que se encuentran más cercanos a las cámaras de grabación). Estas dos suposiciones a pesar de que por lo general no se cumplen en las secuencias de video del mundo real, son aún utilizadas por muchas técnicas de estimación de movimiento.

Las dos siguientes secciones describirán brevemente los métodos basados en la ecuación de flujo óptico y los métodos recursivos de *pels*, así mismo los métodos basados en bloques se cubrirán con más detalle en la sección 5.3

#### 5.1 Métodos basados en la ecuación de flujo óptico

La ecuación de flujo óptico restringe que los cambios de intensidad en un punto de una imagen se producen solamente por movimiento.

Estableciendo  $I(x_1, x_2, t)$  como la distribución de intensidades en el espacio-tiempo continuo, la ecuación de flujo óptico se puede expresar como

$$\frac{\partial I(\mathbf{x}; t)}{\partial x_1} v_1(\mathbf{x}; t) + \frac{\partial I(\mathbf{x}; t)}{\partial x_2} v_2(\mathbf{x}; t) + \frac{\partial I(\mathbf{x}; t)}{\partial t} = 0 \quad (5.1)$$

donde  $v_1(\mathbf{x}; t) = \frac{dx_1}{dt}$  y  $v_2(\mathbf{x}; t) = \frac{dx_2}{dt}$  denotan las componentes del vector de velocidad en términos de coordenadas en el espacio continuo.

Se puede observar claramente que esta ecuación no es suficiente para calcular cada una de las componentes de velocidad. En general los métodos basados en la ecuación de flujo óptico requieren condiciones adicionales para el cálculo de estas dos incógnitas.

## 5.2 Métodos recursivos de *pels*

Los métodos recursivos de *pels* son del tipo predictores-correctores con la siguiente forma:

$$\hat{\mathbf{d}}_a(\mathbf{x}, t; \Delta t) = \hat{\mathbf{d}}_b(\mathbf{x}, t; \Delta t) + \mathbf{u}(\mathbf{x}, t; \Delta t) \quad (5.2)$$

donde  $\hat{\mathbf{d}}_a(\mathbf{x}, t; \Delta t)$  denota un vector de estimación de movimiento en el punto  $\mathbf{x}$  en un tiempo  $t$ ,  $\hat{\mathbf{d}}_b(\mathbf{x}, t; \Delta t)$  un estimador predictivo de movimiento y  $\mathbf{u}(\mathbf{x}, t; \Delta t)$  es un término de actualización.

El estimador es actualizado recursivamente mediante varias iteraciones en la localidad  $(\mathbf{x}, t)$  para después proceder con el siguiente *pixel*.

Los algoritmos recursivos de *pels* tratan de minimizar la diferencia de desplazamiento entre cuadros (distorsión)  $\varepsilon$  la cual está definida como:

$$\varepsilon(\mathbf{x}, \mathbf{d}) = I(\mathbf{x} + \mathbf{d}(\mathbf{x}, t; \Delta t), t + \Delta t) - I(\mathbf{x}, t) \quad (5.3)$$

Netravali y Robbins [19] minimizan el cuadrado de  $\varepsilon(\mathbf{x}, \mathbf{d})$  en vez de utilizar la técnica de *steepest descent*, definiendo al término de actualización  $\mathbf{u}(\mathbf{x}, t; \Delta t)$  como:

$$\mathbf{u}(\mathbf{x}, t; \Delta t) = -\frac{\delta}{2} \nabla_d [\varepsilon^2(\mathbf{x}, \mathbf{d})] \quad (5.4)$$

donde  $\nabla_d$  es el gradiente con respecto a  $\mathbf{d}(\mathbf{x}, t; \Delta t)$ . La razón de convergencia está directamente influenciada por el tamaño del paso  $\delta$ , ya que pasos demasiado grandes provocan un comportamiento oscilatorio mientras que pasos demasiado pequeños causan que la convergencia sea lenta.

## 5.3 Métodos basados en bloques

Los métodos basados en bloques (*block matching*) constituyen el camino más sencillo para la estimación de movimiento en términos de implementación tanto en hardware como en software debido a su estructura uniforme y recursiva, por lo que su complejidad computacional es mucho menor que los métodos descritos anteriormente.

Cada cuadro es dividido en cierto número de bloques (cuadrados generalmente), asumiendo que cada uno de los *pixeles* dentro de cada bloque tiene el mismo movimiento traslacional entre cuadros. La tarea de los algoritmos es encontrar un vector de movimiento para cada bloque del cuadro actual comparándolo con los bloques candidatos dentro de la ventana de búsqueda en el cuadro de referencia. El par de bloques, el actual y el de referencia, que posean menor distorsión en su contenido, es decir, que sean más similares o exactamente iguales son los que determinan la dirección, sentido y magnitud del vector de movimiento.

Para determinar la mínima distorsión entre bloques existen varios criterios que son utilizados dependiendo del método empleado, algunos de ellos son:

- Función de correlación cruzada.
- Error cuadrático medio.

- Error absoluto medio.

Las diferencias entre los métodos de block matching radican principalmente en la naturaleza de sus métodos de búsqueda así como en los criterios de mínima distorsión que utilizan.

El algoritmo por excelencia es el de *full search* o *exhaustive search* el cual puede considerarse de fuerza bruta por su comportamiento que aunque sencillo requiere un gran número de operaciones, pues compara el bloque del cuadro actual con todos y cada uno de los bloques candidatos en la ventana de búsqueda del cuadro de referencia. Esto representa una alta calidad y confiabilidad de los vectores obtenidos sin embargo requiere de un enorme número de operaciones computacionales.

Dado que siempre se ha pretendido un equilibrio entre eficiencia y calidad, existen otros algoritmos llamados subóptimos que aunque computacionalmente son muy rápidos no poseen la misma exactitud que el de *full search*, sin embargo sus resultados son muy aceptables.

Así mismo existen varias técnicas de algoritmos subóptimos entre las cuales destacan las siguientes:

- Algoritmos basados en firmas: Reducen los cálculos utilizando varias etapas, usando criterios diferentes en cada una de ellas. En las primeras etapas se emplean criterios simples descartando bloques candidatos, a medida que avanza el algoritmo la complejidad computacional aumenta, sin embargo el número de candidatos disminuye considerablemente hasta encontrar la solución.
- Algoritmos basados en el Principio de Localidad: Dividen la búsqueda en etapas durante las cuales se va modificando el patrón de búsqueda de acuerdo a algún criterio. Un problema que presentan estos algoritmos es que pueden converger a un mínimo de distorsión local en vez de a uno global arrojando resultados insatisfactorios en la detección del movimiento. Estos algoritmos también se conocen como algoritmos de búsqueda con ventanas dinámicas.
- Algoritmos basados en dependencia espacial: Utilizan la homogeneidad de movimientos que se presentan en una imagen, es decir predicen el movimiento de un bloque en función de los vectores de movimiento de los bloques que lo rodean.
- Algoritmos basados en dependencia temporal: Predicen el movimiento de un bloque en un determinado cuadro en función del vector de ese mismo bloque en el cuadro anterior al analizado.
- Algoritmos basados en jerarquías: Particionan una imagen en componentes de baja resolución y calculan vectores de movimiento para dicho cuadro, que después serán propagados a la misma imagen pero con mayor resolución en cada iteración, obteniendo en cada paso vectores más exactos hasta llegar al nivel original.

Para esta investigación se decidió utilizar los métodos basados en bloques por su facilidad de implementación tanto en hardware como en software debido su estructura uniforme y recursiva.

A continuación se muestra una descripción del funcionamiento de los algoritmos implementados, los cuales fueron obtenidos y modificados de un paquete de métodos de detección de movimiento suministrado por *CityU Image Processing Laboratory*. Dichos algoritmos se encuentran programados en C y la mayoría de ellos se basa en el Principio de Localidad. El nombre del paquete que se modificó es Various Advanced Motion Estimation Research Development Package [11].

### 5.3.1 Algoritmos

#### 5.3.1.1 Full search (FS)

Teniendo cada imagen dividida en bloques de  $N \times N$  *pixeles*, este algoritmo determina un vector de movimiento para cada bloque del cuadro actual, comparándolo con todos los bloques candidatos contenidos en la ventana de búsqueda alrededor del bloque de referencia, obteniendo el vector entre el bloque actual y el bloque en el cuadro de referencia con menor distorsión entre ambos.

Se requieren  $(2p + 1)^2$  cálculos donde  $p$  es el máximo desplazamiento permitido en dirección horizontal o vertical de un bloque definiendo de esta forma la zona de búsqueda (figura 5.1).

Sin embargo debido a su alta complejidad computacional este método no resulta útil para implementaciones en tiempo real.

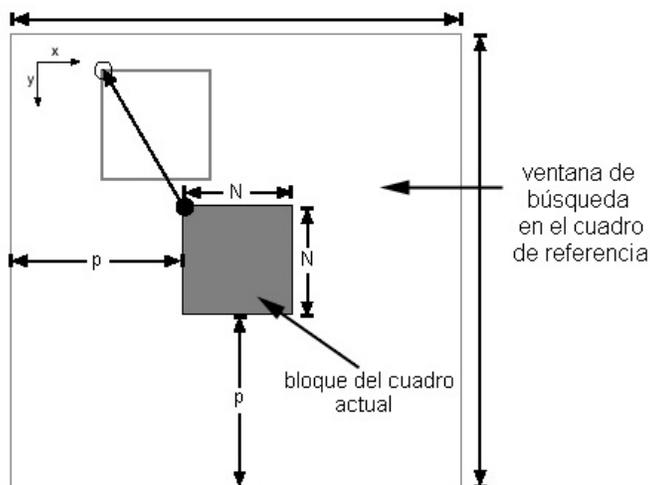


Figura 5.1: Full search.

Para realizar la búsqueda este algoritmo sigue el patrón mostrado en la figura 5.2.

#### 5.3.1.2 Three step search (3SS)

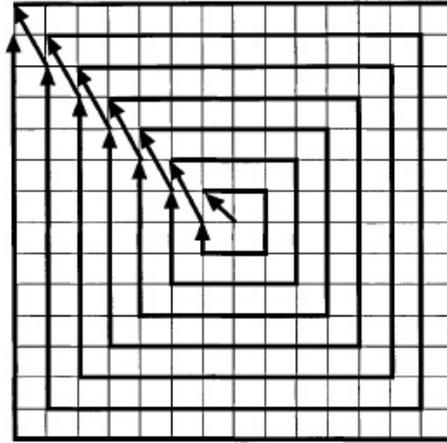
Fue introducido por Koga en 1981 [25]. Es muy popular debido a su simplicidad, a su robustez y a su alta efectividad. Este algoritmo comienza con una búsqueda burda para terminar con un patrón fino.

Su operación se puede describir de la siguiente manera:

Etapa 1: Se establece un paso inicial equivalente a la mitad del máximo desplazamiento permitido  $p$ , es decir  $\lceil \frac{p}{2} \rceil$ <sup>1</sup>. Basándose en el valor de este paso ocho bloques son seleccionados alrededor del bloque central, los cuales son analizados para localizar el que posea la mínima distorsión.

Etapa 2: El paso inicial se reduce a la mitad y el centro del patrón de búsqueda se mueve al punto que tuvo menor distorsión en la etapa anterior. Se utiliza la misma técnica descrita en la etapa 1, analizando nuevamente 9 bloques en este paso.

<sup>1</sup>El operador  $\lceil \cdot \rceil$  indica redondeo del valor obtenido.

Figura 5.2: Patrón de búsqueda de *full search*.

Las etapas 1 y 2 son repetidas hasta que el tamaño del paso sea menor que 1, es decir que se obtenga un valor de redondeo igual a 1.

Normalmente se establece una distancia de máximo desplazamiento  $p$  de 7 *pixeles* por lo que el algoritmo alcanzará un paso menor a 1 en tan solo tres etapas, de ahí que se deriva su nombre.

El número máximo de bloques sobre los cuales es necesario realizar el análisis es de 25 para un desplazamiento máximo de  $p = 7$  *pixeles*.

Para mayores desplazamientos se puede utilizar la siguiente expresión para calcular el número máximo de puntos.

$$1 + 8 \lceil \log_2 (p + 1) \rceil \quad (5.5)$$

En la figura 5.3 se muestra gráficamente la convergencia del método.

#### 5.3.1.3 Four step search (4SS)

Este algoritmo fue propuesto por L. M. Po y W. C. Ma en 1996 [10]. Basándose en los mismos parámetros del algoritmo *three step search* (distancia de máximo desplazamiento  $p$  de 7 *pixeles*) el algoritmo puede describirse de la siguiente manera:

1. Un punto de mínima distorsión es encontrado entre los 9 puntos seleccionados utilizando un paso inicial equivalente a  $\lceil \frac{p}{4} \rceil$ , es decir utiliza un paso más pequeño que el empleado por el método de tres pasos, debido a ello este método utilizará cuatro etapas como máximo para encontrar el vector de movimiento si se fija un desplazamiento máximo de  $p = 7$ . Si el punto mínimo se encuentra en el centro del patrón, se pasa a la etapa 4, si no se procede con la etapa 2.
2. El tamaño del paso no se reduce, sin embargo el patrón de búsqueda es modificado de acuerdo a la posición del punto de mínima distorsión.
  - (a) Si el punto previo de mínima distorsión se encontró en la esquina del patrón anterior de búsqueda se agregan 5 puntos adicionales para el análisis en vez de 8 como en el *three step search*.

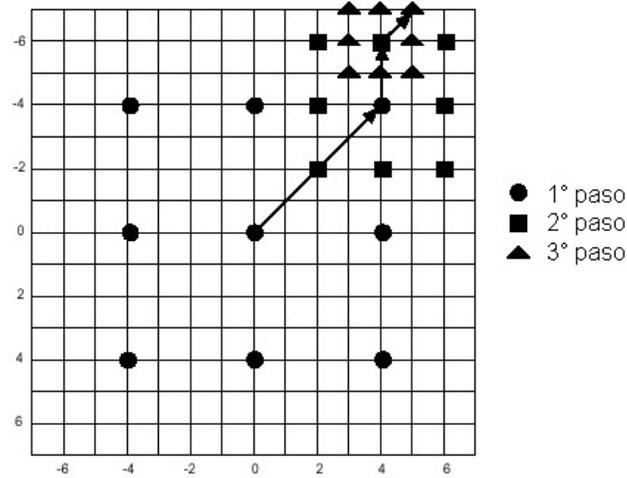


Figura 5.3: Three step search.

- (b) Si el punto previo de mínima distorsión se encontró en el centro de una arista del patrón anterior de búsqueda se agregan solo 3 puntos adicionales para el análisis.
- Si el mínimo se encontró en el centro del patrón se pasa a la etapa 4, si no a la etapa 3
3. La estrategia de búsqueda es la misma que en la etapa anterior, excepto que en cualquier caso se pasa a la etapa 4.
  4. El paso del patrón de búsqueda es reducido a  $\lceil \frac{p}{8} \rceil$  y el punto de mínima distorsión se encuentra analizando el punto obtenido la etapa anterior y los 8 puntos que lo rodean.

El número máximo de puntos (peor caso) sobre los cuales es necesario realizar el análisis es de 27 y esto sucede cuando se tienen movimientos grandes.

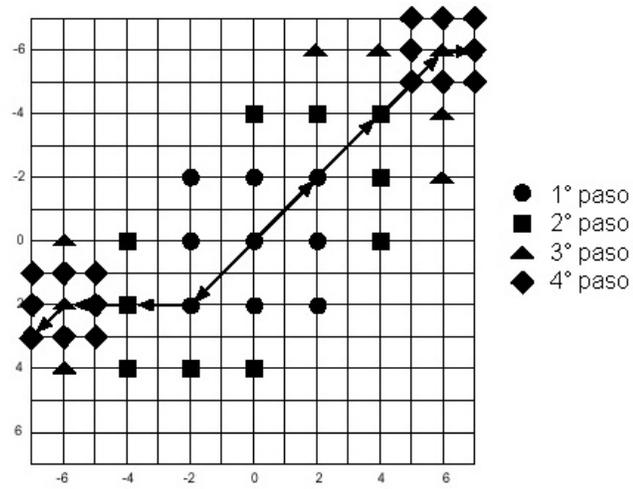
Si después de la cuarta etapa el paso del patrón de búsqueda es mayor a 1, situación que se presenta si el desplazamiento máximo establecido  $p$  es mayor a 7 *pixeles*, se puede realizar otra búsqueda 4SS partiendo del último punto obtenido. En esta situación el número máximo de puntos a verificar (peor caso) se puede obtener de la siguiente expresión:

$$18 \log_2 \left[ \frac{p+1}{4} \right] + 9 \quad (5.6)$$

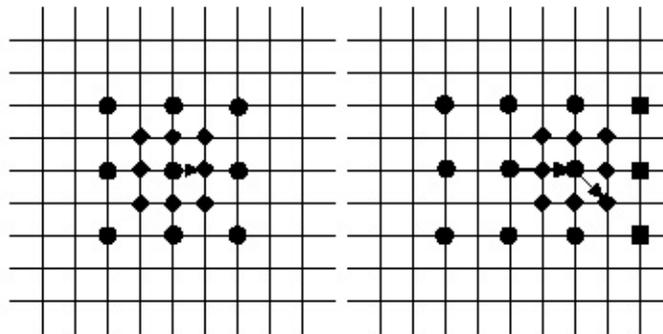
En las figuras 5.4(a) y 5.4(b) se muestra gráficamente la convergencia del método para dos diferentes rutas de búsqueda con movimientos grandes y para dos casos con movimientos pequeños.

#### 5.3.1.4 New three step search (N3SS)

Este algoritmo fue propuesto por Li, Zeng y Liou en 1994 [21]. Es una versión modificada del algoritmo *three step search* para búsquedas en secuencias con movimientos pequeños, cuyos vectores generalmente se encuentran en la parte central de las ventanas de búsqueda.



(a) Desplazamientos grandes



(b) Desplazamientos pequeños

Figura 5.4: Four step search.

Por lo tanto este algoritmo se diferencia del 3SS en que propone agregar los 8 puntos más cercanos al punto central y en que utiliza una técnica de *halfway-stop* al igual que 4SS.

El algoritmo puede detallarse de la siguiente manera:

1. En el primer paso se utilizan 8 puntos adicionales a los utilizados por 3SS. Estos puntos adicionales son los más cercanos al punto central de la ventana de búsqueda.
2. Una técnica de *halfway-stop* es empleada para identificar bloques sin movimiento o con desplazamientos muy pequeños.
  - (a) Si el punto de mínima distorsión en el primer paso se encuentra en el centro de la ventana de búsqueda se termina la búsqueda. (Esto se conoce como *first-step-stop*.)
  - (b) Si el punto de mínima distorsión en el primer paso se encuentra entre uno de los 8 puntos adicionales a los utilizados por 3SS se realiza otro paso de búsqueda y se detiene el algoritmo. (Esto se conoce como *second-step-stop*.)
    - i. Si el punto previo de mínima distorsión se encontró en la esquina del patrón anterior de búsqueda se agregan 5 puntos adicionales para el análisis en vez de 8.
    - ii. Si el punto previo de mínima distorsión se encontró en el centro de una arista del patrón anterior de búsqueda se agregan solamente 3 puntos adicionales para el análisis.
  - (c) Si el punto de mínima distorsión es cualquier otro de los puntos analizados se continua el algoritmo de la misma forma que 3SS.

En la figura 5.5 se muestra la convergencia del método.

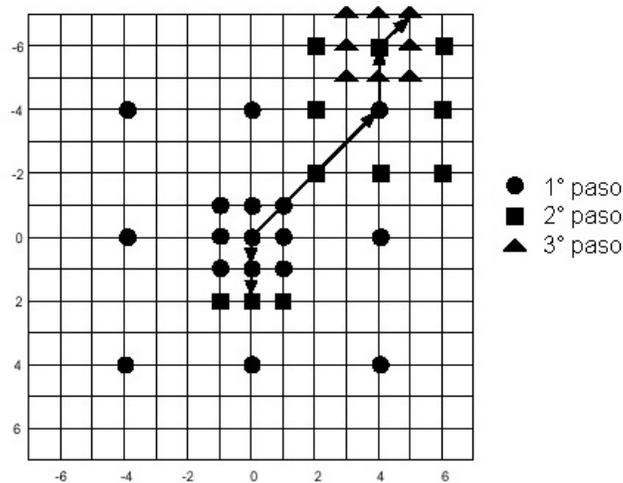


Figura 5.5: New three step search.

### 5.3.1.5 Block-based gradient descent (BBGDS)

Fue propuesto por L. K. Liu y E. Feig en 1996 [13]. Este algoritmo utiliza patrones de búsqueda de 9 puntos separados tan solo por un pixel, esta condición no cambia a lo largo del método. Así mismo no tiene restricciones en cuanto al número de etapas efectuadas, ya que se detiene cuando el punto central del patrón es el que tiene mínima distorsión o cuando se alcanza la orilla de la ventana de búsqueda.

Las búsquedas realizadas con este algoritmo son más efectivas cuando se tienen pequeños movimientos en las secuencias.

En la figura 5.6 se muestra gráficamente la convergencia del método para dos rutas diferentes.

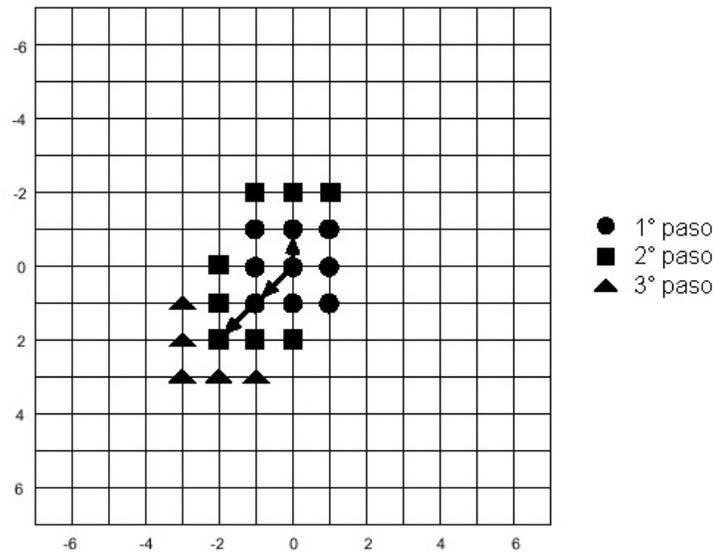


Figura 5.6: Block based gradient search.

### 5.3.1.6 Hybrid search (HS)

Este método se basa en que algunos algoritmos tienen mejor desempeño para movimientos grandes y otros para desplazamientos pequeños, por lo que emplea procedimientos diferentes según la clasificación de movimiento detectado.

Para movimientos grandes utiliza 4SS mientras que para desplazamientos pequeños emplea BBGDS. Para diferenciar entre ambos tipos de movimientos compara la distorsión inicial entre el bloque en el cuadro actual y el bloque ubicado en la misma posición en el cuadro de referencia contra un umbral  $T_n$ , el cual se actualiza después del análisis de cada bloque.

El algoritmo se puede describir de la siguiente manera:

1. Calcula la distorsión total  $D_n(0,0)$  entre el bloque en el cuadro actual y el bloque ubicado en la misma posición en el cuadro de referencia.
2. Si  $D_n(0,0) < T_n$  se clasifica al movimiento del bloque como pequeño y se continúa con el paso 3, en caso contrario se clasifica como grande y se va al paso 4.
3. Se utiliza BBGDS para buscar el vector de movimiento y después se pasa a la etapa 5.

4. Se utiliza 4SS para buscar el vector de movimiento y después se pasa a la etapa 5.
5. Se actualiza el umbral  $T_n$  basándose en la magnitud del vector de salida  $M_n$  y en la distorsión inicial  $D_n(0,0)$ . El procedimiento se repite para el siguiente bloque utilizando el patrón mostrado en la figura 5.7

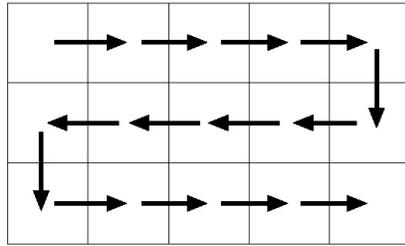


Figura 5.7: Orden utilizado para la búsqueda por HS.

Debido a que las características locales de movimiento cambian a lo largo del cuadro, es necesario actualizar constantemente el umbral  $T_n$

El modelo matemático que muestra la relación (5.8) entre  $D(0,0)$  y  $M$  es el siguiente:

$$D(0,0) = kM + C \quad (5.7)$$

donde  $k$  y  $C$  son el gradiente y el *offset* de la regresión lineal respectivamente.

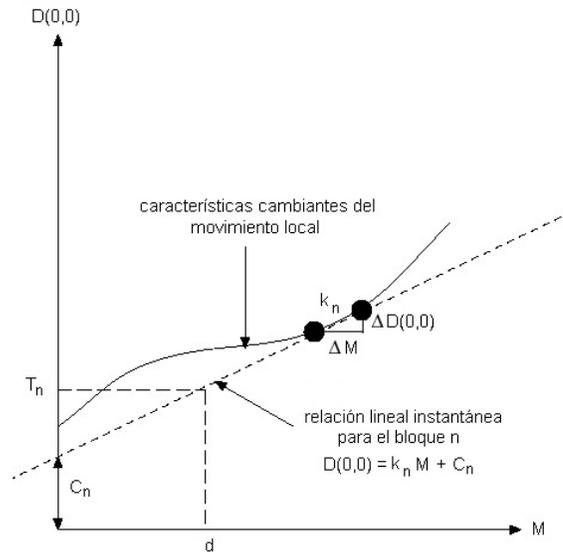


Figura 5.8: Relación entre  $D(0,0)$  y  $M$ .

De acuerdo a esta relación el umbral  $T_n$  se actualiza utilizando la siguiente ecuación:

$$T_n = k_{n-1}d + C_{n-1} \quad (5.8)$$

donde  $d$  es la menor magnitud del vector de movimiento para que un bloque sea clasificado como con desplazamiento grande.

Con el fin de aprovechar la correlación entre bloques, el proceso de búsqueda comienza de izquierda a derecha para el primer renglón y el orden es invertido cada renglón como se muestra en la figura 5.7

Para el bloque  $n$  si la distorsión  $D_n(0,0)$  no se incrementa al aumentar la magnitud del vector de movimiento, es decir si  $\frac{\Delta D(0,0)}{\Delta M} \leq 0$  o si el cambio de la magnitud del vector de movimiento es cero ( $\Delta M = 0$ ), el valor  $C_n$  es actualizado utilizando la siguiente ecuación:

$$C_n = (1 - u)C_{n-1} + u(D_n(0,0) - k_{n-1}M_n) \quad (5.9)$$

Si las dos condiciones anteriores no se cumplen entonces el valor de  $k_n$  se actualiza mediante la siguiente expresión:

$$k_n = (1 - v)k_{n-1} + v \frac{\Delta D(0,0)}{\Delta M} \quad (5.10)$$

donde  $u$  y  $v$  son términos de ajuste para la adaptación de  $C_n$  y  $k_n$  respectivamente. En esta investigación tanto  $u$  como  $v$  se fijaron en 0.002

Así mismo el valor de  $d$  se estableció en  $\sqrt{8}$ , es decir equivale al valor de la distancia entre el centro y una esquina en una ventana de búsqueda de  $5 \times 5$  del algoritmo 4SS.

Los valores iniciales de  $k_n$  y de  $C_n$  se fijaron en 500 y 12800 respectivamente, sin embargo estos valores pueden ser seleccionados arbitrariamente puesto que serán actualizados durante la búsqueda.

### 5.3.1.7 Coherent search (CS)

Existen dos casos específicos donde los algoritmos basados el Principio de Localidad pueden fallar:

1. Cuando no existen patrones distintivos dentro del bloque (*bloques planos*).

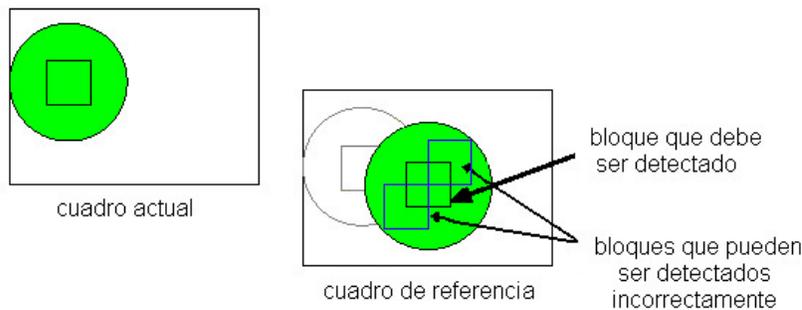


Figura 5.9: Bloques sin patrones distintivos (planos).

2. Cuando solamente hay líneas rectas dentro del bloque, patrón que se puede repetir en varios bloques al igual que en el caso anterior (*bloques repetitivos*).

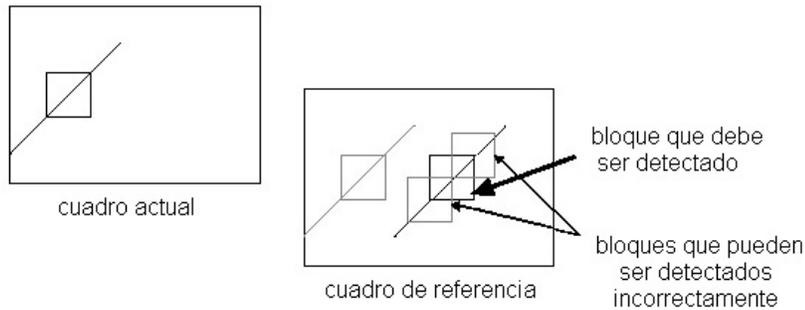


Figura 5.10: Bloques con patrones repetitivos.

En ambos casos aunque los vectores de movimiento son similares para ciertas regiones, los algoritmos basados en el Principio de Localidad generalmente obtienen resultados incorrectos, por lo que la coherencia de los vectores no se mantiene.

Mantener dicha coherencia tiene como ventajas mejorar la exactitud de la estimación del movimiento en los casos especiales mencionados anteriormente y reducir el número de *bits* requeridos para codificar los vectores de movimiento.

El algoritmo de búsqueda coherente propone encontrar vectores de movimiento homogéneo para un bloque tomando la mediana de los vectores de los bloques que lo rodean.

Así mismo se establece una medida de distorsión la cual se utiliza para seleccionar entre el vector encontrado por el algoritmo de Principio de Localidad o el vector de movimiento coherente, de tal suerte que se escoja la mejor solución para la estimación de movimiento.

En la figura 5.11 se muestra los bloques aledaños utilizados para calcular el vector coherente de un determinado bloque:

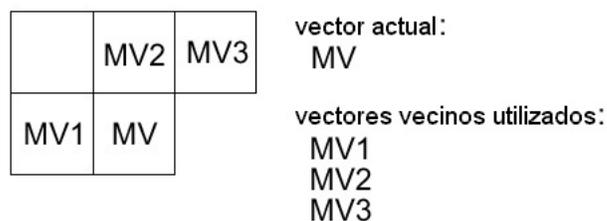


Figura 5.11: Bloques utilizados para el cálculo de la predicción coherente.

### 5.3.1.8 Partial distortion full search (PDFS)

Este método es muy similar al algoritmo de *full search* ya que recorre a través de la misma ruta toda la ventana de búsqueda (figura 5.2).

La principal diferencia con respecto al método de *full search* es que no realiza el cálculo completo

de la distorsión para cada bloque candidato del cuadro de referencia. Se supone que cada vector consiste de  $k$  componentes, por lo que la distorsión total es obtenida sumando todas y cada una de las distorsiones de las  $k$  componentes.

El algoritmo compara las distorsiones parciales normalizadas de cada bloque candidato en lugar de comparar la distorsión total. Un bloque candidato es descartado si su distorsión parcial normalizada es mayor que la mínima distorsión total normalizada encontrada hasta ese momento. De otra manera el cálculo de la distorsión es continuado y el bloque candidato se convierte en el nuevo bloque elegido si su distorsión total es menor a la mínima distorsión total actual.

Utilizando esta técnica, la probabilidad de rechazar más rápidamente un bloque erróneo se incrementa y el número de cálculos para la distorsión se reduce considerablemente.

Este algoritmo divide la distorsión total en 16 distorsiones parciales cada una compuesta por 16 puntos espaciados igualmente como se muestra en la figura 5.12.

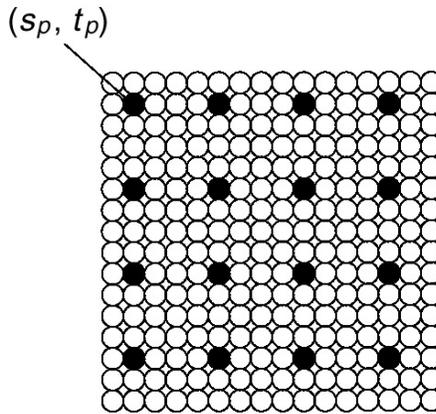


Figura 5.12: Distorsiones parciales.

Bajo este esquema la  $p$ -ésima distorsión utilizando el criterio del error absoluto medio se puede expresar como:

$$d_p(k, l; u, v) = \sum_{i=0}^3 \sum_{j=0}^3 |I_n(k + 4i + s_p, l + 4j + t_p) - I_{n-1}(k + 4i + s_p + u, l + 4j + t_p + v)| \quad (5.11)$$

Esta ecuación no incluye el término divisor  $\frac{1}{MN}$  es decir  $\frac{1}{256}$  para bloques de 16 *pixeles* de arista, ya que de esta manera se puede reducir la complejidad computacional de la implementación en software.

Los valores de  $s_p$  y  $t_p$  son los desplazamientos horizontal y vertical desde la esquina superior izquierda del bloque y sus valores son diferentes según el orden de la distorsión parcial calculada. Esto se ilustra en la tabla 5.1.

En la figura 5.13 se muestra la ubicación definida para cada distorsión parcial en función de los valores de  $s_p$  y  $t_p$ .

Así mismo la distorsión parcial acumulada  $D_p(k, l; u, v)$  se define como:

$$D_p(k, l; u, v) = \sum_{i=1}^p d_i(k, l; u, v) \quad (5.12)$$

$p$	$(s_p, t_p)$	$p$	$(s_p, t_p)$
1	(0,0)	9	(1,0)
2	(2,2)	10	(3,2)
3	(2,0)	11	(0,1)
4	(0,2)	12	(2,3)
5	(1,1)	13	(3,0)
6	(3,3)	14	(1,2)
7	(3,1)	15	(2,1)
8	(1,3)	16	(0,3)

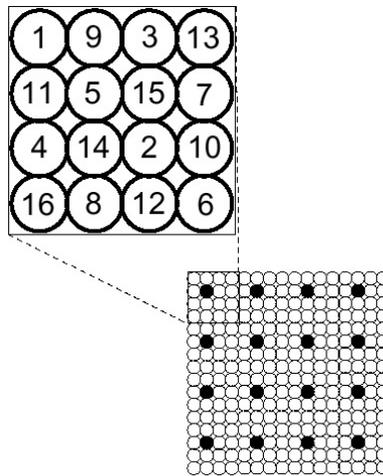
Tabla 5.1: Valores de  $s_p$  y  $t_p$  para cada distorsión  $p$ .

Figura 5.13: Ubicación de cada distorsión parcial.

Finalmente la comparación se realiza entre la distorsión parcial acumulada normalizada  $N_p$  y la distorsión total mínima actual normalizada  $N_{min}$ :

$$N_{min} = pD_{min}(k, l; u_m, v_m) \quad (5.13)$$

$$N_p = 16D_p(k, l; u, v) \quad (5.14)$$

De tal suerte que si para determinada distorsión parcial su acumulación normalizada rebasa el mínimo normalizado actual se elimina el bloque candidato. En caso contrario se continua hasta calcular la totalidad de las distorsiones empleándose el criterio básico de comparación del método *full search*.

### 5.3.1.9 Unrestricted center-biased diamond search (UCBDS)

Fue propuesto por J. Tham, S. Ranganath, M. Ranganath y A. Kassim en 1998 [24]. Su característica de una estructura compacta (diamante) es muy útil para detección de pequeños movimientos en el centro de la ventana de búsqueda.

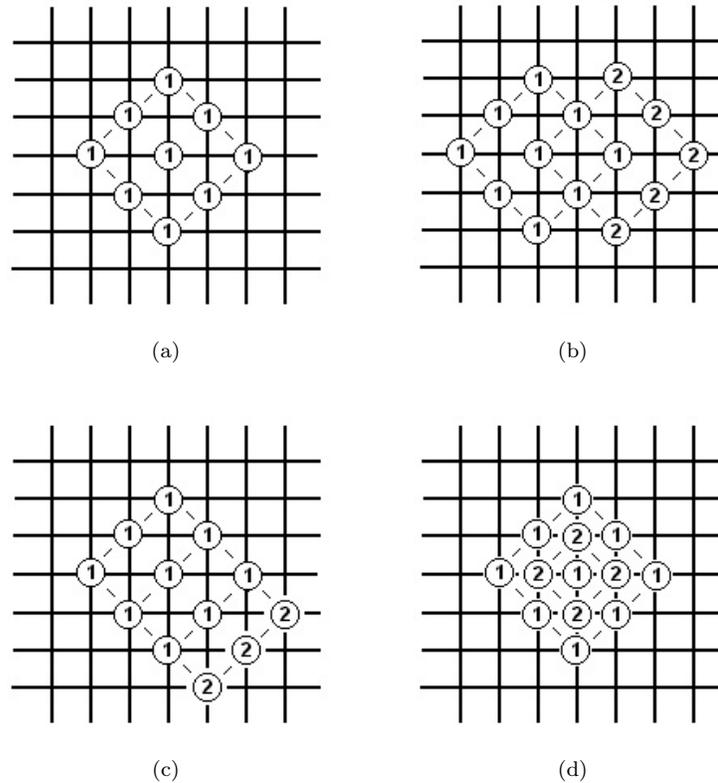


Figura 5.14: Unrestricted center-biased diamond search.

El algoritmo se puede describir de la siguiente manera:

- **Inicio:** Se coloca un patrón de diamante en el centro de la ventana de búsqueda como se muestra en la figura 5.14(a). Se evalúa la distorsión para cada uno de los nueve puntos candidatos. Si el mínimo se encuentra en el punto central se pasa a la etapa de **Fin**; de otra manera se procede con la etapa de **Búsqueda**.
- **Búsqueda:** Si la mínima distorsión de la etapa anterior se encuentra en uno de los cuatro vértices se procede con la etapa **Búsqueda de Vértice**, si por el contrario el mínimo se encuentra en una de las caras del diamante se utiliza la **Búsqueda Lateral**.
  - **Búsqueda de Vértice:** El patrón de diamante se desplaza como se muestra en la figura 5.14(b), de tal suerte que el punto central del diamante es el de mínima distorsión encontrado en la etapa anterior. Ahora cinco nuevos puntos candidatos son evaluados.
  - **Búsqueda Lateral:** El patrón de diamante se desplaza como se muestra en la figura 5.14(c), de tal suerte que el punto central del diamante es el de mínima distorsión encontrado en la etapa anterior. Ahora tres nuevos puntos candidatos son evaluados.

Si alguno de los nuevos puntos se encuentra fuera de la ventana de búsqueda, es ignorado para la comparación de distorsiones.

Nuevamente una vez identificado el mínimo, si éste se encuentra en el punto central del diamante se procede a la etapa de **Fin**, de otra manera se realiza una nueva etapa de **Búsqueda**.

- **Fin:** El patrón de diamante es reducido utilizando el mismo centro. Ahora se compara la distorsión del punto central con la de los cuatro nuevos puntos que eran contenidos por el patrón anterior (figura 5.14(d)). De manera similar cualquier punto que se encuentre fuera de la ventana de búsqueda no es considerado. La posición del punto con menor distorsión es utilizada para calcular el vector de movimiento, por lo que el proceso de búsqueda de movimiento para el bloque actual es terminado.

### 5.3.2 Criterios para medición de distorsión entre bloques

Para medir la similitud entre el bloque del cuadro actual y un bloque candidato del cuadro de referencia se utilizan varios criterios como el error cuadrático medio *mean square error* (MSE) o el error absoluto medio *mean absolute error* (MAE).

Suponiendo que la esquina superior izquierda del cuadro actual  $n$  es  $(k, l)$  y el desplazamiento hacia el bloque candidato en el cuadro de referencia  $n - 1$  es  $(u, v)$  y el tamaño del bloque es  $M \times N$ , los errores MSE y MAE se pueden definir como se menciona a continuación:

#### 5.3.2.1 MSE

$$MSE(k, l; u, v) = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (I_n(k+i, l+j) - I_{n-1}(k+i+u, l+j+v))^2 \quad (5.15)$$

El tener que efectuar el cuadrado en la ecuación 5.15 complica la implementación en hardware, por lo que la mayoría de los codificadores y detectores empleados en video utilizan el error MAE como criterio base para la detección de similitud de bloques.

#### 5.3.2.2 MAE

$$MAE(k, l; u, v) = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} |I_n(k+i, l+j) - I_{n-1}(k+i+u, l+j+v)| \quad (5.16)$$

Tal como se mencionó este criterio es el más popular para implementaciones en hardware ya que su complejidad es mucho menor que el criterio MSE.

### 5.3.3 Parámetros para la medición de desempeño de detección de movimiento entre cuadros

Para la evaluación del desempeño de los diversos algoritmos de *block matching* se calcula la distorsión total entre el cuadro original o actual y el mismo cuadro pero reconstruido a partir del cuadro de referencia y los vectores de movimiento detectados. Para que la efectividad del método sea alta, la distorsión debe ser lo más pequeña posible.

El criterio de error cuadrático medio (MSE) es generalmente el utilizado para medir la exactitud en la detección de movimiento.

Suponiendo que el tamaño de cada cuadro de video es de  $W \times H$  *pixeles*, el MSE entre el cuadro original  $n$  y el cuadro generado por vectores resultantes del análisis del cuadro de referencia es calculado como:

$$D_f(n) = \frac{1}{ST} \sum_{i=0}^{S-1} \sum_{j=0}^{T-1} D_b^n(i, j; u_{ij}, v_{ij}) \quad (5.17)$$

donde  $S = \lceil \frac{W}{B} \rceil$  y  $T = \lceil \frac{H}{B} \rceil$ .

$D_b^n(i, j; u_{ij}, v_{ij})$  constituye la distorsión (MSE) entre el bloque  $(i, j)$  del cuadro original  $n$  y el cuadro obtenido por medio de los vectores de movimiento detectados a partir del cuadro  $n - 1$ . Este distorsión es calculada de la siguiente forma:

$$D_b^n(i, j; u_{ij}, v_{ij}) = \frac{1}{B^2} \sum_{k=0}^{B-1} \sum_{l=0}^{B-1} (I_n(iB + k, jB + l) - I_{n-1}(iB + k + u_{ij}, jB + l + v_{ij}))^2 \quad (5.18)$$

donde  $u_{ij}$  y  $v_{ij}$  son las componentes del vector de movimiento del bloque  $(i, j)$  del cuadro  $n$ .  $B$  es el ancho y el alto de cada bloque, valor que en esta investigación es de 16 *pixeles*.

Así mismo aparte de los criterios de error en la detección de movimiento, la complejidad computacional también es un criterio clave para medir el desempeño de los algoritmos de *block matching*.

Para los algoritmos que pueden controlar o reducir su complejidad computacional limitando el número de puntos a verificar como 3SS, 4SS, N3SS, etc.; la complejidad computacional para determinar la distorsión para cada bloque es la misma. Por lo tanto su valor se puede determinar directamente contando el número de puntos analizados a lo largo del algoritmo.

En cambio para aquellos algoritmos basados en jerarquía o dependencias espacial o temporal, su complejidad se determina en base al número de operaciones realizadas para encontrar el vector de movimiento de cada bloque analizado. Estas operaciones se pueden agrupar en *adiciones, valores absolutos y comparaciones*.

### 5.3.4 Parámetros para la medición de desempeño de detección de movimiento aplicada a la conversión estéreo-artificial

Resulta necesario establecer una métrica distinta de las mencionadas anteriormente para evaluar el desempeño de los algoritmos de detección de movimiento aplicados a la conversión monocular - estereoscópica.

Esto se debe a que solo las componentes horizontales de dichos vectores proporcionan información útil para el proceso estereoscópico, por lo que obtener una medida de distorsión total no refleja necesariamente el desempeño del algoritmo en la conversión, ya que algún código pudiese resultar especialmente útil para detectar movimientos verticales pero con pobre sensibilidad a las componentes horizontales.

Por todo lo anterior se propone tres análisis de desempeño absolutos y relativos, en los cuales se compare la calidad de las detecciones de movimiento horizontal de cada uno de los métodos mencionados *versus* el algoritmo *full search*.

#### 5.3.4.1 Evaluaciones diseñadas

**Análisis visual de desempeño global** Una vez obtenidos los vectores resultantes para una secuencia, se puede graficar cada uno de ellos sobre cada imagen, por lo que se puede apreciar el funcionamiento del detector de movimiento para cada cuadro, aportando información muy útil para la sintonización de los parámetros de sensibilidad de cada método.

Dado que el análisis de movimiento se realiza sobre imágenes de 640 x 480 *pixeles*, es necesario realizar nuevamente una interpolación de la secuencia original, esta vez aplicada también sobre las componentes de crominancia, ya que el tamaño original de las secuencias es de 320 x 240 *pixeles*.

Para realizar esta interpolación se aplica el mismo principio que en la sección 4.2 extendido a las componentes de crominancia.

Mediante esta prueba se pretende ajustar la sensibilidad de cada método a fin de que detecte satisfactoriamente los movimientos de las escenas ignorando texturas o sombras como posibles desplazamientos.

**Análisis analítico de desempeño global** Otra prueba que se propone es el análisis de histogramas para cada cuadro. Las clases de los histogramas son cada uno de los diferentes vectores resultantes (-7 a 7).

De esta forma al realizar un cálculo aproximado del número de bloques que deben permanecer sin movimiento se puede discernir si un método para determinada sensibilidad tiene una respuesta apropiada, con lo que se pueden respaldar matemáticamente los resultados obtenidos mediante el análisis visual.

**Análisis analítico de desempeño relativo** El algoritmo de *full search* es el que genera los vectores de movimiento con mayor exactitud sin embargo debido a su patrón de búsqueda exhaustiva consume una gran cantidad de tiempo, es por ello que es necesario comparar los resultados relativos de los demás métodos a fin de encontrar el que tenga mejor desempeño tanto en calidad de detección como menor tiempo de ejecución utilizando para cada algoritmo la sensibilidad que genera mejor desempeño.

Para ello se propone analizar el porcentaje de movimiento detectado con respecto a *full search* para cada método utilizando la matriz de reglas mostrada en la figura 5.15.

Mediante las reglas anteriores se puede obtener con exactitud el porcentaje de movimiento detectado por cada algoritmo así como el número y tipo de errores que se cometen.

En el siguiente capítulo se muestran los resultados de los tres análisis propuestos.

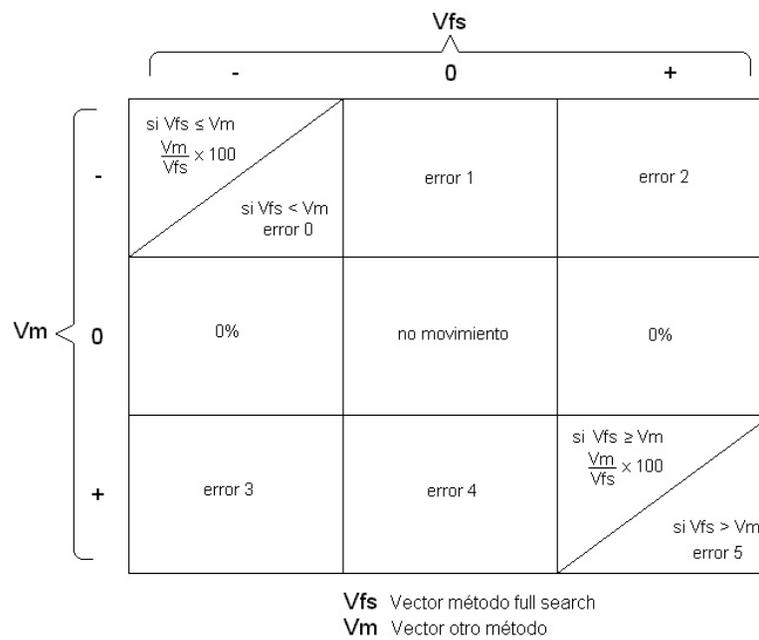


Figura 5.15: Reglas para contabilizar porcentajes de detección de movimiento.



## Capítulo 6

### Resultados de los análisis de movimiento

Dentro de los análisis de desempeño para algoritmos de detección de movimiento el tiempo de ejecución juega un papel muy importante pues actualmente los procesamientos tienden a ser en tiempo real o por lo menos acercarse a ello. Debido a esto se enlistan los tiempos promedio de procesamiento antes de describir los resultados encontrados en cada análisis.

Método	Tiempo <sup>1</sup>
FS	6.33
3SS	0.93
4SS	0.78
N3SS	0.90
BBGDS	0.55
HS	0.82
CS	0.80
PDFS	1.06
UCBDS	0.72

Tabla 6.1: Tiempos de ejecución.

#### 6.1 Resultados de las evaluaciones propuestas

A continuación se muestran los resultados obtenidos por todos los métodos en las tres pruebas propuestas, para fundamentar la clasificación global de los desempeños de detección en la sección 6.2

##### 6.1.1 Análisis visual de desempeño global

Como se comentó en el capítulo anterior mediante esta prueba se analizó el comportamiento de cada método sobre una misma secuencia de video utilizando diferentes sensibilidades.

En la tabla 6.2 se muestra el nivel aproximado de sensibilidad que ofrece mejores resultados para cada método.

Como se puede apreciar ninguno de los métodos obtuvo su mejor desempeño al 100% de su sensibilidad ya que todos confundían las texturas de los objetos con “aparente” movimiento arrojando información incorrecta.

---

<sup>1</sup>seg/cuadro

Método	Sensibilidad
FS	60%
3SS	70%
4SS	70%
N3SS	70%
BBGDS	80%
HS	80%
CS	80%
PDFS	70%
UCBDS	80%

Tabla 6.2: Sensibilidades con mejores desempeños visuales.

Es indispensable sintonizar de manera adecuada cada método para que a una determinada sensibilidad su funcionamiento tienda a uniformizar las direcciones y magnitudes de los vectores obtenidos, por lo que el cálculo de la velocidad del movimiento de la escena tenderá a ser de una forma continua a lo largo de una secuencia, situación que es deseable como se verá en la sección 7.3

En las figuras 6.1, 6.2, 6.3 se aprecian los vectores resultantes obtenidos para un mismo cuadro a distintas sensibilidades utilizando el método de full search.

### 6.1.2 Análisis analítico de desempeño global

Dado que el parámetro que se trata de obtener mediante el detector de movimiento es solamente un vector resultante de movimiento en su componente horizontal para cada cuadro, los resultados del análisis visual pueden resultar muy confiables sin embargo mediante este análisis se demostró de manera analítica los resultados de desempeño *versus* sensibilidad.

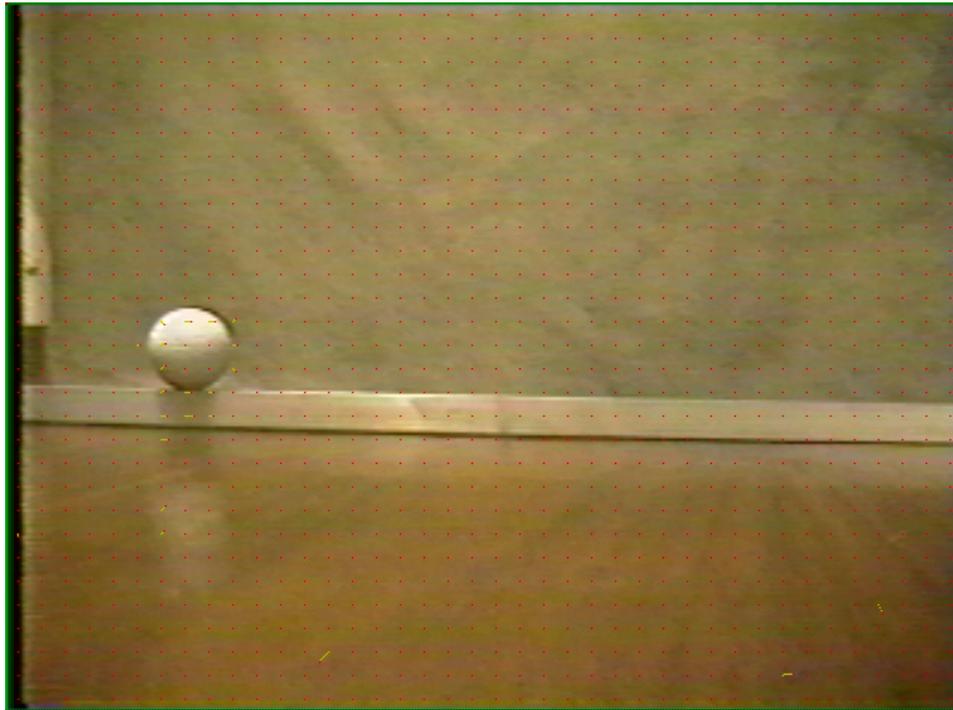
En la figura 6.4 se muestra el comportamiento para el método de full search de los histogramas obtenidos para las diferentes sensibilidades así como el histograma “base” que constituye el resultado ideal esperado de acuerdo al cálculo de bloques que permanecen estáticos.

En este caso se calculó para el cuadro #55 de la secuencia golfnegR que el número aproximado de bloques en movimiento era de 40 y que las magnitudes de los vectores oscilaban entre 1 y 3 unidades. En base a esto se trazó el histograma esperado y se comparó contra los histogramas generados para las 6 diferentes sensibilidades, observándose que a un factor de sensibilidad de 70% se obtenía la mejor respuesta del algoritmo. En las gráficas utilizadas para este análisis se realizó un escalamiento de  $\frac{1}{6}$  para la clase 0 (sin movimiento) para evitar que la escala de la gráfica fuese demasiado grande, ya que los bloques sin movimiento son predominantes para la secuencia analizada.

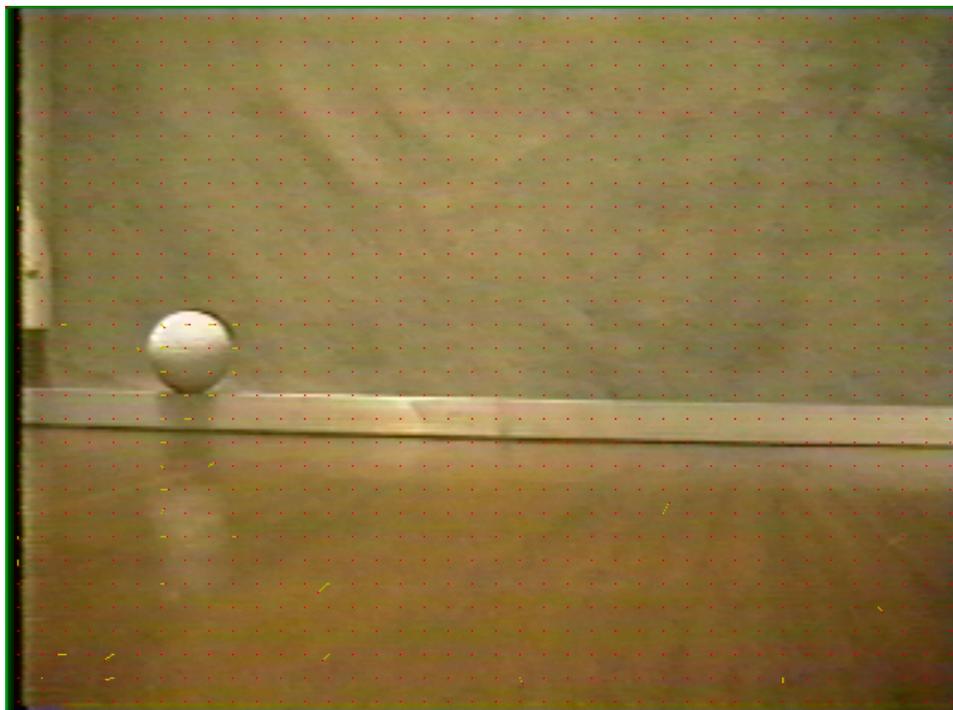
De manera similar se procedió con los demás métodos obteniendo su parámetro de sintonización el cual, tal como se esperaba es muy similar a los calculados por inspección visual, presentándose una variación máxima de  $\pm 10\%$

### 6.1.3 Análisis analítico de desempeño relativo

Mediante este análisis se obtuvo la gráfica 6.5 que muestra la relación entre los porcentajes de detección de los diversos métodos implementados. Los porcentajes mostrados son los valores promedios

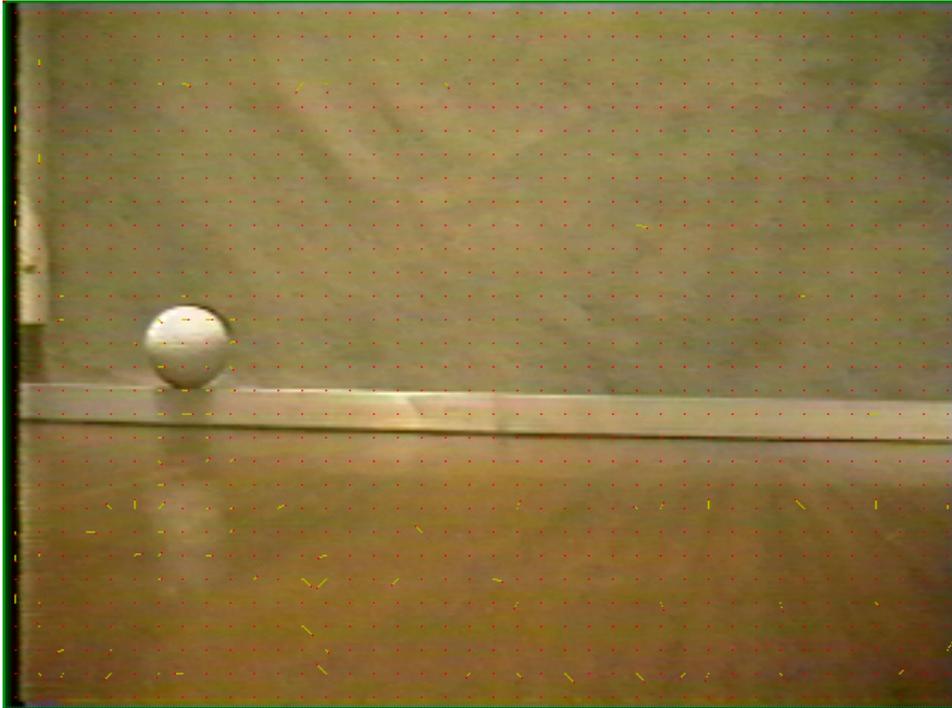


(a) 50%

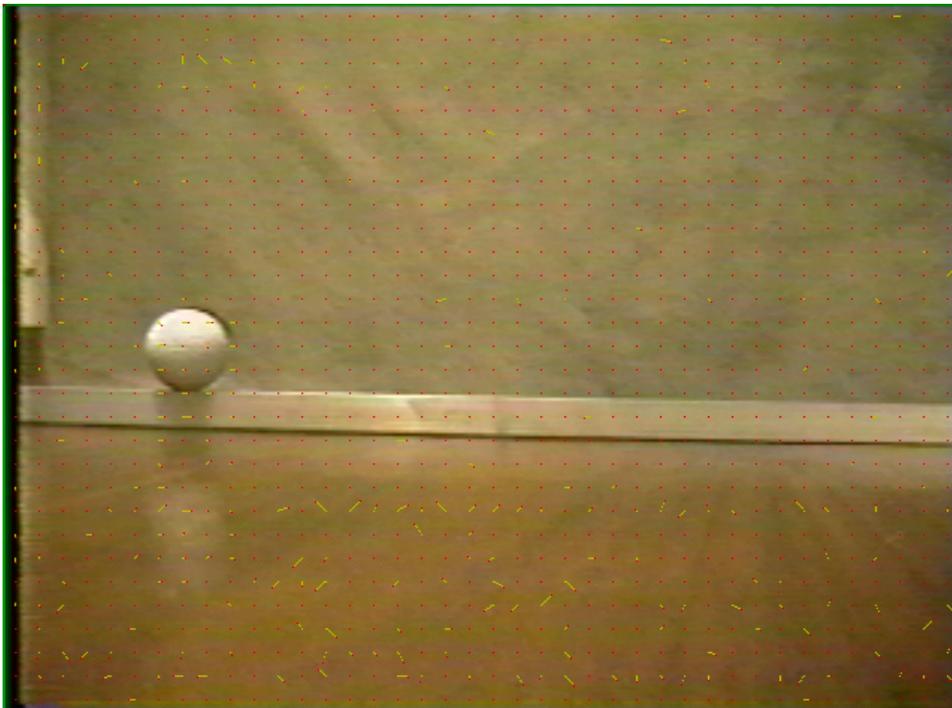


(b) 60%

Figura 6.1: Desempeño visual de *full search* a diferentes sensibilidades (1).

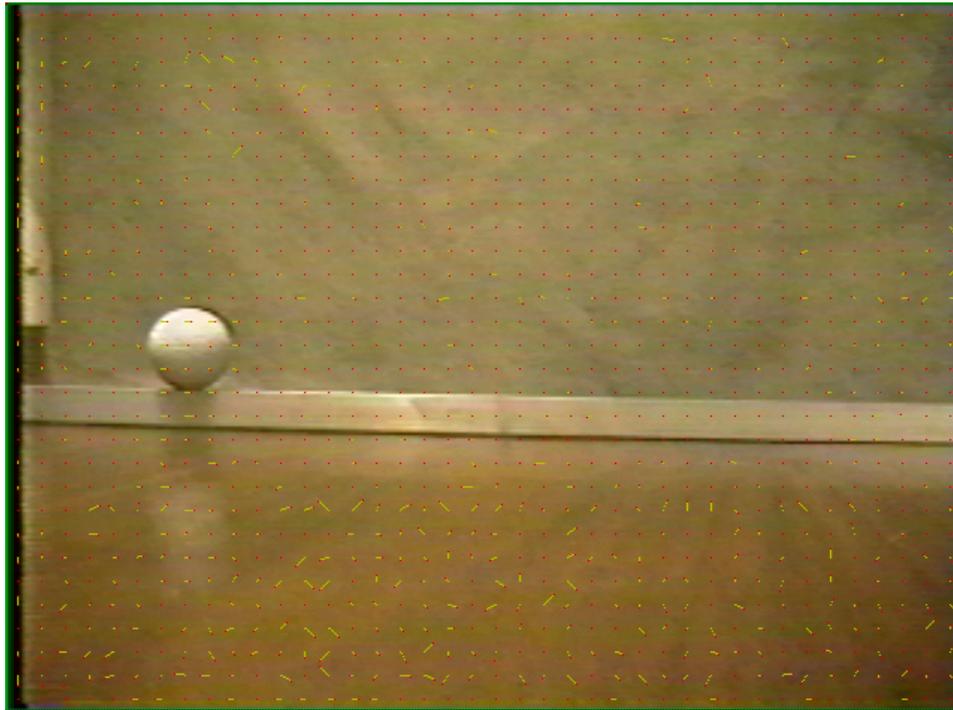


(a) 70%

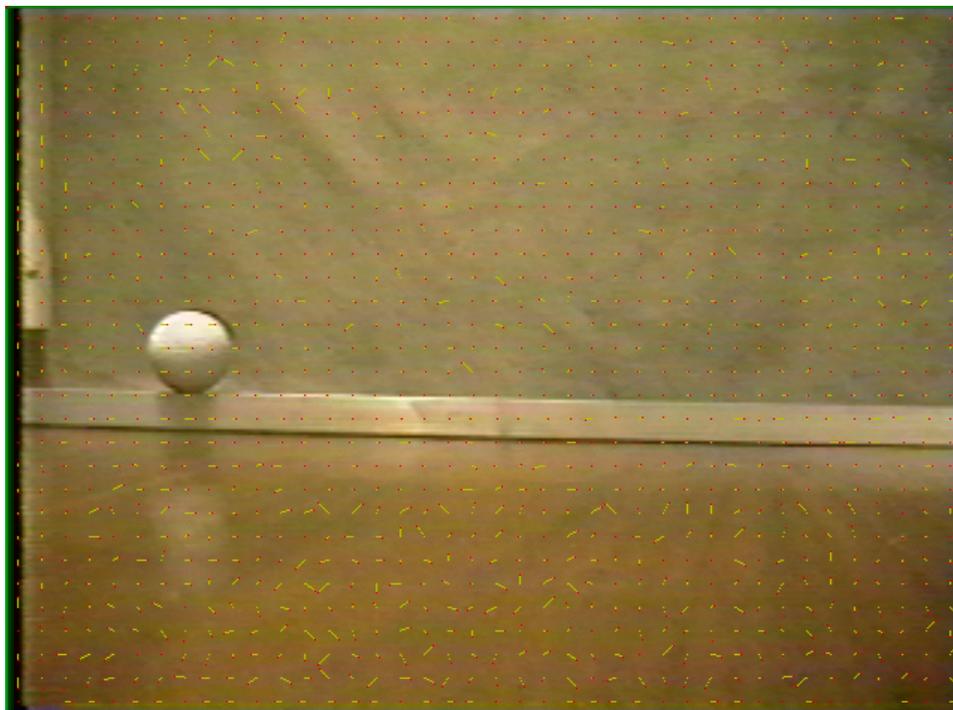


(b) 80%

Figura 6.2: Desempeño visual de *full search* a diferentes sensibilidades (2).



(a) 90%



(b) 100%

Figura 6.3: Desempeño visual de *full search* a diferentes sensibilidades (3).

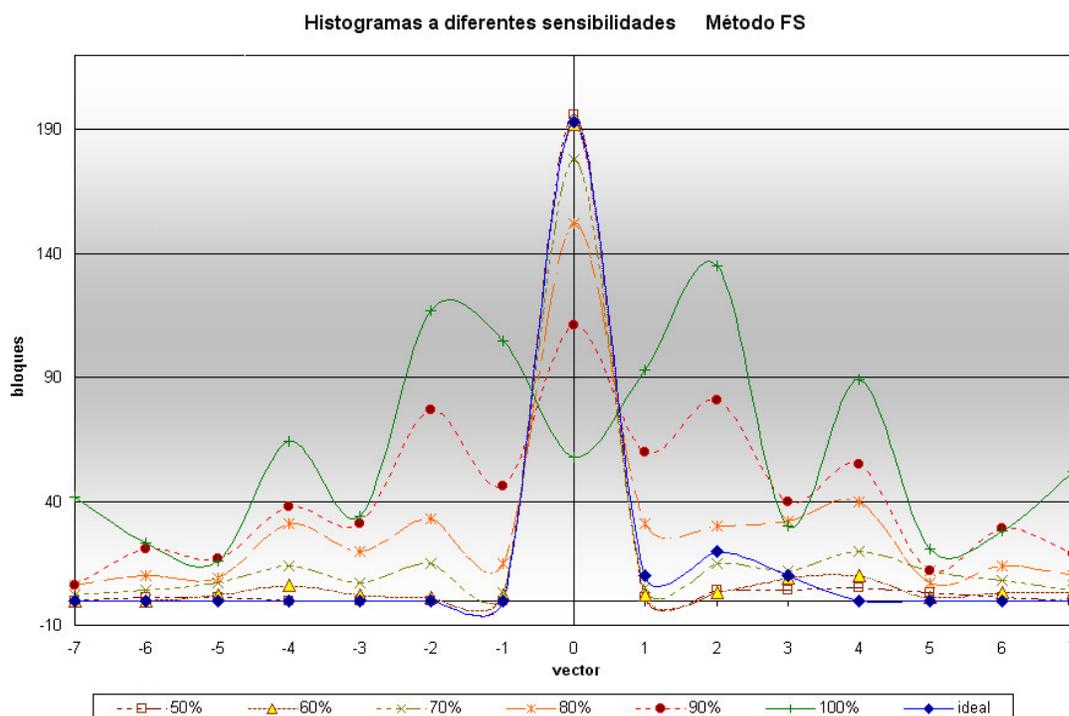


Figura 6.4: Histogramas a diferentes sensibilidades para FS.

encontrados para las diversas secuencias analizadas<sup>2</sup>. Es importante destacar que para cada método se empleó la sensibilidad a la cual se obtuvieron mejores resultados en las dos pruebas anteriores.

En la gráfica 6.5 se considera la penalización de los errores cometidos por cada método de acuerdo a las reglas mencionadas en la sección 5.3.4.1, por lo que los porcentajes de detección son “puros”, es decir en ningún caso se consideró la detección si ésta era errónea en cuanto a sentido o si su magnitud era mayor de la existente.

Como se indicó este análisis se aplicó a 10 secuencias de diferentes longitudes. El desempeño promedio de los métodos para su mejor sensibilidad se puede resumir en la tabla 6.4.

## 6.2 Clasificación final de desempeños de detección

Basándose en la información suministrada por los resultados de las pruebas anteriores y considerando los tiempos promedios de ejecución de cada método, es posible jerarquizar los algoritmos de acuerdo a la tabla 6.5.

Se puede observar claramente que existe una relación entre exactitud y tiempo de ejecución, en donde los algoritmos con mayor velocidad de procesamiento arrojan resultados con menor inexactitud. Para la jerarquización anterior se consideró la exactitud como prioridad esto debido a que los tiempos de ejecución en promedio distan mucho de los necesarios para procesamiento en tiempo real es decir

<sup>2</sup>Para mayor información sobre las características de las secuencias consultar el apéndice A

<sup>3</sup>seg/cuadro

Método	Sensibilidad	Variación
FS	70%	10%
3SS	70%	0%
4SS	70%	0%
N3SS	70%	10%
BBGDS	80%	0%
HS	80%	0%
CS	70%	10%
PDFS	70%	0%
UCBDS	80%	0%

Tabla 6.3: Sensibilidades con mejores desempeños analíticos.

Método	Porcentaje
FS	100%
3SS	35%
N3SS	34%
PDFS	30%
CS	19%
HS	18%
4SS	17%
UCBDS	15%
BBGDS	10%

Tabla 6.4: Porcentajes de movimiento detectado.

Puesto	Método	Exactitud	Tiempo de ejecución <sup>3</sup>
1	3SS	35%	0.93
2	N3SS	34%	0.90
3	PDFS	30%	1.06
4	CS	19%	0.80
5	HS	18%	0.82
6	4SS	17%	0.78
7	UCBDS	15%	0.72
8	BBGDS	10%	0.55

Tabla 6.5: Clasificación de acuerdo a desempeño.

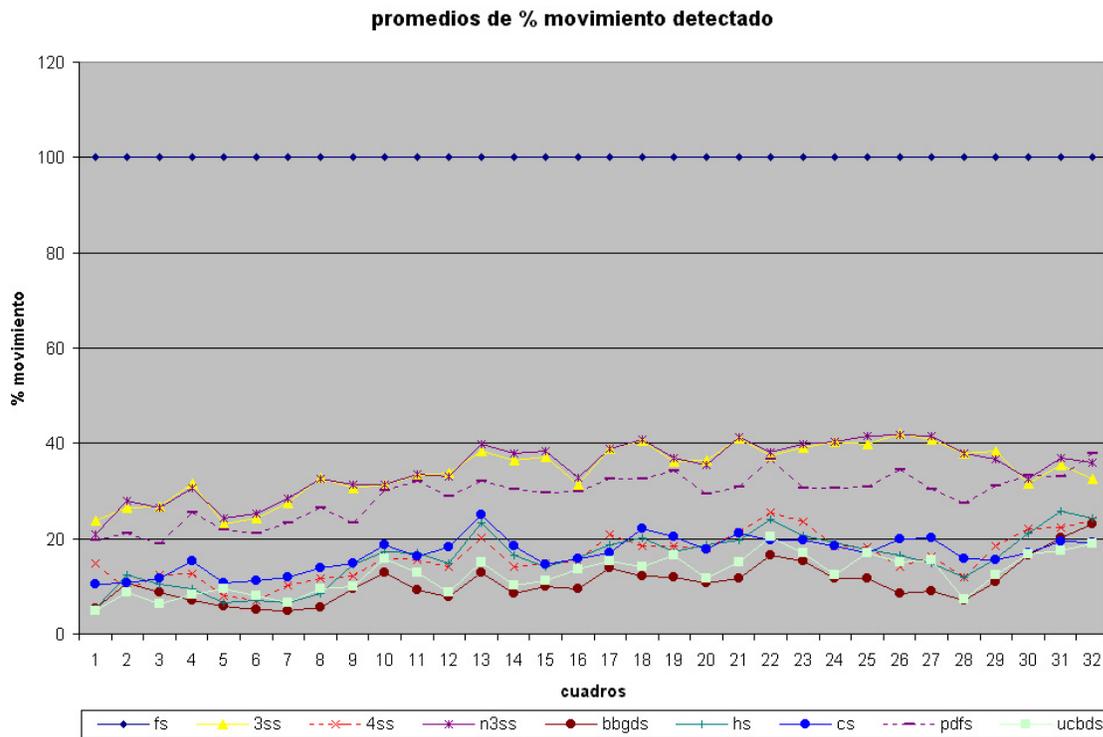


Figura 6.5: Porcentajes de movimiento detectado.

unos 0.033 seg por cuadro<sup>4</sup> como máximo de tal forma que se pudieran desplegar 30 cuadros por segundo.

Cabe destacar que se esperaba que el método CS obtuviera un muy buen desempeño para de esta manera aprovechar sus características de continuidad entre vectores, sin embargo el resultado obtenido no fue del todo satisfactorio ya que por su misma naturaleza cuando se presenta un error de detección este tiende a ser propagado a los bloques vecinos, obteniéndose en algunos casos movimientos inexistentes o en sentido opuesto al esperado.

Así mismo es importante señalar que el método PDFS se podría considerar sin duda como una de las primeras opciones para el proceso de conversión estéreo-artificial, ya que aunado a su desempeño y tiempo de ejecución, posee un alto grado de independencia de los tipos de movimiento que se puedan encontrar en las secuencias, a diferencia de los métodos 3SS y N3SS que en ocasiones la calidad de su detección se encuentra en función de las características del movimiento (3SS para movimientos grandes y N3SS para movimientos muy pequeños o grandes, por lo que fallan cuando se presentan movimientos de magnitud regular, situación que no se manifiesta con PDFS).

<sup>4</sup>Este tiempo es sin considerar el preprocesamiento ni el tiempo necesario para la conversión estéreo-artificial.

## Capítulo 7

### Generación de la imagen estéreo-artificial

Una vez que se determina la velocidad y el sentido del movimiento en la escena es necesario obtener la pareja correspondiente al cuadro en cuestión, a fin de formar un par imágenes con paralaje binocular.

En las sección 7.1 se analizará a detalle la interpretación de los parámetros obtenidos mediante el análisis de movimiento para determinar el retraso necesario entre cuadros.

#### 7.1 Relación entre la dirección de movimiento detectado y la conversión estéreo-artificial

Como se mencionó anteriormente es posible obtener una imagen estereoscópica a partir de una secuencia bidimensional si el cuadro actual es presentado a un ojo mientras que un cuadro retrasado<sup>1</sup> es mostrado al otro ojo.

Para obtener una conversión estéreo-artificial satisfactoria la principal cuestión radica en identificar la dirección y la magnitud del retraso. Por dirección de retraso se entiende el orden en que los cuadros serán desplegados a los ojos, es decir si el cuadro actual será mostrado al derecho o al izquierdo.

Primeramente es necesario identificar que se pueden presentar dos tipos predominantes de secuencias en cuanto a la técnica en que fueron grabadas: secuencias con cámara fija y secuencias con cámara en movimiento.

Es indispensable identificar lo anterior puesto que el tipo de paralaje (positivo o negativo) puede ser determinado de manera errónea si solo se considera la dirección del movimiento sin tomar en cuenta su naturaleza, obteniendo un efecto estereoscópico de baja calidad.

##### 7.1.1 Cámara fija

Se denominan secuencias con cámara fija, aquellas cuyo fondo tiene a permanecer estático y el movimiento detectado es producido por objetos que se desplazan frente a dicho fondo.

Una vez determinada la naturaleza de la secuencia como de cámara fija es necesario interpretar la dirección del movimiento detectado para obtener la dirección del retraso.

###### 7.1.1.1 Dirección del movimiento detectado

En la figura 7.1 se ilustran cuadros de una secuencia bidimensional en donde se muestra una persona y un perro desplazándose de izquierda a derecha sobre un fondo sin movimiento.

---

<sup>1</sup>Por cuadro *retrasado* se entiende cualquier cuadro anterior al actual.

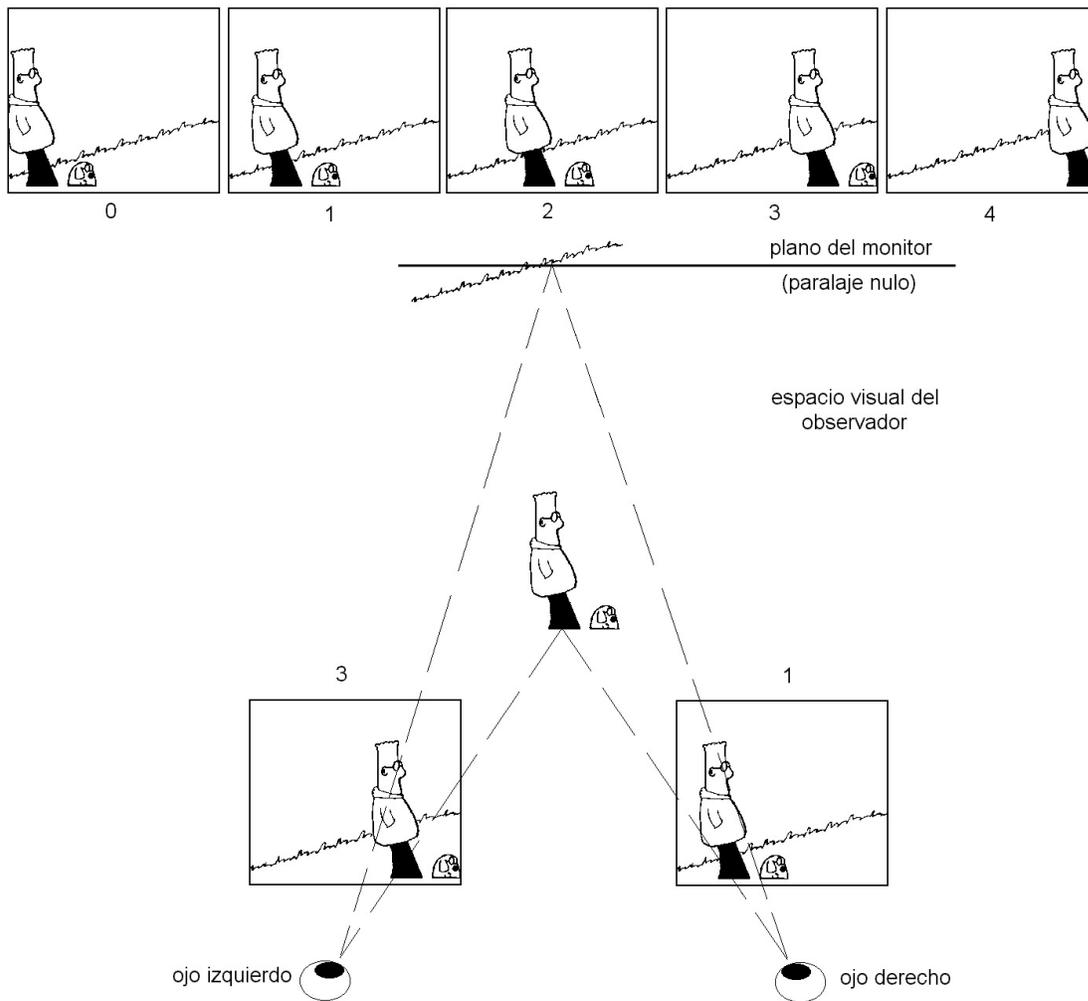


Figura 7.1: Movimiento de izquierda a derecha con paralaje negativo.

Si se identifica al cuadro 3 como el cuadro *actual* se reconocen dos posibilidades en cuanto a la dirección del retraso:

- Si el cuadro actual es presentado al ojo derecho y un cuadro retrasado al ojo izquierdo se obtiene como resultado un paralaje positivo, por lo que el objeto se percibe en el espacio del monitor.
- Por el contrario si el cuadro actual se presenta al ojo izquierdo y el retrasado al derecho se obtiene un paralaje negativo, por lo que el objeto se percibe en el espacio del observador.

En la figura 7.1 se utiliza la última opción por lo que se percibe a la persona y al perro en el espacio visual del observador.

En cambio si el resultado del detector de movimiento establece que la dirección del movimiento en la escena es de derecha a izquierda (figura 7.2), se obtienen paralajes opuestos para los casos del movimiento de izquierda a derecha, es decir:

- Si el cuadro actual es presentado al ojo derecho y un cuadro retrasado al ojo izquierdo se obtiene como resultado un paralaje negativo, por lo que el objeto se percibe en el espacio del observador.
- Si el cuadro actual se presenta al ojo izquierdo y el retrasado al derecho se obtiene un paralaje positivo, por lo que el objeto se percibe en el espacio del monitor.

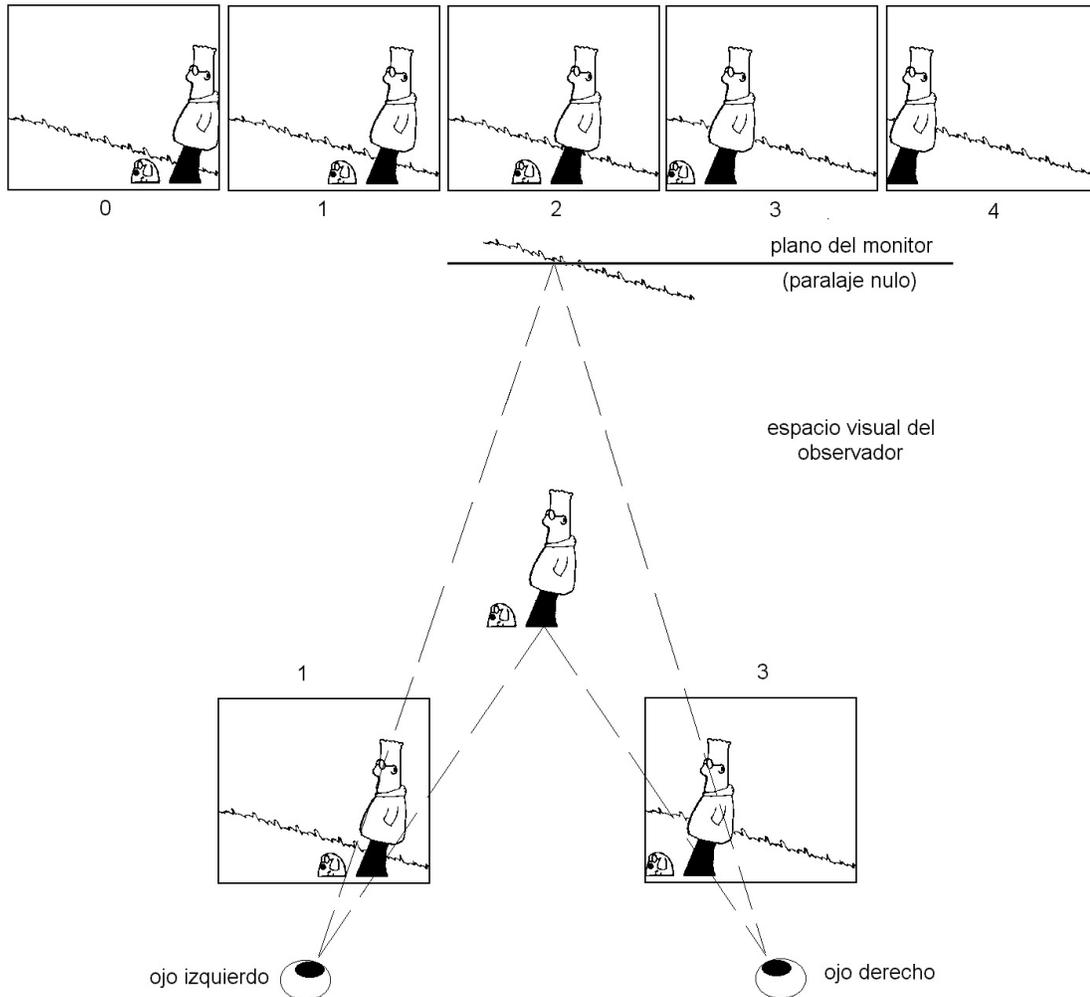


Figura 7.2: Movimiento de derecha a izquierda con paralaje negativo.

En la figura 7.2 se utiliza la opción de paralaje negativo así mismo, por lo que al igual que en el caso anterior, se percibe a la persona y al perro en el espacio visual del observador.

En la tabla 7.1 se muestra el tipo de paralaje obtenido para cada caso mencionado.

### 7.1.2 Cámara en movimiento

Se denominan secuencias con cámara en movimiento, aquellas cuyo fondo tiende a encontrarse en continuo movimiento puesto que la cámara se desplaza a la par de un objeto de la escena conservando

	Ojo Derecho - Cuadro Actual	Ojo Derecho - Cuadro Retrasado
Movimiento I - D	Positivo	Negativo
Movimiento D - I	Negativo	Positivo

Tabla 7.1: Paralajes obtenidos de acuerdo a la relación entre las direcciones de movimiento y retraso para secuencias con cámara fija.

constante su posición relativa, por lo que aparentemente el objeto no se desplaza, o sencillamente el objeto frontal se encuentra estático mientras que los objetos pertenecientes al fondo se desplazan.

En la figura 7.3 se muestra una secuencia con fondo con movimiento y se utiliza la opción de paralaje positivo, por lo que se percibe a la persona y al perro en el espacio visual del monitor.

En la tabla 7.2 se muestra el tipo de paralaje obtenido para cada caso.

	Ojo Derecho - Cuadro Actual	Ojo Derecho - Cuadro Retrasado
Movimiento I - D	Positivo	Negativo
Movimiento D - I	Negativo	Positivo

Tabla 7.2: Paralajes obtenidos de acuerdo a la relación entre las direcciones de movimiento y retraso para secuencias con cámara movimiento.

### Paralaje positivo *versus* paralaje negativo

Como se puede observar los tipos de paralaje obtenidos para cada caso son exactamente los mismos que para escenas con cámara fija, sin embargo al generar imágenes con paralaje negativo para el caso de cámara en movimiento se genera una paradoja visual ya que el paralaje negativo implica que los objetos deben encontrarse en el espacio visual del observador es decir entre el monitor y la persona sin embargo los objetos en la escena bidimensional se encuentran en el fondo de la escena por lo que resulta imposible que un objeto del fondo se encuentre en primer plano. A esta condición se le puede denominar como imágenes de *falso paralaje* ya que a pesar de que pueden ser fusionadas por el cerebro humano no proporcionan una percepción correcta de la profundidad por lo que el efecto estereoscópico resulta incómodo e irreal.

En la figura 7.3 se muestra el caso de movimiento en el fondo, en la cual es necesario utilizar paralaje positivo a diferencia de la figura 7.1 en la que a pesar de ser el mismo movimiento se utilizó paralaje negativo. Como se mencionó es necesario aplicar el paralaje positivo a fin de evitar generar un falso paralaje.

Normalmente el efecto estereoscópico de objetos en el espacio visual del observador tiene mayor impacto que el de espacio del monitor, por lo que de alguna manera resulta deseable manipular los parámetros obtenidos mediante el detector de movimiento para obtener dicho efecto; sin embargo dado que el detector de movimiento no arroja información sobre si el objeto en movimiento se encuentra en primer plano o en el fondo de la escena no es recomendable utilizar paralajes negativos puesto que puede presentarse paralaje falso cuando el movimiento detectado es de un objeto que se encuentra ubicado en el fondo de la escena.

Por lo tanto de acuerdo a lo anterior el proceso de conversión estéreo-artificial ofrece una mejor calidad en su modalidad de espacio visual de monitor (paralaje positivo) ya que a pesar de no ser tan

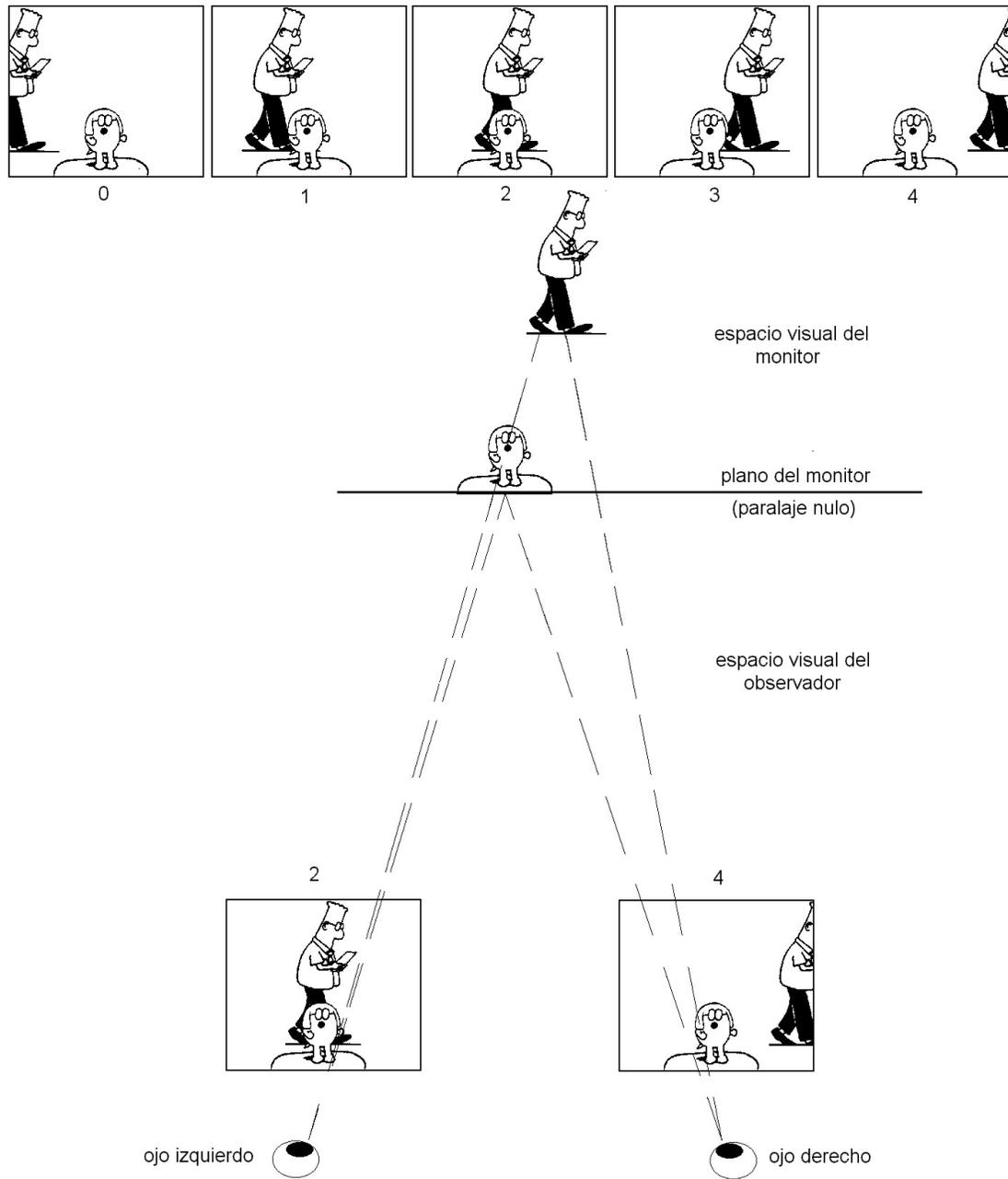


Figura 7.3: Movimiento de izquierda a derecha en el fondo con paralaje positivo.

espectacular como el otro, esta desventaja es aminorada por su constancia en cuanto a la coherencia de sus resultados al no mostrar imágenes con profundidades irreales.

En la figura 7.4 se muestra claramente como al utilizar diferentes paralajes los objetos se encuentran en diferentes espacios visuales.

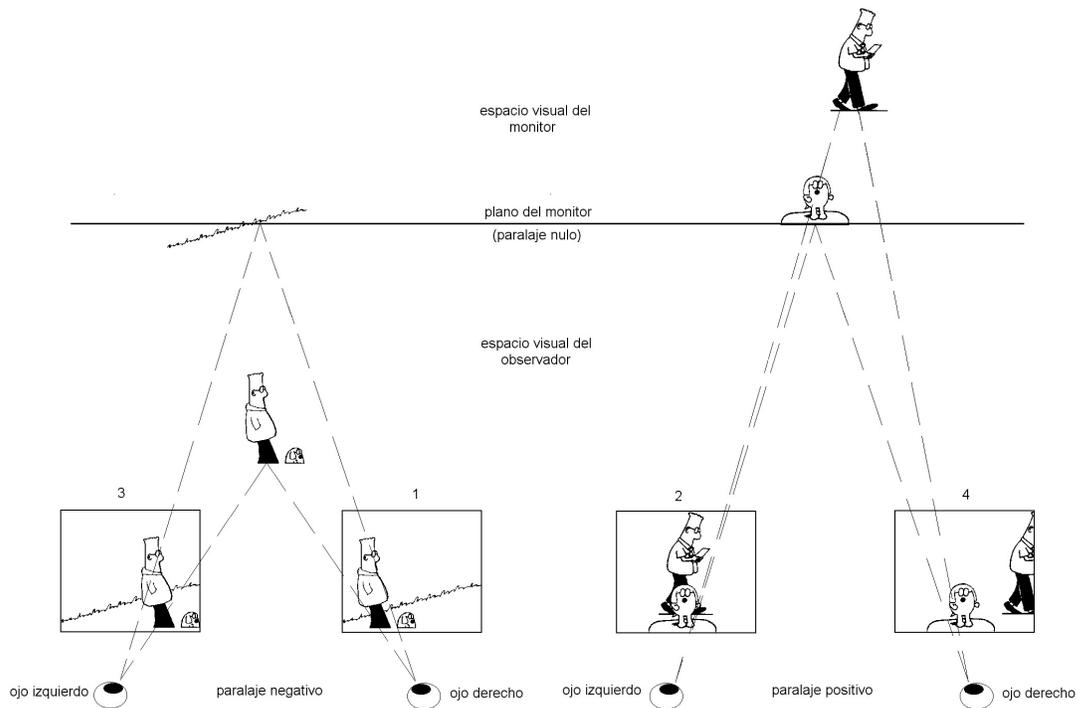


Figura 7.4: Espacios visuales.

## 7.2 Determinación del retraso entre cuadros

Uno de los aspectos fundamentales de la conversión estéreo-artificial lo constituye el retraso entre cuadros pues es en función de este valor, que aspectos básicos de la estereoscopia tales como el máximo paralaje y los límites de la región estereoscópica útil quedan definidos.

El detector de movimiento como se mencionó en el capítulo anterior divide cada cuadro de imagen en bloques de 16 *pixeles* de arista por lo que para cuadros de 640 x 480 *pixeles* se obtienen 1200 bloques, es decir, 1200 vectores.

Dado que cada vector representa el movimiento de una pequeña porción de la imagen total no resulta útil o acertado calcular un vector resultante como la suma de los 1200 vectores obtenidos. En su defecto es mejor generar un histograma en donde cada clase corresponde a un número entero que representa a la componente horizontal del vector. Dado que el rango de valores para los vectores es de -7 a 7 se obtienen 15 clases.

Una vez generado el histograma se calculan las tres clases cuya frecuencia es mayor y mediante la siguiente ecuación se determina el retraso entre cuadros:

$$\text{retraso} = |hm(0)| \quad (7.1)$$

donde:

$hm(0)$  Clase con mayor frecuencia.

El valor del retraso entre cuadros es un parámetro de gran importancia en la conversión ya que determina de manera directa el valor del paralaje que se presenta al observador y como se mencionó en el capítulo 2 es un aspecto primordial en la calidad del efecto estereoscópico ya que debe ser menor al valor de máximo paralaje permisible de tal manera que el observador pueda fusionar ambas perspectivas para percibir la profundidad de la escena.

### 7.3 Continuidad de los parámetros de conversión durante la secuencia

La continuidad es una propiedad que merece especial atención en secuencias de video, ya que es a partir de ella que un observador infiere movimiento y coherencia entre una serie de imágenes.

Si no existe continuidad tanto en la dirección como en la magnitud del retraso se presentan dos problemas:

1. Si la dirección del paralaje se encuentra oscilando entre positiva y negativa debido a “errores” generados por el detector de movimiento, el efecto estereoscópico resulta inaceptable pues los objetos parecen brincar entre el espacio visual del observador y del monitor, producto de paralajes falsos.
2. Si la magnitud del retraso tiene cambios instantáneos consecutivos, los objetos parecen oscilar o brincar dentro de uno de los espacios visuales, provocando que el observador perciba una sensación de mareo.

Para prevenir los problemas anteriores es necesario aplicar heurísticas que generen una histéresis en la conversión.

#### 7.3.1 Histéresis en la dirección del retraso

La metodología utilizada para evitar cambios instantáneos consecutivos en cuanto al tipo de paralaje empleado consiste en almacenar los últimos  $n$  resultados generados en función del análisis del detector de movimiento, y antes de aplicar el resultado actual ( $m$ ) se verifica que concuerde con al menos el 75% de los datos guardados, de otra manera se aplica una corrección que consiste en invertir la dirección del retraso.

En la figura 7.5 se muestra un diagrama de flujo ilustrando lo anteriormente mencionado.

#### 7.3.2 Histéresis en la magnitud del retraso

Para evitar cambios abruptos en la magnitud del retraso se emplea una metodología similar a la de histéresis en dirección, la cual consiste en almacenar los valores de los últimos  $n$  retrasos aplicados de tal forma que la diferencia entre el retraso aplicado en el cuadro actual  $m$  y el promedio del historial no sea mayor a  $k$  unidades, conservando la continuidad entre los valores de retraso a lo largo del tiempo.

En la figura 7.6 se ilustra lo anterior.

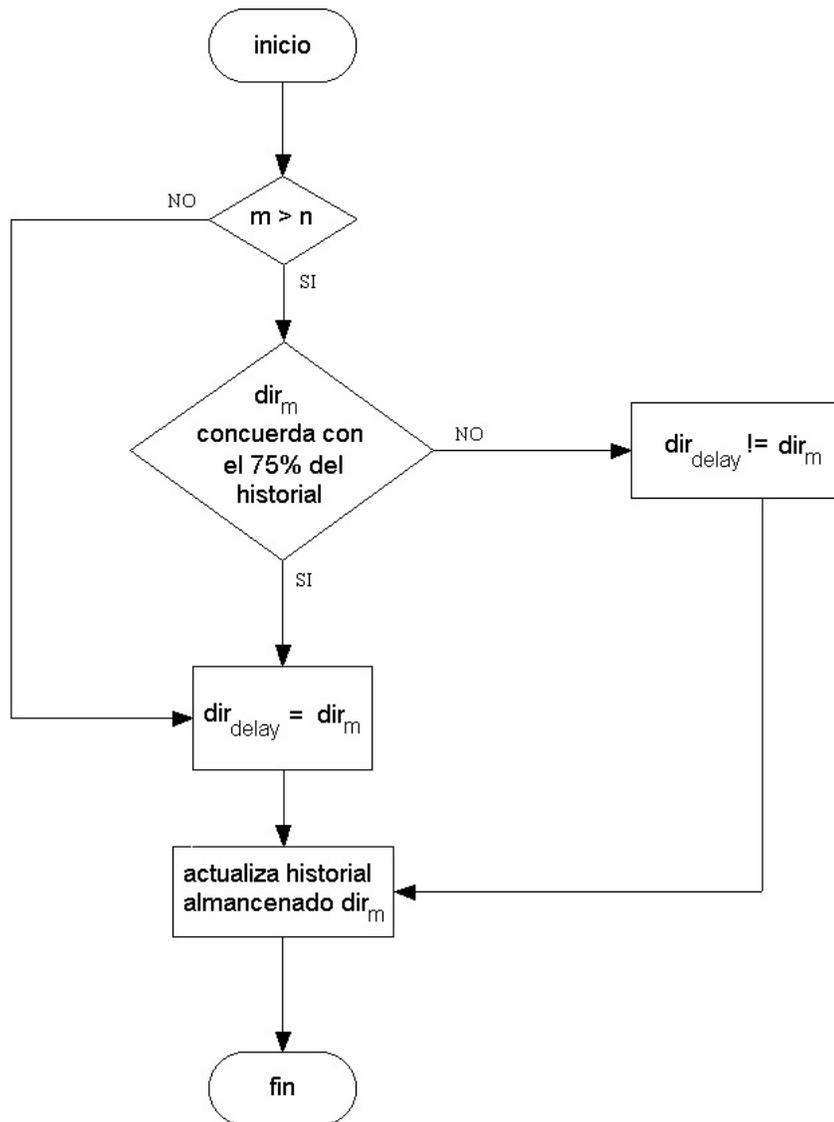


Figura 7.5: Histéresis en dirección.

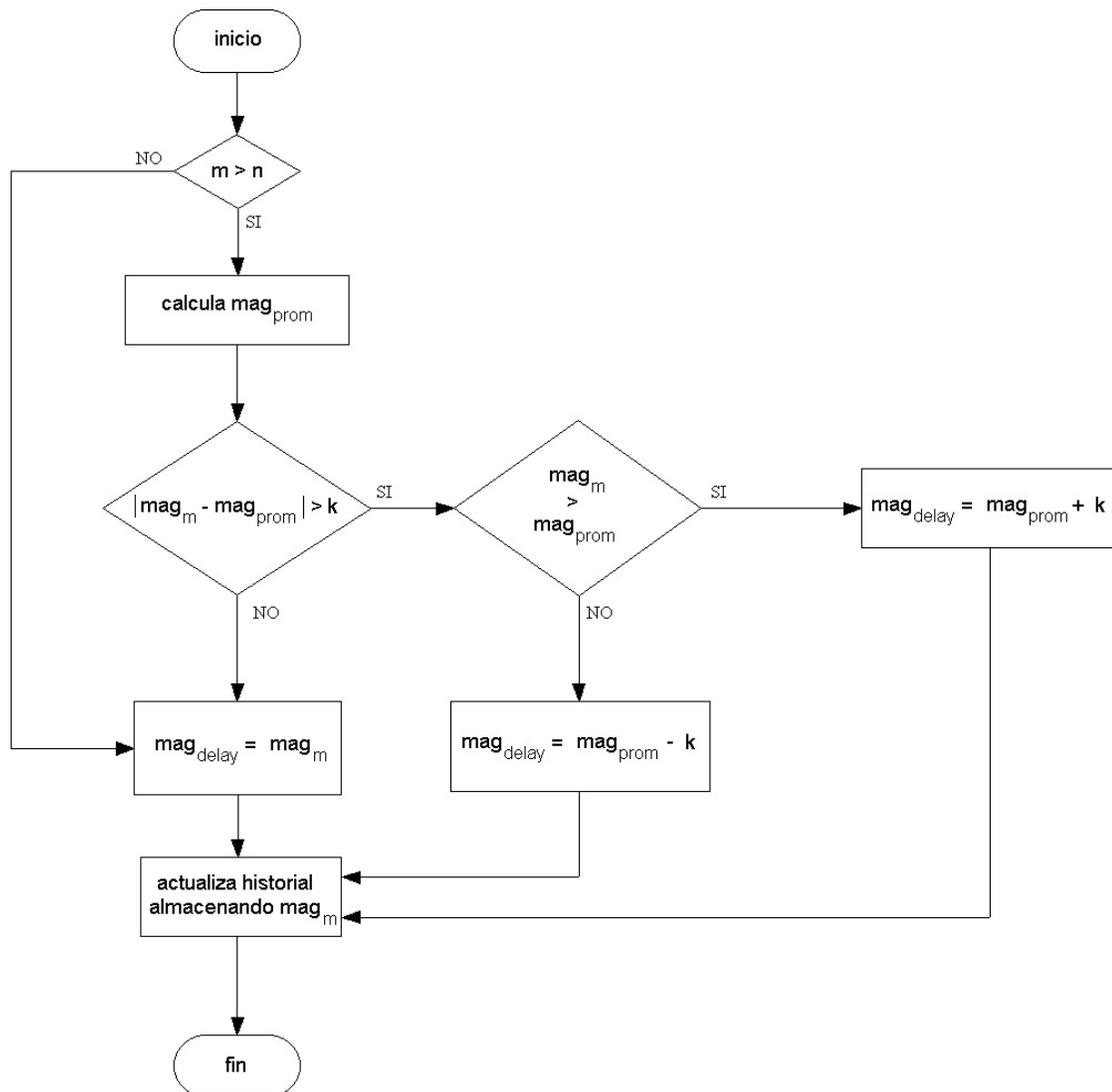


Figura 7.6: Histéresis en magnitud.

## 7.4 Caso especial: Escenas sin movimiento

La metodología propuesta y desarrollada en la presente investigación se centra en la información de movimiento que puede ser obtenida, por lo que cuando se detecta que una secuencia no posee movimiento se presenta un caso especial. Al no existir paralaje de movimiento resulta imposible generar disparidades virtuales mediante retrasos entre cuadros. Como solución alterna a este problema se puede tratar de crear disparidades espaciales mediante corrimientos para cada perspectiva de acuerdo a la siguiente tabla:

	Imagen ojo derecho	Imagen ojo izquierdo
Paralaje positivo	Corrimiento a la derecha	Corrimiento a la izquierda
Paralaje negativo	Corrimiento a la izquierda	Corrimiento a la derecha

Tabla 7.3: Reglas para la creación de paralajes en secuencias sin movimiento

Los corrimientos propuestos son del orden de 2% o 3% del tamaño horizontal de la imagen. Sin embargo dado que el corrimiento es constante para toda la escena el efecto estereoscópico resulta apreciable solamente cuando se tiene un único objeto frente a un fondo sin texturas u otros objetos, obteniéndose una efecto de profundidad pero convirtiendo el objeto en una imagen “acartonada” es decir que se puede distinguir con una profundidad menor que el fondo pero como si fuera un objeto de dos dimensiones en un ambiente tridimensional.

Así mismo estos corrimientos pueden ser utilizados para aumentar el paralaje en secuencias de imágenes con movimientos lentos, con la desventaja de que el corrimiento provoca que se pierda una franja vertical de la imagen equivalente al doble del corrimiento.

## Capítulo 8

### Métodos de despliegue estereoscópico

En el capítulo 2 se comentaron los aspectos básicos de la estereoscopía en cuanto a su naturaleza y su utilización. En las siguientes secciones se describirá brevemente las diferentes técnicas de despliegue estereoscópico que existen y en la sección 8.3 se mostrará a mayor detalle el método de separación de colores que por su simplicidad y su bajo costo fue el seleccionado para ser utilizado en esta investigación.

#### 8.1 Clasificación de métodos para despliegue estereoscópico

Existen numerosas técnicas para despliegue estereoscópico, las cuales pueden ser clasificadas de acuerdo a varios criterios:

La clasificación más simple se basa en el tipo de imágenes que utilizan:

- **3D real:** Despliegan un patrón de rayos de luz en un espacio volumétrico en el mundo real.
- **3D verdadero:** Utilizan dos conjuntos de imágenes bidimensionales, uno para cada ojo.
- **3D no-verdadero:** Utilizan una sola imagen bidimensional para ambos ojos.

De lo anterior los métodos con mejores resultados son los de 3D real y 3D verdadero los cuales a su vez se pueden clasificar en:

- **Autoestereoscópicos:** No requieren de la utilización de lentes especiales o algún otro elemento óptico para ver el efecto 3D.
- **Estereoscópicos:** Al contrario de los autoestereoscópicos una parte fundamental son lentes u otros medios ópticos para la visualización 3D.

#### 8.2 Técnicas para despliegue estereoscópico

Como se mencionó anteriormente existen numerosas tecnologías empleadas para desplegar imágenes en formato 3D. Entre las principales se puede mencionar a las siguientes:

- Holográfica.
- Volumétrica.
- Basada en paralaje.
- Lenticular.
- De separación visual.

### 8.2.1 Holográfica

Esta técnica utiliza patrones de interferencia luminosa a escalas microscópicas para capturar profundidades de escenas. Un holograma es creado combinando dos haces de luz monocromática <sup>1</sup> de una misma fuente, uno proveniente del reflejo del objeto en una escena y el otro, utilizado como referencia, proveniente directamente de la fuente de luz. El patrón de interferencia entre los haces puede ser grabado en un material fotosensible.

Para observar un holograma basta con iluminar el material en que fue plasmado con la misma fuente de luz que fue aplicada al momento de la grabación.

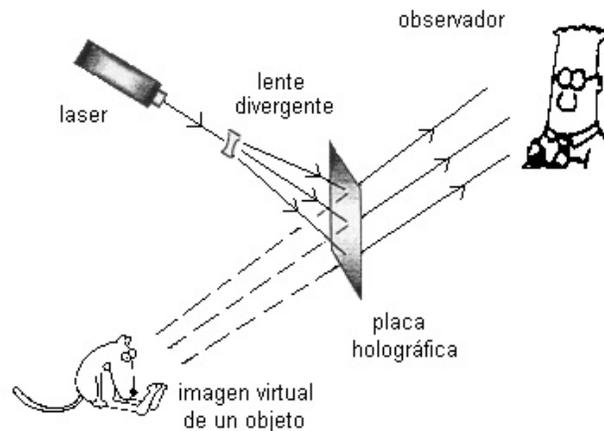


Figura 8.1: Holografía.

El concepto de holografía puede ser aplicado a televisión mediante la utilización de transductores especiales sin embargo tiene grandes desventajas en cuanto a los requerimientos que implica este método:

1. Un cuarto oscuro libre de partículas (polvo) o vibraciones.
2. Un material flexible y borrable para guardar y reconstruir imágenes holográficas.
3. Un ancho de banda enorme para su transmisión: Una imagen de 16 pulgadas cuadradas requiere 100,000 veces el ancho de banda utilizado por TV actualmente.

Las características de esta técnica se pueden resumir como:

- 3D real.
- Autoestereoscópica.
- La escena cambia de acuerdo al punto de vista del observador, permitiendo así mismo paralaje de movimiento.

<sup>1</sup>P.ej. laser de helio-neón, el cual produce una luz de color rojo oscuro.

- Ortoestereoscópico: Reproduce distancias ópticas reales sin deformaciones.
- Monocromático.
- No realizable en tiempo real.
- Requerimientos y condiciones de visualización muy especiales.

### 8.2.2 Volumétrica

Los sistemas volumétricos multiplanares despliegan imágenes compuestas por un gran número de planos, ocupando un espacio real.

Un ejemplo de este tipo de despliegue es el espejo varifocal, el cual se compone de un espejo circular vibrador que es controlado por un *woofer* en sincronía con un monitor.

El espejo cambia su largo focal de acuerdo a la vibraciones recibidas por el woofer de tal forma que las imágenes desplegadas por el monitor parecen encontrarse en diferentes espacios físicos dando la impresión de profundidad. El principio de este sistema puede observarse en la figura 8.2.

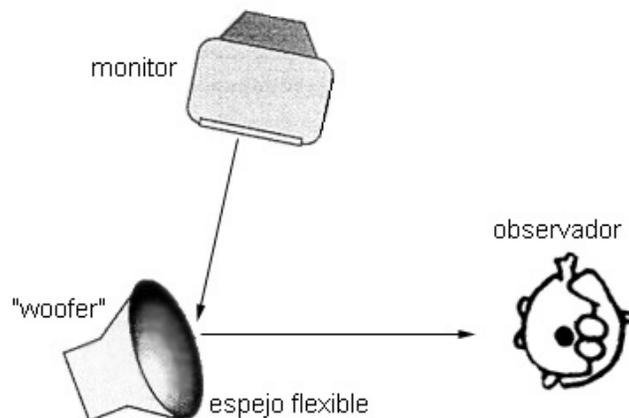


Figura 8.2: Espejo varifocal.

Las características generales de esta técnica son:

- 3D real.
- Autoestereoscópica.
- La escena cambia de acuerdo al punto de vista del observador, permitiendo así mismo paralaje de movimiento.
- Ortoestereoscópico: Reproduce distancias ópticas reales sin deformaciones.
- Generalmente monocromático, con capacidad para despliegue de figuras o estructuras sencillas.
- Tiempo real.

### 8.2.3 Basada en paralaje

Los sistemas basados en paralaje utilizan una barrera de paralaje la cual *filtra* partes de una imagen de acuerdo a su posición.

La imagen que se despliega se compone de franjas verticales que contienen de manera alternante las perspectivas para cada ojo. Entre esta imagen y el observador se coloca una barrera de paralaje que puede ser cristal líquido (LCD), de tal forma que las columnas de la barrera direccionan hacia cada ojo la perspectiva que le corresponde generando el efecto estereoscópico. (figura 8.3).

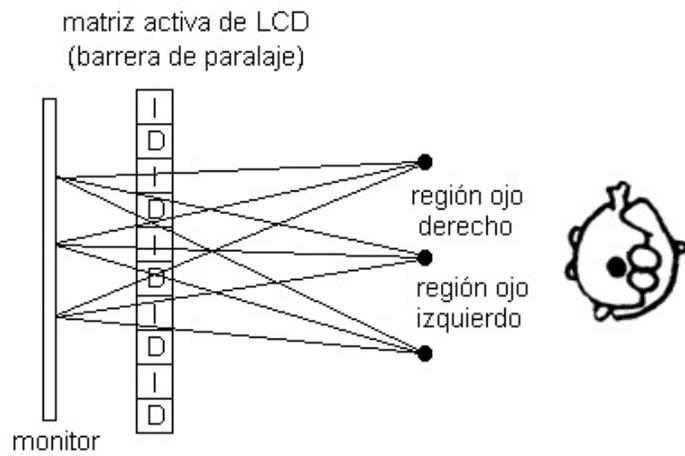


Figura 8.3: Barrera de paralaje.

Entre las desventajas de este sistema se tiene que resulta una imagen de apariencia oscura puesto que la barrera de paralaje tiende a bloquear una cierta cantidad de luz. Así mismo se puede obtener cierta distorsión debido principalmente a que ciertos ángulos los ojos del espectador puede observar porciones de perspectiva que no le corresponden.

Las características generales de esta técnica son:

- 3D verdadero.
- Autoestereoscópica.
- Limitación en cambios de escena de acuerdo la posición del observador, restringiendo así mismo el paralaje de movimiento.
- No ortoestereoscópico: Reproduce distancias ópticas reales con cierto grado de deformación
- A colores pero con imágenes oscuras.
- Tiempo real.

### 8.2.4 Lenticular

Este tipo de técnica es muy similar al basado en paralaje ya que la imagen fuente también contiene ambas perspectivas multiplexadas espacialmente. Sin embargo a diferencia del método de barrera utiliza lentes cilíndricas permitiendo desplegar hasta cuatro pares de perspectivas de una escena convirtiéndose en un sistema multiusuario. Así mismo no genera imágenes oscuras ya que no filtra la luz sino que la redirecciona. (figura 8.4)

Sin embargo una desventaja de este sistema es que la distancia relativa entre las lentes y las franjas de las imágenes debe ser muy exacta ya que de otra manera el observador puede observar más de una perspectiva en cada ojo dando como resultado un efecto 3D deformado.

- 3D verdadero.
- Autoestereoscópica aunque la escena vista depende de la posición del observador.
- Limitación en cambios de escena de acuerdo a posición del observador, restringiendo así mismo el paralaje de movimiento.
- No ortoestereoscópico: Reproduce distancias ópticas reales con cierto grado de deformación
- A colores.
- Tiempo real.

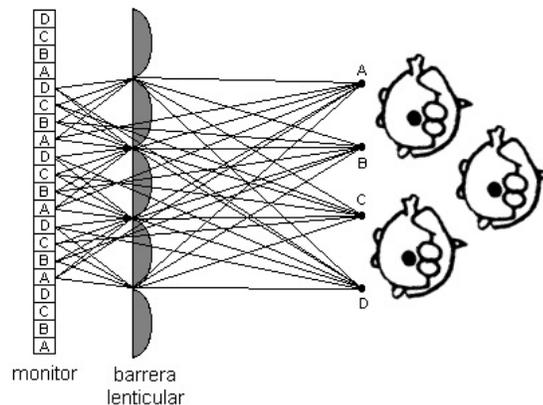


Figura 8.4: Sistema lenticular.

### 8.2.5 De separación visual

Para este tipo de sistemas se utiliza dos conjuntos de imágenes bidimensionales conteniendo cada uno una perspectiva para cada ojo.

Al observar cada ojo la imagen que le corresponde el cerebro humano fusiona las dos perspectivas creando una sensación de profundidad para la persona.

Una característica principal de esta técnica es que el usuario debe utilizar algún tipo especial de lentes.

### 8.2.5.1 Separación pasiva

Este tipo de sistema utiliza imágenes polarizadas, es decir la luz es filtrada de tal forma que solo cierta posición del campo electromagnético es permitida. Para realizar esta tarea se emplea un lente polarizador, el cual tiene “cortes” a nivel molecular, haciendo que el campo electromagnético de la luz que lo atraviesa solo oscile en determinada dirección en vez de varias.

Para desplegar imágenes estereoscópicas es necesario utilizar una polarización diferente para cada ojo. Esto se logra polarizando independientemente cada fuente de luz como se observa en la figura 8.5. Así mismo el usuario debe usar los filtros correspondientes en unos lentes de tal suerte que cada ojo solo perciba las imágenes que le corresponden.

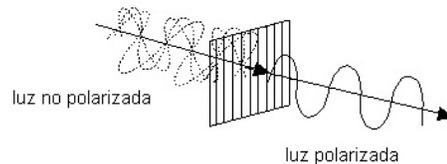


Figura 8.5: Polarización.

Los tipos de polarización que se emplean comercialmente son lineal y circular. La polarización lineal presenta pérdidas de efecto estereoscópico cuando el observar inclina su cabeza de tal suerte que deja de percibir las imágenes ya que cambia el sentido del filtraje, en cambio la polarización circular elimina este problema pues los filtrajes son independientes de la posición de los lentes.

### 8.2.5.2 Separación activa

Esta técnica utiliza lentes activos de cristal líquido en sincronía con las imágenes desplegadas en un monitor. Los lentes de LCD actúan como obturadores alternantes de forma que un ojo es bloqueado mientras se presenta el punto de vista correspondiente al otro y viceversa. (figura 8.7)

Este tipo de separación se llama activa pues los lentes contienen electrónica para realizar las obturaciones a diferencia de los filtrajes de polarización que no emplean más que principios ópticos.

Los lentes de LCD pueden ser alámbricos o inalámbricos, en este último caso es necesario utilizar un emisor infrarrojo para mantener en sincronía a los obturadores con las imágenes presentadas.

### 8.2.5.3 Separación activo-pasiva

Este método como su nombre lo menciona es una combinación de las dos técnicas anteriores. El usuario utiliza lentes polarizadores pasivos mientras que se coloca una placa polarizadora alternante frente al monitor la cual cambia de sentido de polarización en sincronía con las imágenes desplegadas.

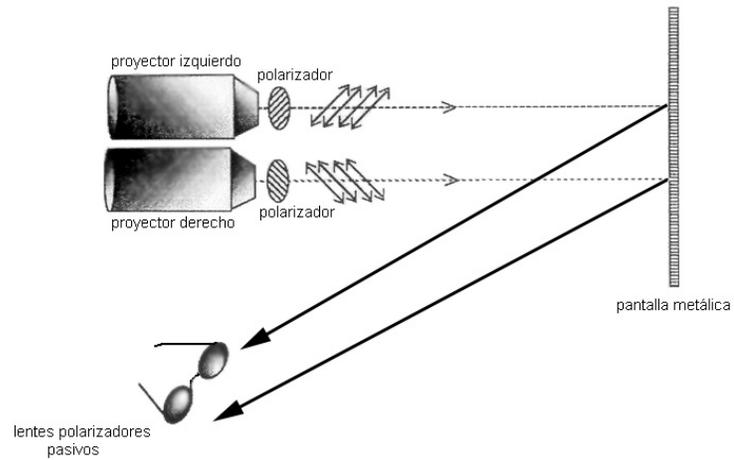


Figura 8.6: Separación pasiva.

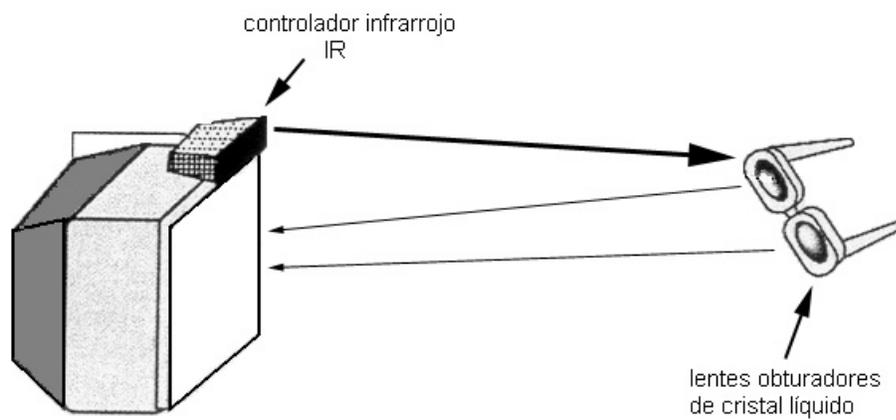


Figura 8.7: Separación activa.

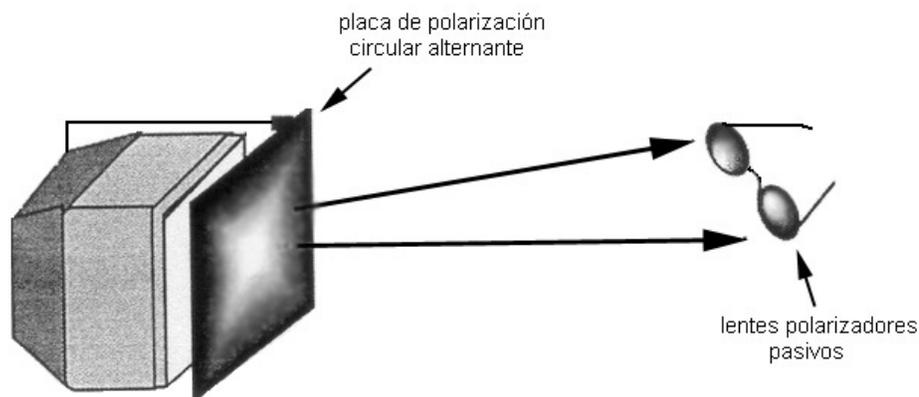


Figura 8.8: Separación activo-pasiva.

#### 8.2.5.4 Separación por colores

De todas las técnicas mencionadas anteriormente la separación por colores es una de las pioneras en el mundo de la estereoscopia, así mismo constituye una de las más simples y económicas por lo que los efectos estereoscópicos generados no son de excelente calidad.

Este método utiliza lentes de colores azul y rojo o verde y rojo; normalmente se le conoce como método anaglífico.

Al igual que todos los métodos de separación, utiliza dos conjuntos de imágenes. Las imágenes para el ojo izquierdo se encuentran “pintadas” de rojo mientras que las imágenes para el ojo derecho son de color azul. El par de lentes con un filtro rojo en el ojo izquierdo y uno azul en el ojo derecho se utilizan para enmascarar al color opuesto del otro ojo.

La elección de estos colores proviene del hecho de que sus frecuencias se encuentran muy separadas en el espectro por lo que el costo de los filtros es muy bajo. Sin embargo el filtrado no es perfecto por lo que existe lo que se conoce como *cross-talk* o interferencia entre ambos puntos de vista aminorando el efecto estereoscópico.

### 8.3 Implementación del método de separación por colores

Al momento de realizar esta investigación el autor tenía acceso a utilizar un sistema de separación activa que desarrolló en años anteriores, sin embargo dado que la tarjeta de digitalización utilizada no posee salida de video analógico fue necesario implementar un método económico para observar los resultados obtenidos por el algoritmo de transformación estéreo-artificial.

Como se mencionó en el capítulo 7 las perspectivas de ambos ojos se encuentran en cuadros intercalados a lo largo de la secuencia, de tal forma que los cuadros pares contienen la información del ojo derecho y los impares los del izquierdo. Estas imágenes originalmente se encontraban a colores por lo que fue necesario “teñirlas” del color que les correspondía.

El proceso de tinte consiste en eliminar las componentes de crominancias verde y azul para los cuadros pares a fin de obtener una imagen en “rojos”; mientras que para los cuadros impares se

elimina la componente roja conservando la información de color verde y azul, dando como resultado una imagen de tonalidades azules.

Este proceso no puede ser directamente aplicado a los archivos en formato UYVY pues como se recordará del capítulo 4 no se tiene acceso directo a las componentes de color por lo que es necesario convertir la información al espacio de color RGB (red-green-blue).

Para ello se utilizan las siguientes ecuaciones:

$$R = 1.164(Y - 16) + 1.793Cr - 128 \quad (8.1)$$

$$G = 1.164(Y - 16) - 0.534Cr - 128 - 0.213(Cb - 128) \quad (8.2)$$

$$B = 1.164(Y - 16) + 2.115(Cb - 128) \quad (8.3)$$

Estas ecuaciones implican que cada pixel tiene su correspondiente muestra de luminancia y de las diferencias de crominancia. Sin embargo dado que el espacio UYVY es del tipo 4:2:2 solo los *pixeles* pares tienen toda esta información por lo que fue necesario realizar una interpolación inicial para las componentes de color a fin de obtener un muestreo 4:4:4 así mismo se aprovechó este paso para convertir de formato empaquetado a planar de forma que cada componente pudiese ser manipulada en arreglos independientes.

Una vez que la información se encuentra en el espacio de color RGB mediante las siguientes ecuaciones es posible suprimir las bandas de colores necesarias regresando al espacio YUV 4:4:4

Para los cuadros pares (ojo derecho):

$$Y = 0.183R + 16 \quad (8.4)$$

$$U = -0.101R + 128 \quad (8.5)$$

$$V = 0.439R + 128 \quad (8.6)$$

Para los cuadros impares (ojo izquierdo):

$$Y = 0.614G + 0.062B + 16 \quad (8.7)$$

$$U = -0.338G + 0.439B + 128 \quad (8.8)$$

$$V = -0.399G - 0.040B + 128 \quad (8.9)$$

Una vez aplicadas las ecuaciones anteriores sobre los arreglos planares es necesario realizar la conversión a 4:2:2 en forma empaquetada para ello solo se agrupan los *pixeles* de acuerdo a la figura 4.1 mostrada en el capítulo 4.

Finalmente es necesario entrelazar cada par de cuadros a fin de formar una sola imagen conteniendo ambos puntos de vista. Esto debe realizarse debido principalmente a que el algoritmo explicado en el capítulo 7 duplica el tamaño de las secuencias puesto que incorpora un segundo conjunto de imágenes a la secuencia original por lo que si el *frame rate* no se duplica así mismo los videos resultantes se observarán a la mitad de la velocidad original. Así mismo resulta indispensable este entrelazado

porque al desplegar secuencialmente cuadros de diferentes colores (rojo, azul, rojo . . .) se presenta un parpadeo que puede resultar molesto para el observador, situación que no se presenta al entrelazar los cuadros correspondientes.

El entrelazado se realiza de la misma forma que en televisión las líneas pares contienen los cuadros “rojos” ahora convertidos en campos, mientras que las líneas impares los “azules”.

El proceso completo de “tintado” se puede observar en la figura [8.9](#)

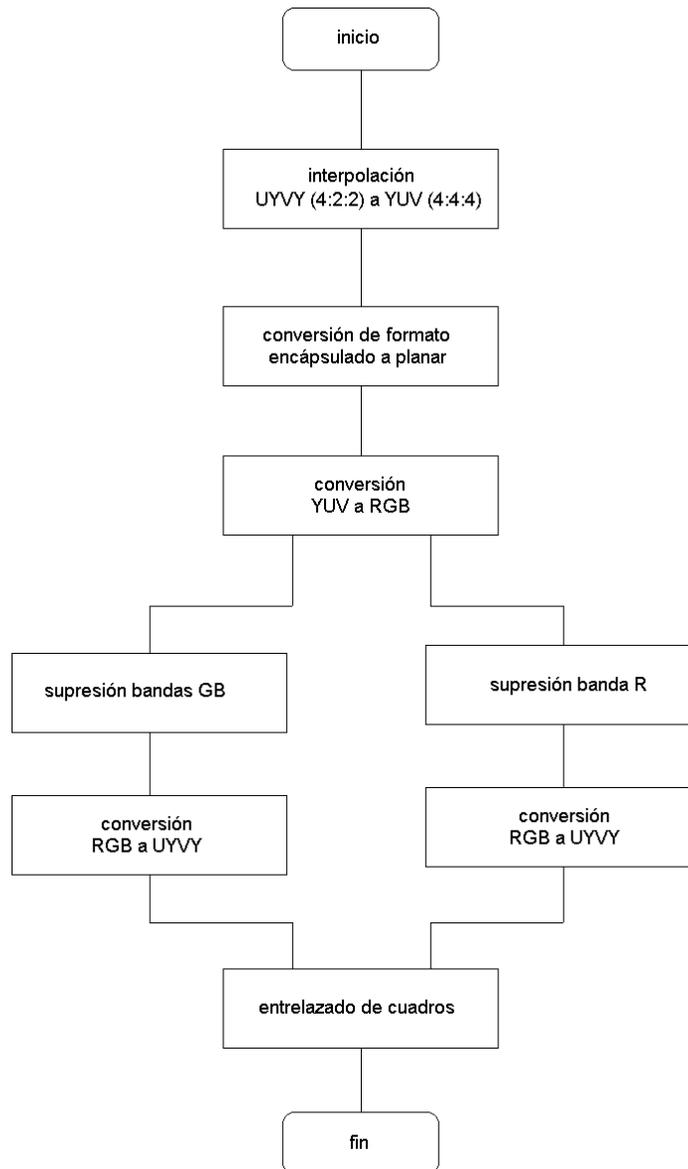


Figura 8.9: Proceso de “tintado”.



## Capítulo 9

### Análisis de resultados de la conversión estéreo-artificial

En este capítulo se analizan los parámetros de estereoscopía obtenidos mediante la conversión estéreo-artificial, así mismo se discuten aspectos relevantes acerca de las características globales del efecto estereoscópico obtenido mediante la conversión versus secuencias estéreo-naturales.

#### 9.1 Análisis de los parámetros críticos en estereoscopía

Como se mencionó en el capítulo 2 existen 4 parámetros críticos para determinar la calidad de un imagen estereoscópica.

##### 9.1.1 Máximo paralaje obtenido

El máximo paralaje obtenido mediante la conversión se encuentra directamente ligado al retraso entre cuadros de imágenes, el cual a su vez es función de la detección de velocidad y dirección de movimiento.

Para fines de esta investigación se observó que un paralaje de hasta 27 mm a una distancia de observación de 1 m es aún fusionable por el cerebro humano obteniendo un muy buen efecto estereoscópico. Uno de los problemas que puede presentarse es que el movimiento no sea netamente horizontal, es decir que también posea una componente vertical, al ocurrir esto si se trata de sintonizar el convertidor a fin de obtener los máximos paralajes (máximo efecto estereoscópico) se corre el riesgo de que se presenten disparidades verticales las cuales pueden ser fusionadas por el observador pero en un rango mucho menor (hasta un máximo de 10 mm) por lo que el efecto es contraproducente empobreciendo la calidad de la imagen estereoscópica. Por lo tanto es preferible generar paralajes del orden del 75% del máximo obtenible a fin de mantener secuencias de muy buena calidad estereoscópica.

##### 9.1.2 Límites de cercanía y lejanía

Los límites de cercanía y lejanía se encuentran estrechamente ligados a los máximos paralajes obtenidos. Como se mencionó en el capítulo 7 la conversión estéreo-artificial autónoma se restringió al espacio del monitor es decir se fijó como límite de cercanía el plano del monitor. Lo anterior obedece a que el detector de movimiento no es capaz de proporcionar información acerca de si el objeto en movimiento pertenece al fondo de la imagen o a un plano más cercano, por lo que si se trata de obtener que las imágenes salgan del monitor se pueden generar en algunos casos falsos paralajes.

Sin embargo basándose en que se conoce *a priori* las características de la secuencia i3d, se aplicó la conversión generando paralajes negativos. Los resultados obtenidos fueron satisfactorios pues se logró crear la ilusión de que los objetos *salieran* del monitor. Sin embargo se detectó que cuando el objeto en cuestión se encuentra parcialmente en la imagen, es decir que se encuentra *cortado* por

alguna de las orillas del monitor se produce un conflicto visual. Por un lado el paralaje negativo genera la ilusión de que el objeto sale de la pantalla pero por otro lado la interposición de la orilla del monitor indica que el objeto se encuentra detrás del plano del monitor “jalando” el objeto, este conflicto genera un efecto que podría considerarse como *fuera de foco* empobreciendo la estereoscopia de la imagen.

Resumiendo lo anterior, para generar imágenes estereoscópicas en el espacio del observador mediante la conversión estéreo-artificial es indispensable determinar la posición relativa de los objetos en movimiento con respecto al fondo y así mismo determinar si su contorno no se encuentra obstruido por las orillas del monitor. Dado que por sus características, el detector de movimiento basado en bloques no es capaz de dar tal información, para esta investigación se limitó a utilizar el espacio del monitor, sin embargo la utilización del espacio del observador como límite de cercanía se puede proponer como un trabajo de extensión de la presente tesis.

### 9.1.3 Resolución de profundidad y distorsión de despliegue

Se encontró que la resolución de profundidad obtenida mediante la conversión estereoscópica es un parámetro con gran variabilidad. Esto se debe a que la razón de cambio en disparidad se encuentra en función de la velocidad de movimiento detectado y dado que éste no es constante a lo largo de las secuencias o en ocasiones es nulo, no se puede establecer una resolución constante de profundidad. En un sentido estricto de la palabra no se recomienda la utilización de la conversión estéreo-artificial para procesos automatizados en los que, por ejemplo, un brazo mecánico calcule sus movimientos en base a la disparidad generada mediante esta técnica, ya que puede cometer fallos al no contar con un parámetro de disparidad con variaciones mínimas, es decir las profundidades calculadas se pueden encontrar ligeramente distorsionadas de la realidad.

Sin embargo para aplicaciones de entretenimiento y educación o para procesos donde un humano puede compensar de manera inmediata la distorsión de despliegue, la resolución de profundidad obtenida resulta aceptable.

## 9.2 Estéreo-artificial *vs* estéreo-natural

Sin duda una de la evaluaciones de mayor importancia resulta la comparación de una imagen estéreo-artificial *versus* una imagen estéreo-natural obtenida mediante dos cámaras.

El criterio utilizado para comparar ambas imágenes fue la inspección visual. Para ello sin pretender realizar una encuesta exhaustiva o extremadamente representativa se presentaron ambas secuencias a un grupo de 8 personas, de las cuales solo tres habían observado imágenes estereoscópicas anteriormente con cierta regularidad.

Las personas con experiencia previa en estereoscopia lograron acertar en la mayoría de las ocasiones sobre cuál imagen era artificial y cuál la natural, sobretudo cuando los movimientos de las secuencias eran algo complicados generando disparidades verticales. Sin embargo describieron a las secuencias estéreo-artificiales como más “suaves” que las naturales, refiriéndose a que ofrecen mayor comodidad visual al no generar efectos de profundidad exagerados presentes muchas veces en las secuencia grabadas con dos cámaras.

En cuanto a los observadores *novatos* se refirieron en la mayoría de los casos a que ambas secuencias se encontraban en 3D sin comentar importantes diferencias entre ambos tipos de imágenes. Incluso, al igual que los observadores expertos, algunos expresaron sentir menor estrés ocular al contemplar imágenes estéreo-artificiales.

### 9.3 Secuencia de laparoscopia

Desde el comienzo de la investigación se determinó que las secuencias de video generadas mediante laparoscopia son candidatas para la conversión estéreo-artificial pues la mayoría de los movimientos que poseen son horizontales, en algunas ocasiones generadas por desplazamientos de la cámara introducida en el paciente y en otros por los instrumentos médicos al generar suturas, cortes y otras operaciones.

La secuencia que se utilizó corresponde a una laparoscopia para realizar desprendimientos entre el intestino grueso y la matriz de una paciente. Claramente se logra apreciar el volumen de la matriz y de otros órganos, por lo que puede resultar muy útil para ser mostrada posteriormente a los alumnos, para enseñar la técnica operatoria, facilitando el reconocimiento de los diversos órganos al proveer profundidad en las imágenes.

Sería deseable aplicar la conversión estéreo-artificial en una operación, para ello es necesario que el convertidor opere en tiempo real y así mismo es importante que el médico reciba un entrenamiento de acondicionamiento a las imágenes estereoscópicas generadas a fin de que logre compensar las distorsiones de despliegue que pudieran presentarse como se comentó en la sección 9.1.3.



## Capítulo 10

### Conclusiones

La conversión estéreo-artificial constituye una valiosa herramienta para la generación de videos estereoscópicos a partir de imágenes monoculares con movimientos horizontales, ampliando la videoteca estereoscópica en gran medida sobretodo para tipos específicos de videos como lo son: laparoscopías, tomas realizadas desde vehículos en movimiento, partidos de futbol, baloncesto y en general cualquier secuencia con dicho tipo de desplazamiento.

La televisión ha constituido a lo largo de varios años una tecnología que provee la capacidad de transmitir una *realidad visual* a distancia, comunicando al mundo entero al mostrarle los sucesos acontecidos en los más recónditos lugares. Dicha realidad visual tiende a ser cada vez más fiel a las imágenes verdaderas que el hombre percibe diariamente, una muestra de ello lo constituye la televisión de alta definición cuyos principales objetivos radican en transmitir a mayor detalle, sin interferencias visuales y con colores más verídicos imágenes de acontecimientos ocurridos a distancia.

Esta tendencia tecnológica invita a pensar en la inclusión de la estereoscopia como una variable que puede ser adoptada por la televisión comercial en un futuro no muy lejano. En ese entonces resultará más que deseable poder utilizar todo el material monocular que ha sido recopilado a lo largo de estos años, por lo que la conversión estéreo-artificial cobrará gran importancia.

El algoritmo de conversión desarrollado en esta investigación ofrece gran calidad de efecto estereoscópico e inclusive dadas sus características puede ser utilizado para suavizar efectos exagerados en secuencias estéreo-naturales.

El buen desempeño de este algoritmo se basa en su capacidad de generación de oclusión de objetos, ya que esta característica se encuentra presente en todas las imágenes estéreo-naturales.

#### 10.1 Extensión de resultados y futuras investigaciones

Este trabajo de tesis no tuvo contemplado como punto prioritario el generar un algoritmo convertidor en tiempo real, por lo que queda abierta la posibilidad de optimizar e implementar en hardware dedicado el convertidor para tiempo real.

Dado que el mayor tiempo de ejecución se centra en la detección de movimiento, este módulo del algoritmo pudiese ser notablemente mejorado al emplear alguna otra técnica que no sea tan exhaustiva en la evaluación de diversos puntos sobre la imagen.

Así mismo los resultados de esta investigación pueden ser fácilmente extendidos a otros formatos de despliegue de video como lo son: PAL, SECAM, *Above and Below*, etc.

Una mejora sustancial que puede realizarse como complemento a esta investigación es la conversión para imágenes sin movimiento ya que a partir de dichas imágenes es imposible generar un par estereoscópico mediante retrasos entre cuadros. Para ello se propone emplear técnicas de detección de

profundidad basadas en enfoque y desenfoco de objetos, de tal suerte que se pueda complementar la información de un algoritmo de detección de movimiento pudiendo utilizar libremente el espacio del observador para generar efectos estereoscópicos sin producir falsos paralajes.

## Apéndice A

### Secuencias de video

#### A.1 Secuencias estereoscópicas

Para la grabación de las secuencias estereoscópicas *naturales* se emplearon dos videocámaras Sony con las siguientes características:

Perspectiva	izquierda	derecha
Modelo	CCD-TR45	CCD-TR6
Formato	NTSC V8	NTSC V8
Zoom máximo	6x	8x
focus	7-42mm	7-56mm

Tabla A.1: Características las cámaras utilizadas.

Para la grabación de la escenas se utilizó iluminación artificial, proveniente de tres lámparas, dos localizadas en los extremos y la restante al frente de la escena.

En cuanto a los parámetros de las cámaras, se desactivó la opción de *autofocus* y se ajustaron los *zooms* de ambas cámaras al mismo nivel a fin de obtener una imagen estereoscópica de buena calidad.

La distancia entre los lentes de las cámaras era de 9 cms que es la separación mínima que se puede obtener debido al tamaño de las cámaras y se encontraban a una distancia de 85 cms del punto de convergencia, por lo que se contó con suficiente espacio para realizar tomas con paralaje negativo es decir delante del punto de convergencia.

Ambas perspectivas fueron digitalizadas mediante una tarjeta SunVideo Plus y después fueron entrelazadas y procesadas mediante el método de separación de colores desarrollado, a fin de poder ser desplegadas en una estación de trabajo SUN.

Todas las secuencias fueron digitalizadas a 320 x 240 *pixeles* a menos que se especifique otro tamaño.

#### A.2 Secuencias monoscópicas

Para las secuencias monoscópicas simplemente se utilizó una de las perspectivas obtenidas mediante una de las cámaras.

Nombre	# cuadros	Descripción
bola	175	Desplazamiento de una pelota de golf sobre una mesa
golfneg	128	Desplazamiento de una pelota de golf sobre un riel, esta secuencia fue grabada utilizando paralaje negativo
golfpos	120	Desplazamiento de una pelota de golf sobre un riel, esta secuencia fue grabada utilizando paralaje positivo
manoneg	345	Desplazamiento horizontal de una mano, esta secuencia fue grabada utilizando paralaje negativo
manopos	335	Desplazamiento horizontal de una mano, esta secuencia fue grabada utilizando paralaje positivo
carrusel <sup>a</sup>	115	Carrusel girando en sentido en contra de las manecillas del reloj
i3d	315	Animación computarizada de una vista aérea sobre un campo
vidal <sup>b</sup>	200	Personas desplazándose en un parque de diversiones

Tabla A.2: Secuencias estereoscópicas.

<sup>a</sup>Grabada utilizando 640 x 480 *píxeles* a 18 fps

<sup>b</sup>Grabada utilizando 640 x 480 *píxeles* a 19 fps

Nombre	# cuadros
bola	175
golfneg	128
golfpos	120
manoneg	345
manopos	335
carrusel	115
i3d	315
vidal	200

Tabla A.3: Secuencias monoscópicas.

### A.2.1 Secuencias monoscópicas sobre medicina

Para la secuencia de laparoscopia se utilizó material proveído por el Dr. Carlos Edmundo Covarrubias Camacho en formato VHS.

Nombre	# cuadros	Descripción
lap1	211	Animación computarizada sobre el efecto de un medicamento a niveles microscópicos dentro del cuerpo humano
lap2	298	Operación utilizando laparoscopia para desprender una matriz de un intestino

Tabla A.4: Secuencias laparoscopia.

### A.2.2 Otras secuencias monoscópicas utilizadas

Las secuencias originalmente utilizadas en esta investigación fueron digitalizadas a 640 x 480 *pixeles* sin embargo dado que el frame rate (fps) obtenido era menor a 30 cuadros por segundo, se decidió utilizarlas solamente para complementar los análisis de desempeño de los algoritmos de detección de movimiento. Algunas de ellas fueron utilizadas para pruebas en los procesos de conversión estéreo-artificial.

Nombre	# cuadros	fps	Descripción
golf1	84	19	Desplazamiento lento de una pelota de golf sobre una superficie plana
golf2	56	18	Desplazamiento rápido de izquierda a derecha de una pelota de golf sobre una superficie plana
mano_hor	141	18	Movimiento horizontal de una mano cerrada
mano1	323	18	Movimientos de diferentes tipos de una mano cerrada
carro1	295	18	Desplazamiento lento de un carrito de juguete sobre una superficie plana
carro2	102	19	Desplazamiento rápido de un carrito de juguete sobre una superficie plana
pesa1	200	19	Desplazamiento de una pesa de gimnasia rodando sobre una superficie plana
pesa2	547	19	Desplazamiento de una pesa de gimnasia mediante una mano.
frijol1	270	18	Desplazamiento de la cámara al enfocar una maceta con frijoles

Tabla A.5: Otras secuencias monoscópicas de prueba.



## Apéndice B

### Glosario

**área fusionable de Panum** Región restrictiva sobre la cual deben ubicarse los puntos homólogos de una imagen estereoscópica para poder ser fusionados por el cerebro.

**convergencia** Rotación de los ojos en dirección horizontal produciendo fusión.

***crosstalk*** Aislamiento incompleto de los canales con las perspectivas izquierda y derecha.

**disparidad retinal** Distancia entre puntos homólogos entre las perspectivas izquierda y derecha en las retinas del observador. Su parte análoga en la pantalla de despliegue lo constituye el paralaje.

**distancia interocular** Distancia promedio entre los ojos humanos.

**espacio visual del monitor** Región estereoscópica comprendida detrás de la superficie del monitor. Las imágenes con paralaje positivo se perciben en el espacio del monitor.

**espacio visual del observador** Región estereoscópica comprendida entre el observador y la superficie del monitor. Las imágenes con paralaje negativo se perciben en el espacio del observador.

**estereopsis** Sentido de percepción de profundidad.

**estereoscopio** Instrumento óptico desarrollado para observar imágenes fijas en 3D.

**fusión** Combinación de las perspectivas izquierda y derecha en una sola imagen realizada por el cerebro.

**paralaje** Distancia entre puntos homólogos entre las perspectivas izquierda y derecha en una pantalla de despliegue.



## Bibliografia

- [1] Creating anaglyphs with adobe photoshop. [Online] <http://dogfeathers.com/3d/3dhowto.html>.
- [2] The FOURCC Definition List. [Online] <http://www.webartz.com/fourcc/indexyuv.htm>.
- [3] David C. Burr and John Ross. How does Binocular Delay Give Information about Depth? *Vision Research*, 19:523–532, july 1978.
- [4] Cheung Chok Kwan. Fast Motion Estimation Techniques for Video Compression. Master’s thesis, City University of Hong Kong, 1998.
- [5] Cheung Chok Kwan and Po Lai Man. Normalized Partial Distortion Search Algorithm for Block Motion Estimation. *IEEE Transactions on Circuits and Systems for Video Technology*, 10(3):417–422, april 2001.
- [6] Daniel B. Diner. A New Definition of Orthostereopsis for 3-D Television. Technical report, Jet Propulsion Laboratory / NASA, Pasadena, California 91109, USA, 1991.
- [7] J.P.O. Evans et al. The Development of 3D (Stereoscopic) Imaging Systems for Security Applications. Technical report, Department of Electrical and Electronic Engineering, Nottingham Trent University, 1995.
- [8] Earl F. Glynn II. YUV Lab Reports. [Online] <http://www.efg2.com/Lab/Graphics/Colors/YUV.htm>.
- [9] Keith Jack. *Video Demystified*, chapter Color Spaces, pages 15–34. Llh Technology Pub, 3rd edition, 2001.
- [10] Po Lai Man, Cheung Chok Kwan, and Ma Wing Chung. A Novel Fast Block Motion Estimation Using Center-Biased Search Pattern. Technical report, Department of Electronic Engineering, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong, 1996.
- [11] Po Lai Man et al. Various Advanced Motion Estimation Research Development Package, 2000. [Online] <http://www.image.ee.cityu.edu.hk/>.
- [12] Lenny Lipton. Developers’ Handbook. Technical report, StereoGraphics Corporation, 1997.
- [13] Liu Lurng Kuo and Feig Ephraim. A Block-Based Gradient Descent Search Algorithm for Block Motion Estimation in Video Coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 6(4):419–422, august 1996.
- [14] Y. Matsumoto et al. Conversion System of Monocular Image Sequence to Stereo Using Motion Parallax. In *Stereoscopic Displays and Virtual Reality Systems IV*, volume 3012, pages 108–115. SPIE, february 1997.

- [15] Suzanne P. McKee. The Spatial Requirements for Fine Stereoacuity. *Vision Research*, 23(2):191–198, june 1983.
- [16] Vasily G. Moshnyaga. An Adaptive Block Matching Algorithm for Motion Estimation. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1953–1956, march 1999.
- [17] Arun Netravali et al. *Digital Video : An Introduction to MPEG-2*, chapter Three-Dimensional TV, pages 323–360. Chapman & Hall, 1997.
- [18] Arun Netravali and K.A. Prabhu. Motion-Compensated Component Color Coding. *IEEE Transactions on Communications*, 30(12):2519–2527, december 1982.
- [19] Arun Netravali and J.D. Robbins. Motion-Compensated Television Coding: Part i. *The Bell System Technical Journal*, 58(3):631–669, march 1979.
- [20] T. Okoshi. *Three Dimensional Imaging Techniques*, 1976. Academic Press.
- [21] Li Renxiang et al. A New Three-Step Search Algorithm for Block Motion Estimation. *IEEE Transactions on Circuits and Systems for Video Technology*, 4(4):438–442, august 1994.
- [22] John Ross and J.H. Hogben. Short-term Memory in Stereopsis. *Vision Research*, 14:1195–1201, january 1974.
- [23] John Ross and J.H. Hogben. The Pulfrich Effect and Short-term Memory in Stereopsis. *Vision Research*, 15:1289–1290, january 1975.
- [24] Jo Yew Tham et al. A Novel Unrestricted Center-Biased Diamond Search Algorithm for Block Motion Estimation. *IEEE Transactions on Circuits and Systems for Video Technology*, 8(4):369–377, august 1998.
- [25] Deepak Turaga and Mohamed Alkanhal. Search Algorithms for Block Matching in Motion Estimation, 1989. [Online] Available [http://www.ece.cmu.edu/~ee899/project/deepak\\_mid.htm](http://www.ece.cmu.edu/~ee899/project/deepak_mid.htm).