

01130
16



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO



FACULTAD DE INGENIERIA



REALCE Y SEGMENTACION PARA EL ANALISIS DE REGIONES EN MOVIMIENTO EN UN VIDEO DIGITAL.

T E S I S

QUE PARA OBTENER EL TITULO DE:
INGENIERO EN TELECOMUNICACIONES

P R E S E N T A N :

BEATRIZ FLORES SANCHEZ
VERONICA VILLANUEVA ZUNIGA

ASESOR

DR. FRANCISCO J. GARCIA UGALDE



MEXICO, D. F.

En la Dirección General de Bibliotecas de ...
difundir en formato electrónico e impreso e
lo de mi trabajo recepcional

FECHA: JUNIO DEL 2003

Veronica Villanueva Zúñiga
16 - Junio - 2003



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

***Esto no es el final de una etapa, sino el
comienzo de muchas otras más...***

***Queremos dedicar el presente trabajo a
nuestros padres y hermanos por su ayuda y
apoyo incondicional.***

***Agradecemos a todos nuestros compañeros y
amigos, en especial a Marcos y Gustavo.
Y por supuesto jamás dejaremos de agradecer
lo que nuestra alma mater nos dio...***

¡Gracias UNAM!

Beatriz y Verónica.



INDICE

CAPÍTULO 1

INTRODUCCIÓN.....	1
-------------------	---

CAPÍTULO 2

CONCEPTOS DE IMÁGENES DIGITALES Y TEORÍA DE LA INFORMACIÓN.....	3
---	---

2.1	ESTRUCTURA DEL OJO HUMANO.....	3
2.1.1	Formación de imágenes en el ojo.....	4
2.1.2	Adaptación al brillo y discriminación.....	5
2.1.3	Luz y luminancia.....	5
2.1.4	Criterio de fidelidad de la imagen.....	5
2.2	CONCEPTOS BÁSICOS DEL VÍDEO.....	5
2.2.1	Vídeo analógico.....	6
2.2.2	Revolución digital.....	6
2.2.3	Vídeo digital.....	6
2.3	PROCESO DE CONVERSIÓN DE VÍDEO ANALÓGICO A DIGITAL.....	7
2.3.1	Teoría de muestreo bidimensional.....	7
2.3.2	Filtrado.....	8
2.3.3	Muestreo.....	8
2.3.4	Cuantización.....	9
2.3.5	Cuantización visual.....	10
2.3.6	Codificación PCM.....	10
2.4	INTRODUCCIÓN AL PROCESAMIENTO DE IMÁGENES DIGITALES.....	11
2.4.1	Definiciones básicas.....	11
2.5	PROCESAMIENTO DE VÍDEO DIGITAL.....	14
2.6	RESUMEN DE APLICACIONES.....	15
2.6.1	Exploración planetaria.....	15
2.6.2	Aplicaciones terrestres con sensado remoto.....	15
2.6.3	Aplicaciones médicas.....	16
2.6.4	Astronomía.....	16
2.7	ELEMENTOS DE UN SISTEMA DE COMUNICACIÓN.....	16
2.8	TEORÍA DE LA INFORMACIÓN.....	17

CAPÍTULO 3

TÉCNICAS DE REALCE DE IMÁGENES DIGITALES..... 19

3.1	DESARROLLO HISTÓRICO DE LOS MODELOS DE PERCEPCIÓN VISUAL.....	19
3.2	MÉTODOS DE MEJORA Y REALCE DE LA IMAGEN.....	20
3.3	CLASIFICACIÓN DE LOS MÉTODOS DE MEJORA Y REALCE DE IMÁGENES.....	21
3.4	MÉTODOS EN EL DOMINIO ESPACIAL.....	21
3.4.1	Métodos basados en los píxeles.....	22
	Imágenes de negativos.....	22
	Mejora de contraste.....	22
3.4.2	Métodos basados en el histograma.....	22
3.4.3	Filtros de alisamiento.....	23
	Filtro de mediana.....	23
3.4.4	Filtros de realce.....	23
	Filtro máscara de enfoque (<i>Unsharp Masking</i>).....	23
3.5	MÉTODOS EN EL DOMINIO DE LA FRECUENCIA.....	25
3.6	MORFOLOGÍA MATEMÁTICA.....	25
3.6.1	Transformaciones morfológicas.....	26

CAPÍTULO 4

JUSTIFICACIÓN DEL USO DE MÉTODOS PEL-RECURSIVOS Y DE SEGMENTACIÓN DE OBJETOS EN MOVIMIENTO CON MIRAS A LA COMPRESIÓN DE VÍDEO..... 29

4.1	ESTIMACIÓN DE MOVIMIENTO.....	29
4.1.1	Diferencia entre movimiento 2D y movimiento aparente.....	29
4.1.2	Problemas en la estimación de movimiento.....	31
	Problema de oclusión.....	32
	Problema de abertura.....	32
4.2	MÉTODOS DE ESTIMACIÓN DE MOVIMIENTO APARENTE.....	34
4.3	CLASIFICACIÓN SEGÚN LOS MODELOS DE MOVIMIENTO.....	34
4.3.1	Modelos lineales de movimiento.....	35
	Modelos paramétricos.....	36
	Modelos quasiparamétricos.....	36
	Modelos no paramétricos.....	36
4.4	CLASIFICACIÓN SEGÚN LA ESTRATEGIA DE MEDICIÓN DE MOVIMIENTO.....	37
4.4.1	Métodos por puesta en correspondencia.....	37
	Métodos por puesta en correspondencia de bloques.....	37
	Métodos por puesta en correspondencia de rasgos de la imagen.....	37
	Métodos diferenciales.....	37
	Métodos bayesianos.....	39
4.5	ALGORITMOS PEL-RECURSIVOS DE ESTIMACIÓN DE MOVIMIENTO.....	39
4.5.1	Diferencia de trama desplazada (DFD).....	40
4.6	ALGORITMO PEL-RECURSIVO BÁSICO.....	40
4.6.1	Método biemond.....	43
4.7	LÓGICA DE LOS ALGORITMOS PEL-RECURSIVOS DE ESTIMACIÓN DE MOVIMIENTO.....	45
4.8	SEGMENTACIÓN BASADA EN MOVIMIENTO.....	47
4.9	MÉTODOS DE SBM.....	48
4.9.1	Métodos directos de SBM.....	49
4.9.2	Métodos indirectos de SBM.....	49
4.9.3	Métodos bayesianos.....	50
4.9.4	Métodos de estimación y segmentación simultáneas.....	50
4.10	ASPECTOS A CONSIDERAR EN LA ELECCIÓN DEL MÉTODO DE SBM.....	51

CAPÍTULO 5

MÉTODO EMPLEADO Y RESULTADOS OBTENIDOS.....	53
5.1 PLANTEAMIENTO GENERAL DEL PROBLEMA.....	53
5.1.1 Campos de Markov y distribuciones de Gibbs.....	53
5.1.2 Estimación bayesiana global usando MRF.....	54
5.1.3 Aplicación.....	55
5.1.4 Modelado de las funciones de energía.....	57
Modelos de las observaciones.....	57
Modelo a priori de la segmentación.....	59
5.2 SECUENCIAS DE PRUEBA.....	63
5.2.1 Secuencia Miss América.....	63
5.2.2 Secuencia Claire.....	63
5.3 SELECCIÓN DE LOS PARÁMETROS DEL MODELO DE ESTIMACIÓN.....	64
5.4 RESULTADOS DEL ALGORITMO DE ESTIMACIÓN DE MOVIMIENTO.....	66
5.5 SELECCIÓN DE LOS PARÁMETROS DEL MODELO DE SEGMENTACIÓN.....	68
5.6 RESULTADOS DE LA APLICACIÓN DEL ALGORITMO DE SEGMENTACIÓN.....	68
5.6.1 Secuencia Miss América tramas 3 y 4.....	70
5.6.2 Secuencia Claire tramas 6 y 7.....	81
5.6.3 Secuencia Miss América tramas 1 y 2.....	92

CAPÍTULO 6

CONCLUSIONES.....	103
-------------------	-----

ANEXO

MÉTODOS DE COMPRESIÓN DE SECUENCIAS DE IMÁGENES.....	105
SIMETRÍA.....	109
COMPRESIÓN SIN PÉRDIDAS.....	112
COMPRESIÓN CON PÉRDIDAS.....	113

BIBLIOGRAFÍA.....	119
--------------------------	------------

PAGINACIÓN DISCONTINUA

Capítulo 1

INTRODUCCIÓN.

El desarrollo de la ciencia y la tecnología en las comunicaciones y la computación han llevado a la aparición de sistemas de procesamiento y análisis de secuencias de imágenes en áreas tan diversas como la industria del entretenimiento, los servicios de comunicaciones visuales, los servicios de presentación múltiple (multimedios), la educación, la medicina, el control a distancia y la investigación científica.

Históricamente la primera vez que se hizo uso de las técnicas de imágenes digitales fue en los años Veinte en la transmisión de imágenes de fotografías periodísticas a través de cable submarino entre Londres y New York. Esto supuso que el tiempo entre emisión y recepción de las imágenes se redujera drásticamente de una semana (por barco) a tres horas.

Durante los años ochenta las técnicas de análisis de imágenes se desarrollaron de forma vertiginosa como consecuencia de la gran cantidad de aplicaciones que aparecieron y la madurez alcanzada en el diseño de arquitecturas de computadores. Las mayores contribuciones se han centrado en el desarrollo de algoritmos para la detección de características (bordes, líneas, texturas), así como en el desarrollo de técnicas globales de segmentación de una imagen en regiones.

A partir de la segunda mitad de la década de los 80 se inició un creciente interés hacia el desarrollo de nuevas teorías y algoritmos para la interpretación de imágenes bidimensionales como proyecciones de escenas tridimensionales, donde el movimiento es una importante fuente de información acerca del ambiente de la escena y de la estructura tridimensional de los objetos que la conforman. Los trabajos realizados se centraron principalmente en:

El estudio del flujo óptico y sus múltiples aplicaciones: caracterización de superficies 3D, caracterización de superficies a partir del estudio del movimiento y técnicas de representación y búsqueda de objetos.

Precisamente es en ésta década en la que el método de estimación en el que se basa el presente trabajo, pel-recursivo, alcanza su mayor desarrollo. En un principio se empleó para la codificación de señales de televisión, donde los mejores resultados se obtuvieron en aplicaciones con poco movimiento. En particular, para el tipo de secuencias que usamos en éste trabajo, el algoritmo pel-recursivo que proporciona mejores resultados después del análisis realizado en [DGP96] es el del investigador Biemond.

Una vez realizada la estimación de movimiento utilizamos la información generada para iniciar el algoritmo de segmentación, segunda etapa del procesamiento que efectuamos,

basándonos en el movimiento relativo entre dos tramas consecutivas de la secuencia de imágenes analizadas.

En aplicaciones donde no se es muy exigente con la calidad del vídeo, deben encontrarse métodos que se ajusten a estas limitantes. Nuestra propuesta es la introducción de un preprocesamiento basado en un filtrado que simplifique los datos de la imagen y que ayude a obtener mejores resultados en la estimación y segmentación de la misma.

Este tipo de procesamiento y preprocesamiento, lo enfocamos principalmente para aplicaciones donde se cuenta con un ancho de banda limitado; por lo que sería necesaria la optimización de la capacidad del canal de transmisión (cable coaxial, microondas, satélites o fibra óptica). Por ejemplo, la transmisión de imágenes de baja resolución y movimiento limitado (vídeoconferencia) constituye un área de desarrollo en la cual se pueden obtener buenos resultados empleando los métodos que proponemos en éste trabajo.

La vídeoconferencia es un sistema de comunicación diseñado para llevar a cabo educación a distancia, la cual, nos permite la interacción visual, auditiva y verbal con personas de cualquier parte del mundo, siempre y cuando los sitios a distancia tengan equipos compatibles y un enlace de transmisión adecuado entre ellos.

Así entonces, el objeto de estudio del presente trabajo de tesis se centra en secuencias de imágenes donde aparezcan figuras humanas (rostro y torso) con movimientos lentos. Por lo que, resumiendo, abordaremos el estudio de la segmentación basada en movimiento tras haber empleado un algoritmo de estimación de movimiento pel-recursivo de Biernond y estudiaremos también los efectos que causa en la segmentación el empleo de un preprocesamiento con filtros morfológicos y de realce de imágenes. Hemos tomando como punto de partida el trabajo realizado en la tesis [DGP96] para los algoritmos de estimación y segmentación de regiones en movimiento que aquí utilizamos.

Capítulo 2

CONCEPTOS DE IMÁGENES DIGITALES Y TEORÍA DE LA INFORMACIÓN.

La visión es el más avanzado de nuestros sentidos, por lo tanto, no es de sorprender que las imágenes jueguen un importante papel en la percepción humana [GRC 02].

2.1 ESTRUCTURA DEL OJO HUMANO

El ojo humano tiene una forma casi esférica de 20mm de radio que se encuentra rodeado por tres membranas: la *esclerocórnea*, la *ívea* y la *retina* [GRC 02].

La esclerocórnea es la capa formada por la esclerótica y la córnea. La córnea es un tejido transparente muy resistente que cubre la superficie anterior del ojo. La *esclerótica* es la continuación de la córnea y es una membrana opaca que cubre el resto del globo óptico. Tiene la función de protección, es resistente, soporta grandes presiones y mantiene la forma del ojo.

La *ívea* está debajo de la *esclerótica* y es una membrana que contiene una red de vasos sanguíneos fuente de nutrición del ojo. El revestimiento de la *ívea* está altamente pigmentado, por lo tanto, ayuda a reducir la cantidad extra de luz que entra en el ojo y la dispersión dentro del globo óptico. En su extremo anterior, la *ívea* se divide en el *cuerpo de ciliary* y el *diafragma del iris*; que se encargan de controlar la cantidad de luz que entra en el ojo.

La *retina* es la membrana más profunda situada en la parte posterior del ojo; es la zona sensible a la luz, en ella se extienden las terminaciones del nervio óptico. El punto de máxima sensibilidad es una pequeña hendidura llamada *fóvea*, que es donde se encuentran en mayor concentración los receptores discretos de luz llamados *conos* y *bastones*. Los conos son altamente sensibles al color mientras que los *bastones* proporcionan un panorama general del cuadro del campo de vista, no se encuentran involucrados en la visión del color y son sensibles a bajos niveles de iluminación.

En condiciones de buena iluminación (más de 3 cd/m^2)¹, como ocurre de día, la visión es nítida, detallada y se distinguen muy bien los colores; es la *visión fotópica*. Para niveles inferiores a 0.25 cd/m^2 desaparece la sensación de color y la visión es más sensible a los tonos azules y a la intensidad de la luz (*visión escotópica*). En situaciones intermedias, la

¹ Una candela es la intensidad luminosa emitida normalmente por una abertura de $1/60 \text{ cm}^2$, practicada en un recinto que se encuentra a la temperatura de fusión del platino [FEA91].

capacidad para distinguir los colores disminuye a medida que baja la cantidad de luz pasando de una gran sensibilidad hacia el amarillo a una hacia el azul (visión *mesiópica*).

El *lente* se constituye por capas concéntricas de células fibrosas y se encuentra suspendido por fibras que lo unen al *corpo de ciliary*. Absorbe aproximadamente el 8% del espectro de la luz visible (incluyendo la luz ultravioleta y la infrarroja), con relativamente una mayor absorción en las longitudes de onda más cortas.

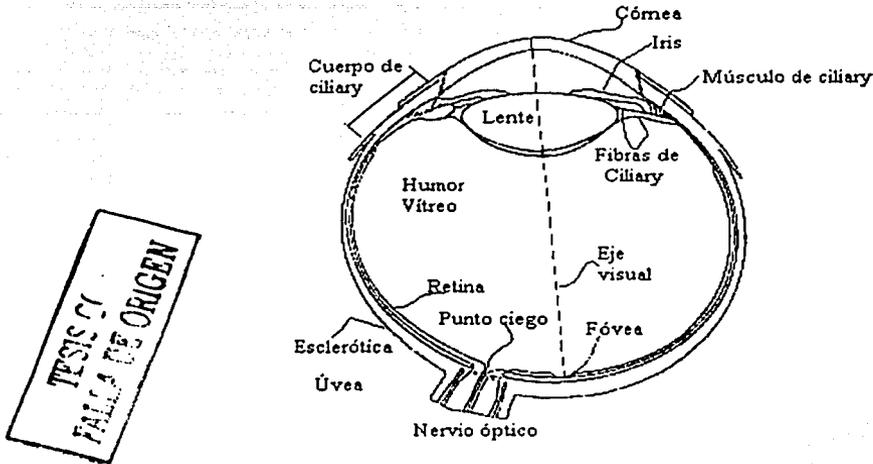


Fig 2.1. Diagrama simplificado de la estructura del ojo.

2.1.1 FORMACIÓN DE IMÁGENES EN EL OJO

El ojo presenta un sistema de lentes, equivalente a una lente convergente, que hace que se forme en la parte posterior (la retina) la imagen de los objetos que se encuentran frente a él [FEA91].

La diferencia entre el lente de una cámara y el del ojo es la flexibilidad. El radio de curvatura de la superficie anterior es mayor que en la superficie posterior. La forma del *lente* es controlada por la tensión en las fibras del *corpo de ciliary* que hacen que el *lente* se contraiga o se expanda para enfocar objetos cercanos o lejanos al ojo.

Cuando enfocamos nuestra visión a un objeto en particular, la imagen retinal es reflejada en el área de la *fovea* (porción central de la retina) excitando los receptores luminosos que a su vez, transforman la energía radiante en impulsos eléctricos que serán decodificados por el cerebro.

2.1.2 ADAPTACIÓN AL BRILLO Y DISCRIMINACIÓN

Las imágenes digitales son desplegadas como un conjunto discreto de puntos luminosos. La capacidad del ojo para discriminar entre diferentes niveles de brillantez es una importante consideración en los resultados del procesamiento de imágenes [GRC02].

El *brillo* de una región, como la percibe el ojo, no sólo depende de la luz radiada de una cierta región sino que, en términos del procesamiento de imágenes, uno de los fenómenos más interesantes es que la respuesta del sistema visual humano a la percepción de la brillantez tiende al sobrepaso (*overshoot*) alrededor de las fronteras de las regiones con intensidades diferentes. Esto provoca que las regiones con intensidad constante aparezcan como si tuvieran variaciones en su brillantez.

2.1.3 LUZ Y LUMINANCIA

La luz es la radiación electromagnética que estimula nuestra respuesta visual que se expresa como la distribución espectral de energía $L(\lambda)$, donde λ es la longitud de onda que se encuentra dentro de la región de luz visible del espectro electromagnético (350 nm a 780 nm) [GRC02].

La *luminancia* o intensidad de un objeto espacialmente distribuido con una distribución de iluminación $I(x, y, \lambda)$ se define como:

$$f(x, y) = \int_0^{\infty} I(x, y, \lambda) V(\lambda) d\lambda \quad 2.1$$

donde $V(\lambda)$ es la *función de eficiencia luminosa relativa* del sistema visual. Para el ojo humano $V(\lambda)$ es una curva con forma de campana cuyas características dependen de que se trate de visión *escotópica* o *fotópica*.

2.1.4 CRITERIO DE FIDELIDAD DE LA IMAGEN

Este criterio es muy útil para medir la calidad de una imagen y para evaluar el desempeño de una técnica de procesamiento [GRC02]. Existen dos tipos: subjetivo y cuantitativo. El criterio subjetivo utiliza escalas de bondad (*goodness*), que es una escala global o de grupo cuyos rangos de evaluación van desde excelente hasta insatisfactorio, y escalas de deterioro que se basa en el nivel de degradación en una imagen cuando se compara con una imagen ideal (muy utilizada en aplicaciones tales como codificación de imágenes).

2.2 CONCEPTOS BÁSICOS DEL VÍDEO

El término vídeo se refiere a la información visual que incluye imágenes variantes en el tiempo. Una imagen variante tiene una distribución espacial de intensidad que cambia con respecto al tiempo.

La señal de vídeo es una señal de tiempo analógica o digital, donde la información espacio-temporal está ordenada como una función del tiempo de acuerdo con una convención de exploración predefinida.

El vídeo tradicionalmente ha sido grabado, almacenado y transmitido en forma analógica, por lo que comenzaremos esta sección con una breve descripción de las señales de vídeo analógicas y después con la representación digital del vídeo, que brinda la oportunidad de comunicaciones y servicios de vídeo interactivo.

2.2.1 VÍDEO ANALÓGICO

En un sistema analógico, la información es transportada por la variación infinita de un parámetro continuo (que puede ser el voltaje en un cable) dentro del ancho de banda disponible [WAT94].

El vídeo analógico es entonces, una señal eléctrica unidimensional $f(t)$ que se obtiene mediante la exploración en las coordenadas vertical y temporal $s_c(x_1, x_2, t)$. La señal $f(t)$ captura solamente los cambios de intensidad de la imagen variante con el tiempo s_c a lo largo de las líneas de exploración, además de información de sincronía y las señales necesarias para alinear los cuadros correctamente.

Sin embargo, el equipo de producción analógica se ha vuelto una tecnología madura y ha alcanzado los límites determinados por las leyes de la física. Las limitaciones de ciertos formatos de vídeo son debido a que las señales de vídeo digitalizadas de fuentes analógicas están generalmente limitadas por la resolución y los dispositivos dispuestos por el respectivo estándar analógico [HHG97].

2.2.2 REVOLUCIÓN DIGITAL

Casi todas las señales son analógicas en el punto donde se originan o en el punto donde se reciben [HHG97].

El bajo costo de los circuitos integrados, redes de comunicaciones de alta velocidad, medios de almacenamiento de rápido acceso, que fácilmente pueden manipular datos de vídeo, están provocando que el estándar analógico se vuelva obsoleto rápidamente.

Actualmente la tecnología se encuentra lista para el uso común del vídeo digital, que además de tener una forma más robusta propia de la señal digital, la principal ventaja de la representación y transmisión digital es que hacen más fácil manipular y procesar prácticamente sin errores información con el único requisito de que se pueda distinguir un bit nulo de otro que no lo es.

2.2.3 VÍDEO DIGITAL

El vídeo digital es una alternativa para transportar una forma de onda de vídeo. Idealmente, los sistemas digitales y los sistemas analógicos de vídeo tienen las mismas características [WAT94].

Esencialmente, el vídeo digital transporta la forma de onda digital numéricamente. El número de muestras es un análogo del tiempo, el cual por sí mismo, es un análogo de la posición a lo largo de la pantalla; y la magnitud de la muestra es (en el caso de la luminancia) un análogo de la brillantez en un determinado punto de la imagen.

Por lo tanto, la sucesión de muestras en un sistema digital es, de cierta forma, un análogo a la forma de onda analógica original; donde la calidad está únicamente determinada por lo apropiado de la conversión y es independiente de la calidad de la trayectoria de la señal.

Para transportar el vídeo numéricamente se utiliza un sistema binario, porque es el que cuenta con un menor número de dígitos (0 y 1), conveniente para conmutar circuitos de un estado de encendido a uno de apagado y viceversa; además de que con sólo dos estados, y con una energía adecuada la posibilidad de error se reduce.

Entre las razones para digitalizar el vídeo analógico está que la calidad de reproducción de un sistema digital es independiente del medio, además que la conversión del vídeo al dominio digital brinda oportunidades que no serían posibles en señales analógicas como las que se enlistan a continuación:

- ✓ No existe pérdida de la calidad en una copia digital.
- ✓ Tasa de transmisión variable o en demanda.
- ✓ Integración de aplicaciones de vídeo tales como T.V. videófono, etc. en una misma plataforma multimedia.
- ✓ Mayor robustez al ruido del canal, corrección de errores y la posibilidad de encriptación.
- ✓ Costos de almacenamiento reducidos.
- ✓ Se les pueden aplicar sistemas y técnicas desarrollados previamente.
- ✓ En sistemas de televisión, eliminan la interferencia, el desvanecimiento y la multitrayectoria frecuentes en la transmisión de T.V. analógica.

2.3 PROCESO DE CONVERSIÓN DE VÍDEO ANALÓGICO A DIGITAL.

Un requisito básico para el procesamiento de imágenes es que éstas se encuentren en un formato digital [CKR96], esto es, que se encuentren como arreglos de palabras binarias de longitud finita. El proceso de digitalización del vídeo consiste de operaciones de prefiltrado, muestreo, cuantización y codificación de la imagen para que pueda ser procesada por una computadora.

2.3.1 TEORÍA DE MUESTREO BIDIMENSIONAL

El proceso de digitalización de imágenes puede ser entendido si modelamos las señales como si fueran de banda limitada; aunque en el mundo real rara vez se encuentran limitadas, se pueden aproximar arbitrariamente por funciones de banda limitada [ANIL91].

Una función $f(x,y)$ es de banda limitada si su transformada de Fourier $F(\xi_1, \xi_2)$ es cero fuera de una región limitada en el plano de la frecuencia.

Cuando una función es periódica, puede ser muestreada a una tasa que iguale o exceda al doble de su frecuencia máxima, y entonces es posible recuperar la forma que sus muestras proveen si el muestreo captura exactamente un número entero de periodos de la función.

La teoría del muestreo puede entenderse recordando el hecho de que la transformada de Fourier de una función muestreada arbitrariamente es una réplica escalada y periódica de la transformada de Fourier de la función original.

2.3.2 FILTRADO

La operación de filtrado o prefiltrado es una operación que se realiza antes del muestreo para reducir las frecuencias no deseadas en la señal, así como el ruido de la misma. El filtrado más simple se obtiene promediando la intensidad de la imagen dentro de una pequeña área alrededor del punto de interés y reemplazando la intensidad del punto original por el promedio calculado [CKR96].

2.3.3 MUESTREO

El siguiente paso es muestrear la señal prefiltrada en un determinado número de puntos de muestreo. Esta operación se puede imaginar como un conmutador que se cierra en un periodo uniforme de tiempo y sensa el valor de amplitud de la señal de entrada, después de lo cual se abre el conmutador esperando al siguiente cierre en un tiempo apropiado [CKR96].

Un método muy común para muestrear una imagen es explorarla renglón por renglón y muestrear luego cada renglón. En otra técnica llamada *scan-in*, el objeto es explorado con un delgado haz de luz colimada la cual ilumina solamente una pequeña región a un determinado tiempo.

La tasa mínima a la cual una señal analógica debe ser muestreada se le conoce como tasa de Nyquist y corresponde al doble de la frecuencia más alta de la señal. Aunque es una práctica generalizada muestrear a una tasa un poco mayor para recuperar más fácilmente la señal de entrada [JAK89].

El fundamento teórico de lo anterior proviene de la unicidad de la transformada de Fourier en el cual, el espectro de la imagen original se puede recuperar de alguna forma a partir del espectro de la imagen muestreada interpolando la imagen continua a partir de una imagen muestreada.

A los límites inferiores de las tasas de muestreo se les conoce como tasas de Nyquist o frecuencias de Nyquist y sus recíprocos como intervalos de Nyquist. La teoría del muestreo establece que una imagen de banda limitada muestreada por encima de sus tasas de Nyquist puede ser recuperada sin error haciendo pasar la imagen muestreada a través de un filtro paso bajas.

Sin embargo, si las frecuencias de muestreo se encuentran por debajo de las frecuencias de Nyquist, las replicas periódicas de $F(\xi_1, \xi_2)$ se superpondrán o encimarán provocando la distorsión del espectro $F_s(\xi_1, \xi_2)$ a partir del cual será imposible recuperar la señal $F(\xi_1, \xi_2)$ original [GRC02]. Este fenómeno se conoce como "*aliasing*" y se encuentra siempre presente en toda imagen muestreada; aunque se pueden reducir sus efectos eliminando las

componentes espectrales altas con un filtro paso bajas antes del muestreo. El efecto del *aliasing* se observa en forma de *Moiré*² en las imágenes, como se muestra en la figura:

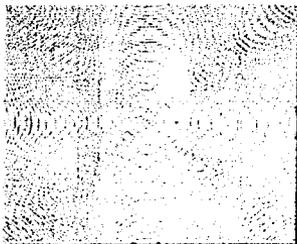


Fig 2.2. Muestra del efecto conocido como *Moiré*.

La teoría de muestreo se basa en varias idealizaciones. Las imágenes reales no son de banda limitada, provocando una superposición de los espectros en el dominio de la frecuencia. La pérdida de resolución también es consecuencia de aspectos que incluyen la apertura finita de los dispositivos de exploración de la imagen y que los sistemas de reconstrucción de la imagen no poseen el filtro paso bajas ideal necesario para dicha operación.

2.3.4 CUANTIZACIÓN.

La señal muestreada, que ahora es una secuencia discreta de valores de *pels* o *pixeles*, es cuantizada. Un cuantizador transforma una variable continua u en una variable discreta v la cual, toma sus valores de un conjunto finito de números [HHG97]. Esta transformación generalmente es una función escalonada con una cantidad determinada de niveles de transición o de decisión.

Esto significa que cuando el valor de un píxel de entrada u se encuentra dentro de un cierto intervalo se le asigna el valor del nivel de reconstrucción que representa a dicho rango de valores. Con la cuantización se obtiene la versión discretizada v de los valores de los píxeles obtenidos en el muestreo.

El proceso de cuantización es irreversible e implica pérdida de información (distorsión) ya que el *píxel* cuantizado v puede ser menos exacto que su valor u de entrada. A la diferencia entre el valor de entrada y su representación cuantizada se le denomina error de cuantización. La elección del número de niveles de cuantización involucra un compromiso entre la exactitud en la representación y la visibilidad de ruido de cuantización en la señal.

² La palabra *Moiré* proviene de la palabra *mohair*, un tipo de ropa hecha del fino cabello de una cabra de Angora.

¡FALLA!
FALLA DE ORIGEN

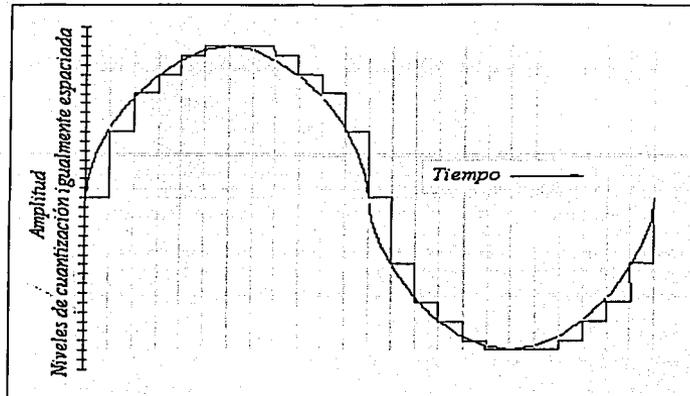


Fig 2.3. Esquema de cuantización.

2.3.5 CUANTIZACIÓN VISUAL.

En una cuantización en escala de grises aplicada en imágenes monocromáticas, si el número de niveles de cuantización no es suficiente entonces se hace visible el fenómeno conocido como de contorno o *contouring* [JAK89]. En este fenómeno grupos de píxeles vecinos son cuantizados hacia el mismo valor, formando regiones con niveles de gris constante cuyas fronteras son los *contornos* creados.

Una cuantización uniforme de imágenes donde los píxeles representan la función de luminancia requiere aproximadamente de 256 niveles de gris (8 bits/píxel) para que no aparezca el efecto de contorno que se manifiesta al usar tasas por debajo de 6 bits/píxel.

En la evaluación de imágenes cuantizadas, el ojo parece ser bastante sensible a los *contornos* y a los errores que afectan la estructura local. Por lo tanto, un esquema de cuantización visual debe tratar de mantener los contornos de cuantización por debajo del nivel de visibilidad en el rango de la luminancia de las imágenes proyectadas.

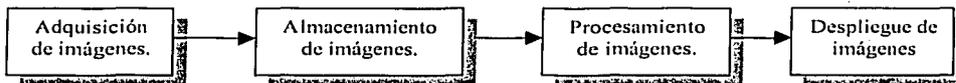
2.3.6 CODIFICACIÓN PCM.

El último paso en la conversión analógico-digital es codificar los valores cuantizados, esto es, obtener una representación digital de dichos valores [HHG97]. El tipo de cuantización que generalmente se utiliza en los píxeles de vídeo es la modulación por codificación de pulsos PCM (Pulse Code Modulation), representándolos con palabras de 8 bits. Esto significa que con éste tipo de codificación es posible asignar a cada pel uno de los $2^8=256$ posibles valores en un rango de 0-255.

La mayoría de los sistemas que proveen representaciones digitales muestreadas de una escena introducen un grado de ruido aleatorio dentro de la imagen, causado por los sistemas electrónicos y por el proceso de conversión analógico-digital.

2.4 INTRODUCCIÓN AL PROCESAMIENTO DE IMÁGENES DIGITALES.

Cada sistema de procesamiento digital de imágenes se puede representar a través de un diagrama de bloques que contenga cuatro elementos principales [GREEN83] como se muestra en la figura siguiente:



El objetivo de la manipulación de imágenes puede dividirse en tres categorías:

1. **Procesamiento:** A partir de una imagen fuente, se obtiene una versión modificada de esa imagen con el resultado deseado.
2. **Análisis:** A partir de una imagen fuente, se obtienen medidas interesantes de ella.
3. **Compresión:** A partir de una imagen fuente, se obtiene una descripción de alto nivel.

A diferencia del estudio de los mecanismos de la visión humana, el procesamiento de imágenes digitales (que frecuentemente incluye su análisis) nace en el momento en que se dispone de los recursos tecnológicos para captar y manipular grandes cantidades de información espacial en forma de matrices de valores.

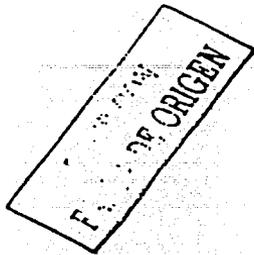
Debemos considerar que el estado natural de una imagen es continuo y que los resultados del procesamiento serán interpretados en forma analógica; por lo que la imagen será discreta sólo por un breve periodo de tiempo suficiente para implementar nuestros algoritmos [CKR96]. Sin embargo, cada paso dentro del proceso puede degradar la imagen, por lo tanto, se debe diseñar cada uno de ellos de la mejor manera posible para minimizar dichos efectos.

2.4.1 DEFINICIONES BÁSICAS.

Una *imagen* es la representación de algo más, contiene la información descriptiva del objeto que ésta representa [CKR96]. Por ejemplo, una fotografía de alguien es la representación de esa persona delante de la cámara en ese momento.

Una *imagen digital* es la representación numérica de un objeto o imagen que puede representarse como una matriz de números [GREEN83]. Un ejemplo sencillo de una imagen digital sería:

TESIS CON
FALLA DE ORIGEN



1	1	1	1	1	1	1	1	1	1
1	0	0	0	1	1	0	0	0	1
1	1	0	1	1	1	1	0	1	1
1	1	0	1	1	1	1	0	1	1
1	1	0	1	1	1	1	0	1	1
1	1	0	0	0	0	0	0	1	1
1	1	0	1	1	1	1	0	1	1
1	1	0	1	1	1	1	0	1	1
1	1	0	1	1	1	1	0	1	1
1	0	0	0	1	1	0	0	0	1
1	1	1	1	1	1	1	1	1	1

Fig 2.4. Ejemplo de imagen digitalizada.

La representación digital de la escena es una versión muestreada de la escena continua del objeto en el espacio. Cada punto dentro de la representación digital corresponde a un área del objeto al cual se le asigna un valor digital relacionado con la intensidad de dicha área.

Líneas	Muestras →					
	135	127	125	122	95	20
	129	14	122	121	87	35
	125	127	120	118	92	31

La intensidad digital del píxel es 125

Fig 2.5. Componentes de las imágenes digitales.

La Figura 2.5, ilustra algunos términos básicos para referirse a las componentes de las imágenes digitales. A la matriz bidimensional de números se le conoce como *imagen digital*. Para establecer la orientación espacial de la imagen muestreada, los renglones de la matriz son las *líneas* y las columnas son las *muestras*. Estos elementos se incrementan hacia abajo y a la derecha respectivamente.

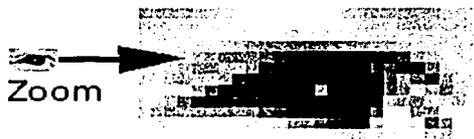
Los componentes individuales que se encuentran dentro de la imagen digital son los *elementos de la imagen*, píxeles. La *intensidad digital* es el valor de cada píxel que representa la intensidad de la imagen, donde cero representa al negro y 255 el blanco.

Por lo tanto, podemos definir una *imagen digital* como una función de dos dimensiones generada por medios ópticos, muestreada en una trayectoria en forma de una malla rectangular igualmente espaciada y cuantizada en intervalos de amplitudes iguales, convirtiéndose así, en un arreglo rectangular de dos dimensiones.

Digitalización es el proceso de convertir una imagen de su forma original a una forma digital. *Conversión* es el proceso no destructivo para la generación de imágenes digitales. *Despliegue* es el proceso inverso a la conversión, en la que se genera una imagen visible a partir de una imagen digital.

La *resolución* es la capacidad de distinguir los detalles espaciales finos. Se define como el número mínimo de pares de líneas distinguibles por unidad de distancia. La frecuencia de muestreo es el factor que determina la resolución espacial de una imagen.

La resolución espacial dentro de una imagen digital es menor a la de la escena continua. Las causas de la degradación son el proceso de muestreo por sí mismo y la óptica utilizada en el dispositivo de adquisición de la imagen. La mayoría de los sistemas de imágenes digitales están diseñados de tal manera que el factor limitante en éste aspecto sea el muestrear la imagen y no el sistema de adquisición.



TESIS CON
FALLA DE ORIGEN

Fig 2.6. Resolución espacial.

Una frecuencia de muestreo alta podría proporcionar una mejor definición en las fronteras, aunque el efecto de degradación en la resolución continuaría presente pero sería menos severo. Sin embargo, esta alternativa genera una representación digital más grande y cara de la misma escena.

El *rango dinámico* de una imagen es el rango de valores a lo largo de la escala de grises que posee una imagen. Por ejemplo, se dice que una imagen tiene un rango dinámico alto, si posee una porción significativa de niveles de gris dentro de la escala de niveles de gris; por lo tanto tendrá un contraste alto. De lo contrario, una imagen con un rango dinámico bajo tendrá una apariencia apagada.

La *profundidad de bits* es determinada por la cantidad de bits utilizados para definir cada píxel. Cuanto mayor sea la profundidad de bits, tanto mayor será la cantidad de tonos (escala de grises o color) que puedan ser representados. Las imágenes digitales se pueden producir en blanco y negro (en forma bitonal), a escala de grises o a color.

Una *imagen bitonal* está representada por píxeles que constan de 1 bit cada uno, que pueden representar dos tonos (típicamente negro y blanco), utilizando los valores 0 para el negro y 1 para el blanco o viceversa.

Una *imagen a escala de grises* está compuesta por píxeles representados por múltiples bits de información. Por ejemplo, 8 bits se tiene una resolución de 256 niveles de gris.

Una *imagen a color* está típicamente representada por una profundidad de bits entre 8 y 24 o superior a ésta. En una imagen de 24 bits, los bits por lo general están divididos en tres grupos: 8 para el rojo, 8 para el verde, y 8 para el azul. Para representar otros colores se utilizan combinaciones de esos bits.

Se le llama *región R* a un subconjunto conectado de píxeles en una imagen. Para establecer si dos píxeles están conectados, se debe determinar si son vecinos y si sus niveles de gris satisfacen un criterio de similitud.

La *frontera o contorno (border)* de una región *R* es el conjunto de píxeles en la región, que tienen uno o más vecinos que no se encuentran dentro de *R*. El contorno de una región finita forma una trayectoria cerrada. Las *orillas (edges)* están formadas con píxeles que exceden un cierto umbral basándose en la medida de la discontinuidad de nivel de gris en un cierto punto.

2.5 PROCESAMIENTO DE VÍDEO DIGITAL.

El procesamiento de vídeo digital se refiere a la manipulación de las tramas de vídeo digital. Lo que hace diferente al procesamiento de vídeo digital del procesamiento de imágenes fijas es que las imágenes de vídeo contienen una cantidad significativa de correlación temporal (redundancia) entre las tramas. Podríamos analizar las imágenes de vídeo como una secuencia de imágenes fijas, donde cada trama es procesada independientemente. Pero explotar la redundancia temporal existente nos lleva a técnicas de procesamiento multitrama que permiten el desarrollo de algoritmos más efectivos.

Operaciones en el procesamiento de secuencias de imágenes:

- **Compresión.** Reducción de la cantidad de información a transmitir, en el caso del vídeo, explotando la redundancia temporal presente en una secuencia de imágenes por medio de la información de movimiento. De donde se genera una gran variedad de algoritmos de compresión, conocidos como técnicas de compresión por *estimación y compensación de movimiento* [DGP96].
- **Reducción de ruido.** Técnicas de filtrado temporales y espacio-temporales con compensación de movimiento se emplean con frecuencia como métodos de reducción de ruido buscando no solo mejorar la calidad de la secuencia sino también para incrementar potencialmente el desempeño de funciones posteriores [DGP96].
- **Restauración.** Empleando la correlación temporal y la información de movimiento entre tramas se elaboran métodos multitrama para restauración. Entendiendo por esta última la acción en la que se depura una secuencia de imágenes degradada por difuminaciones ópticas, las cuales pueden ser ocasionadas por el movimiento relativo entre la cámara y la escena, o por la adquisición desenfocada de la secuencia, provocando entonces que un solo punto de un objeto se esparza sobre varios píxeles de la imagen [DGP96].
- **Conversión de estándares de vídeo digital.** Esquema mediante los cuales se permite el intercambio de información entre sistemas de vídeo con formatos diferentes de

almacenamiento, transmisión y despliegue; puede incluir operaciones de interpolación o submuestreo de la señal de vídeo [DGP96].

El análisis de movimiento involucra funciones como: *detección y estimación* del movimiento aparente en el plano de la imagen, *segmentación* de la escena basada en movimiento, *reconocimiento* de patrones y su *seguimiento* a través del tiempo, *extracción* de estructura y movimiento tridimensional de los objetos de la escena, e *interpretación* de ésta [DGP96].

Una de las características más importantes de las señales de vídeo es que presentan una gran redundancia espacial y temporal. El objetivo de las técnicas de compresión de vídeo es reducir esa redundancia para disminuir así el ancho de banda necesario para transmitir la señal.

La redundancia espacial se da dentro de una misma imagen y supone que el valor de cada píxel está muy correlacionado con el de sus píxeles vecinos. Las imágenes naturales están compuestas básicamente por áreas limitadas por contornos. Estas áreas suelen ocupar la mayor parte de la imagen, y tienen la característica de que al recorrer sus píxeles, el color y la iluminación cambian muy suavemente. Actualmente se emplea la transformada wavelet para reducir la redundancia espacial. La redundancia temporal se refiere a la fuerte correlación que hay entre píxeles vecinos a lo largo del tiempo. En una secuencia se lleva a cabo un muestreo temporal, que debe cumplir el criterio de Nyquist para que el observador no detecte dicho muestreo. Esto supone que la diferencia entre dos cuadros consecutivos de una secuencia sea muy baja, existiendo a veces áreas que no cambian en toda la secuencia. Además, los cambios entre cuadros suelen obedecer más al movimiento de los objetos de la imagen, que a la aparición de nuevos contenidos. Las técnicas de estimación y compensación de movimiento son las que tratan de reducir la redundancia temporal.

2.6 RESUMEN DE APLICACIONES.

Esta sección contiene un pequeño resumen de la variedad de aplicaciones que tiene el procesamiento digital de imágenes para ilustrar que estas técnicas son de empleo general en muy diversas áreas de interés.

2.6.1 EXPLORACIÓN PLANETARIA.

El procesamiento digital de imágenes se emplea como apoyo en algunas de las disciplinas científicas dentro de la exploración planetaria, las cuales incluyen:

- Geología
- Cartografía
- Ciencias atmosféricas

2.6.2 APLICACIONES TERRESTRES CON SENSADO REMOTO.

Esta área involucra el uso de sensores y otros instrumentos montados en una nave aérea, espacial, vehículos y plataformas submarinos, o cualquier otro sistema para monitorear las

características de la Tierra; cuyos instrumentos obtengan datos que puedan ser procesados utilizando la manipulación de imágenes.

- Exploración mineral
- Aplicaciones agrícolas y forestales
- Monitoreo del océano
- Análisis del clima
- Exploración submarina

2.6.3 APLICACIONES MÉDICAS.

El procesamiento digital de imágenes se ha convertido en una importante herramienta para el diagnóstico médico. Existen diversas áreas en donde ésta tecnología juega un papel muy importante como lo son:

- Aplicaciones biomédicas
- Análisis del cromosoma
- Radiografía
- Tomografía y Ultrasonido

2.6.4 ASTRONOMÍA.

Las técnicas digitales han sido ampliamente utilizadas en aplicaciones astronómicas para proveer una mejor resolución y rango dinámico. Estas técnicas incluyen análisis radiométrico, realce de imágenes, filtrado digital, clasificación multispectral, etc.

2.7 ELEMENTOS DE UN SISTEMA DE COMUNICACIÓN.

La información representa cierto cambio con respecto a un estado anterior [EWL56]. Proviene de dos fuentes principales:

1. El cerebro del hombre (ideas).
2. Modificaciones del medio físico ambiental.

La información sólo es útil si se la hace llegar a un receptor capaz de interpretarla y aplicarla. En correspondencia, existen dos tipos de receptor de información:

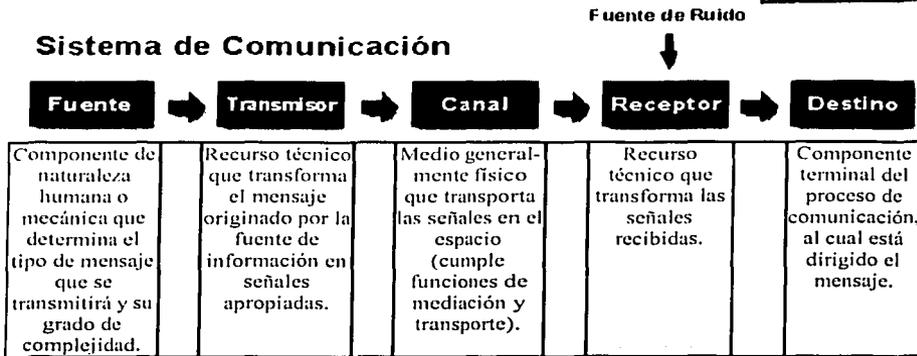
1. El cerebro del hombre
2. Los dispositivos físicos que pueden responder a una señal o a una versión transformada de ella misma.

El más simple de los sistemas de comunicación contiene cinco elementos:

1. Una fuente de información.
2. Un codificador o transductor generalmente incluido dentro del transmisor.
3. Un canal o sistema de transmisión.
4. Un decodificador o transductor generalmente incluido dentro del receptor.

5. Un receptor o destino de la información.

TESIS CON
FALLA DE ORIGEN



RUIDO

Expresión genérica utilizada para referirse a las distorsiones originadas en forma externa al proceso de comunicación.

En estos sistemas, como en cualquier otro, la información puede alterarse durante el proceso de transporte. Los errores tienen dos orígenes posibles:

- La distorsión de la información en los distintos elementos del canal de información.
- La introducción de señales espurias o ruido que el receptor interpreta como parte de la señal originada por la fuente. Este ruido puede aparecer en cualquiera de los eslabones del sistema.

2.8 TEORÍA DE LA INFORMACIÓN.

Esta teoría fue formulada en 1940 por Claude E. Shannon, y aparece publicada por primera vez en octubre de 1948. Posteriormente el sociólogo Warren Weaver redactó un ensayo destinado a enfatizar las bondades de esta propuesta, el cual fue publicado en 1949.

Weaver, define los 3 niveles en que se abordan generalmente los problemas de comunicación:

- **TÉCNICO:** En este nivel se enfrentan los problemas relacionados a la fidelidad con que cierta información puede ser transmitida desde un emisor a un receptor.
- **SEMÁNTICO:** Se estudian cuestiones relativas al significado e interpretación de un mensaje.
- **PRAGMÁTICO:** Se enfoca la comunicación desde el punto de vista de sus consecuencias en el comportamiento de las personas.

Los problemas que plantea Shannon, tienen que ver con la cantidad de información, la capacidad del canal de comunicación, el proceso de codificación que puede utilizarse para cambiar el mensaje en una señal y los efectos del "ruido" [ZIEM81].

El problema básico es: "Dada la fuente del mensaje, ¿Cómo se representarán los mensajes producidos, para llevar al máximo la información transportada por un canal determinado?". La consideración de esta teoría es la medida de la información y como siguiente paso, definir la capacidad para llevar la información.

El resultado de ésta teoría es que, si se reestructura de manera apropiada la señal transmitida, es posible enviar información a través de una canal a cualquier velocidad que sea menor que la capacidad de canal y con un error muy pequeño, a pesar de la presencia del ruido, siempre y cuando se pueda disponer de un tiempo arbitrariamente largo para la transmisión.

Lo anterior se conoce como *Segundo Teorema De Shannon*, según el cual, la probabilidad de error al recibir una secuencia binaria determinada de n dígitos es arbitrariamente pequeña si n se vuelve relativamente grande. Así entonces, existen muchos códigos pero no se nos dice cómo hallarlos, lo que constituye el dilema de la teoría de la información.

Por lo tanto, el modelo propuesto no se refiere a las personas como protagonistas de la comunicación, sino al proceso desde la perspectiva de:

- Sus aspectos mensurables.
- A las condiciones idóneas de transmisión de información entre máquinas.
- Al cálculo de la pérdida de información a través de un canal.

Capítulo 3

Técnicas de realce de imágenes digitales.

En éste capítulo hablaremos de las diferentes técnicas que se utilizan para el realce de imágenes y para la extracción de sus características, poniendo especial énfasis en los filtros morfológicos y en el filtro mascara de enfoque (*unsharp mask*) que serán utilizados en una etapa de preprocesamiento en la experimentación realizada en esta tesis.

3.1 DESARROLLO HISTÓRICO DE LOS MODELOS DE PERCEPCIÓN VISUAL.

Desde la época de los griegos (Platón y Aristóteles) hasta nuestros días se han formulado teorías sobre como percibe el ser humano su realidad exterior y de que manera usa y representa la información capturada por sus sentidos.

Los estudios de Kepler sobre el análisis geométrico de la formación de la imagen en el ojo, los de Newton sobre visión en color y más recientemente los de Helmholtz (1910) sobre óptica fisiológica y de Wertheimer (1912) sobre el movimiento aparente de agrupaciones de puntos o campos. Han establecido las bases de las actuales teorías de la percepción visual.

Desde la aparición de las primeras computadoras en los años 60's se puso de manifiesto la gran potencialidad de éstas para el procesamiento de información espacial en campos de aplicación relacionados con el estudio de propiedades del sistema de visión humano.

Pero debido a la gran complejidad de este sistema de visión y del lento avance de las teorías y algoritmos que explicaban su funcionamiento y propiedades, se intento abordar estos problemas de formas más directas a partir de tres enfoques distintos:

- En un *primer enfoque*, comenzó el desarrollo de técnicas empíricas basadas en criterios matemáticos que aproximaban los problemas de estimación de bordes y líneas usando distintos criterios. Sugiere el uso simultáneo de operadores de distinto tamaño, pero es incapaz de proponer métodos para la evaluación de los distintos algoritmos.
- El *segundo enfoque* profundizaba en el problema reduciendo su alcance a un mundo de bloques blancos mate iluminados sobre fondo negro. Los bloques podían tener cualquier forma, siempre que todas sus superficies fueran planas y todos sus bordes rectos. La simplificación en el modelo se hizo con la idea de que un conocimiento profundo del problema en un mundo simplificado permitiría extrapolar dichos conocimientos a mundos más complejos. Sin embargo, esto no ocurrió así y hubo que estudiar un nuevo enfoque.

- El *tercer enfoque* lo constituyen los trabajos de Horn sobre la formación de la imagen, donde establece modelos de cálculo que expresan la formación de la imagen a través de ecuaciones diferenciales que relacionan los valores de intensidad de la imagen con la geometría de la superficie, tras estudiar el modo en que la iluminación, geometría, la reflectancia de la superficie y el punto de vista del observador actuaban de forma conjunta para crear los valores de intensidad medidos en la imagen.

La tercera aproximación plantea que: *debe existir un nivel adicional de comprensión en el que el carácter de las tareas de procesamiento de la información llevadas a cabo durante la percepción se analicen y comprendan de modo independiente a los mecanismos y estructuras particulares que los implementan en nuestros cerebros.*

Aunque el enfoque dado por Horn en sus trabajos es el adecuado para abordar los problemas de cálculo asociados a los mecanismos de visión, faltaba determinar de qué manera se implementaban los distintos modelos de cálculo dentro del sistema de visión humano.

En este sentido, Julesz (1975) a partir de estereogramas aleatorios en los que demostró que los mecanismos de la visión estereoscópica se realizan en la retina en una etapa previa del proceso de visión. O que el mecanismo de visión humana tiene la posibilidad de interpretar imágenes en 3D usando solamente las informaciones sobre profundidad, distancia y textura.

Este fue el inicio del establecimiento de nuevas teorías en las que se analiza el proceso de visión como *una sucesión de transformaciones que a partir de las imágenes del mundo externo produce una descripción que es útil al observador y no esta ensombrecida por información irrelevante.*

3.2 MÉTODOS DE MEJORA Y REALCE DE LA IMAGEN.

Las técnicas de mejora y realce tienen como objetivo destacar los detalles finos de una imagen o intensificar detalles que han sido difuminados de forma tal que la imagen pueda ser visualizada a través de una pantalla de la forma más adecuada para que el sistema visual humano pueda hacer la mejor interpretación posible de lo que en ella aparece, o que pueda ser usada por un sistema de reconocimiento de formas.

Así, estas técnicas tratan de la eliminación de ruido en la imagen, realce de bordes, corrección de la iluminación, elección de los mejores valores de brillo y contraste, eliminación de los efectos de distorsión introducidos por el mecanismo de captura de la imagen, etc.

Los métodos y objetivos del realce varían con la aplicación. Cuando una imagen es realzada para ser visualizada por el ojo humano, como es el caso de la televisión, el objetivo es la mejora de la apariencia visual. En otras aplicaciones como la identificación realizada por una máquina, la imagen puede ser realzada y se considera este realce como un preprocesamiento que ayuda a la identificación.

El realce de las imágenes incluye aplicaciones que van desde la impresión electrónica de las imágenes médicas, hasta las inspecciones industriales e incluso la detección autónoma de objetivos en las armas “inteligentes”; o como en nuestro caso, para el mejoramiento de los resultados de los algoritmos de estimación y segmentación de secuencias de imágenes.

El realce es una técnica muy cercana e incluso confundible con la restauración. En la restauración, el objetivo es hacer que una imagen degradada se parezca lo más posible a la imagen original; mientras que en el realce, el objetivo es conseguir que la imagen procesada sea mejor que la imagen original. El resultado de cada técnica dependerá de lo que el observador quiera ver.

3.3 CLASIFICACIÓN DE LOS MÉTODOS DE MEJORA Y REALCE DE IMÁGENES.

Una imagen puede ser analizada como una señal en el dominio del espacio o en el dominio de la frecuencia. Por esta razón existen técnicas que permiten llevar a cabo operaciones de realce y/o mejora de la imagen en ambos dominios.

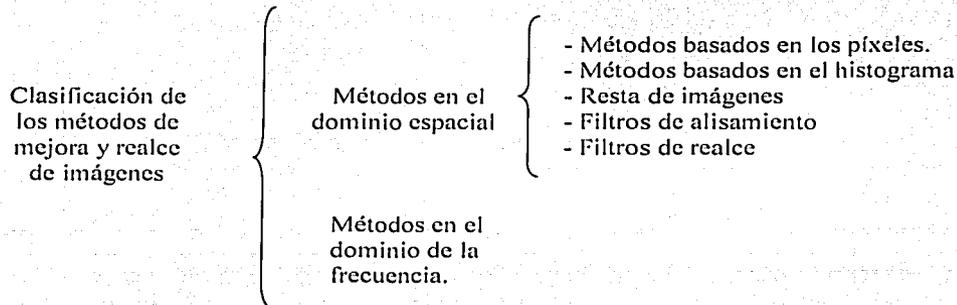


Tabla 3.1 Clasificación de los métodos de mejora y realce de imágenes

En éste trabajo se pondrá mayor énfasis en los filtros en el dominio espacial, pues son los que utilizaremos en la experimentación; mientras que los filtros en el dominio de la frecuencia sólo los mencionaremos y proporcionaremos un breve resumen.

3.4 MÉTODOS EN EL DOMINIO ESPACIAL.

Utilizan los valores de nivel de gris de los píxeles para mejorar la imagen. Este tipo de métodos estudian qué tipo de transformaciones hemos de aplicar a una imagen para que el resultado sea otra imagen que tenga unas características visuales mucho mejores. Es decir,

$$g(x, y) = T\{f(x, y)\} \quad 3.1$$

en donde el cálculo de un valor de g en el punto (x,y) depende en principio de todos los valores de $f(x,y)$; aunque en la realidad, se utiliza la hipótesis de que el valor de la imagen en un punto determinado sólo está relacionado con los píxeles que lo rodean, esto permite la utilización de ventanas, la mayor parte del tiempo rectangulares o cuadradas.

3.4.1 MÉTODOS BASADOS EN LOS PÍXELES.

Se basan en transformaciones simples que con base en un solo píxel permiten obtener mejoras en la visualización de la imagen.

IMÁGENES DE NEGATIVOS.

Dadas las características no-lineales del sistema de visión a veces se calcula la imagen invertida de una imagen dada, ya que aspectos que resultaban difíciles de percibir pueden estar más claros en la imagen invertida como en las radiografías. La expresión que calcula la inversa de una imagen de niveles de gris está dada por:

$$g(x, y) = 255 - f(x, y) \quad 3.2$$

MEJORA DE CONTRASTE.

En ciertas ocasiones es necesario mejorar el contraste de una imagen, debido a causas diversas entre las que podemos encontrar:

- Ausencia de iluminación,
- sensores con un rango de valores de iluminación bajo o,
- cualquier otro factor equivalente.

El objetivo de éste tipo de transformaciones es modificar los valores de los píxeles de manera que se produzca un aumento en el rango dinámico de los valores de niveles de gris de la imagen.

Esto quiere decir que debemos de elegir dos valores extremos de nivel de gris uno por la parte alta de los blancos b , y otro por la parte baja de los negros n .

3.4.2 MÉTODOS BASADOS EN EL HISTOGRAMA.

El histograma de una imagen de niveles de gris es una gráfica que a cada valor posible de nivel de gris entre 0 y 255 le asigna la proporción de píxeles en la imagen con dicho valor.

Los métodos de realce basados en el histograma tienen como fin modificar la masa de puntos asignados a los distintos valores de niveles de gris de forma que la imagen mejore en su visualización. Cuando el rango de valores del histograma es el máximo posible y cuando la masa de puntos está repartida entre 0 y 255 de forma tal que el histograma se asemeje a la forma de una campana, entonces el sistema visual humano consigue una máxima respuesta en su apreciación de la imagen.

3.4.3 FILTROS DE ALISAMIENTO.

Estos filtros sirven para desvanecer una imagen o para reducir su ruido. Un ejemplo de éstos filtros es el de mediana que se describe a continuación.

El principal inconveniente de los filtros de alisamiento que utilizan la convolución es la paulatina desaparición de los saltos entre los niveles de gris de píxeles vecinos. Dichos saltos definen las fronteras presentes en la imagen, por tanto la convolución hace decrecer la "fuerza" con que las fronteras se presentan en la imagen.

FILTRO DE MEDIANA

TESIS CON
FALLA DE ORIGEN

Cuando el objetivo es reducir el ruido sin desvanecer las fronteras, una alternativa es el uso de filtros de mediana, en los que el nivel de gris de cada píxel se reemplaza por la mediana de los niveles de gris en un entorno de este píxel. Este método es particularmente efectivo cuando el patrón de ruido consiste en componentes fuertes y concentradas, y la característica que se desea preservar es la agudeza de los bordes.

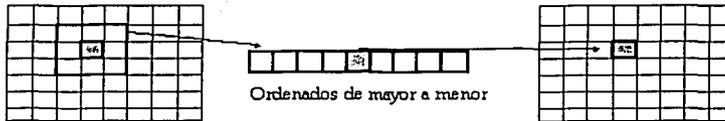


Fig 3.1 Estructura del filtro de mediana

Éste tipo de filtro es el adecuado para suprimir ruidos aleatorios y picos sin significado, ya que obligan que un píxel sea lo más parecido posible a los píxeles de su entorno. El tamaño de la máscara fijará que clase de picos estamos considerando como no deseables dentro de la imagen.

3.4.4 FILTROS DE REALCE

El principal objetivo de los filtros de realce es resaltar aquellas características de la imagen que por causa del mecanismo de captación o por error hayan quedado difuminados en la imagen. Con mucha frecuencia la característica más importante a realzar son las fronteras que definen los objetos presentes en la imagen.

UNSHARP MASKING

Experimentos psicofísicos muestran que una fotografía o señal visual con los bordes acentuados es subjetivamente más agradable a la vista que una reproducción exacta. Para obtener este tipo de imágenes acentuadas se utiliza la técnica conocida como "máscara de enfoque" (*unsharp masking*) [PRATT91].

Existen diferentes algoritmos para lograr esto, uno de ellos es el que se basa en que se tiene una imagen borrosa (con pendiente pequeña), y se le resta otra con una pendiente aún más pequeña. Además a esto se le puede multiplicar por un factor de ponderación c .

Luego se suma la señal de la diferencia con la original para obtener una pendiente mayor y por lo tanto una resolución más alta.

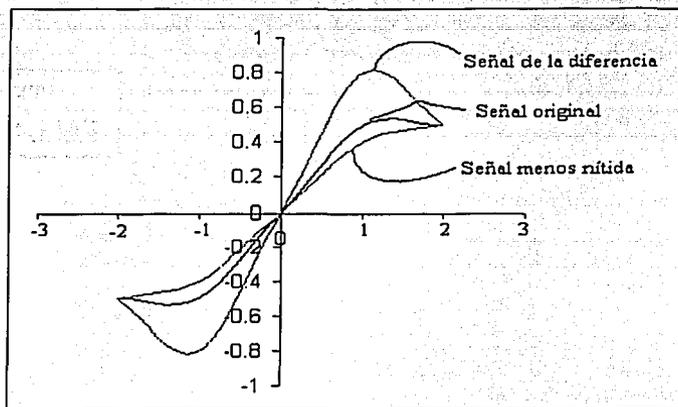


Fig 3.2 Principio del filtro Máscara de enfoque

El proceso es el siguiente: la imagen es digitalizada mediante dos aberturas superpuestas, una a resolución normal y otra con una resolución espacial más baja, las cuales una vez muestreadas producen dos imágenes con resolución normal y baja: $F(j,k)$ y $F_L(j,k)$, respectivamente. Una imagen con máscara de enfoque toma la diferencia ponderada entre las dos imágenes de la manera siguiente:

$$G(j,k) = \frac{c}{2c-1} F(j,k) - \frac{1-c}{2c-1} F_L(j,k) \quad 3.3$$

donde c es la constante de ponderación con un rango de $3/5$ a $5/6$, de tal forma que la razón entre las componentes de la imagen con resolución normal y la de baja resolución es de $1.5:1$ a $5:1$.

Subjetivamente, la imagen procesada con esta máscara resulta más nítida en sus bordes.

El filtro máscara de enfoque se usa para resaltar detalles de las imágenes. A pesar de su nombre (*unsharp mask*) no tiene nada que ver con las máscaras de desenfoco. Su nombre es una reminiscencia de la operación de superposición de un negativo levemente desenfocado sobre un positivo enfocado para resaltar los bordes de las zonas de contraste.

TESIS CON
FALLA DE ORIGEN

3.5 MÉTODOS EN EL DOMINIO DE LA FRECUENCIA.

Se basan en el uso del teorema de convolución. Parten de la hipótesis de que la imagen que queremos mejorar, debido al mecanismo de captación de la imagen, ha sufrido una convolución con una determinada función definida por las lentes del aparato de medida, por el medio en que se ha capturado la imagen, o cualquier otro efecto, y que además está contaminada por ruido aleatorio.

El proceso de mejora en el dominio de la frecuencia se basa en los siguientes pasos:

1. Cálculo de la transformada de Fourier de la imagen a mejorar,
2. Multiplicación por la función de transferencia de un filtro y,
3. Cálculo de la transformada inversa de Fourier para obtener la imagen mejorada.

Este razonamiento matemático aunque es muy simple, presenta serios problemas de inestabilidad en el cálculo de la solución como consecuencia del ruido de la imagen y de los errores de redondeo y truncamiento en los cálculos.

3.6 MORFOLOGÍA MATEMÁTICA.

La palabra morfología generalmente se refiere a la rama de la Biología que se dedica al estudio de la forma y estructura de los animales y plantas. En éste trabajo usaremos la misma palabra pero en el contexto de la morfología matemática.

La morfología matemática se basa en geometría y forma. Las operaciones morfológicas simplifican imágenes y conservan las principales características de formas de los objetos [GRC02].

Los operadores que utiliza la morfología matemática son útiles para:

- Descomponer formas complejas en partes que tienen sentido y separarlas de las partes que le son extrañas,
- Permitir que las formas subyacentes sean identificadas y reconstruidas de forma óptima a partir de sus formas distorsionadas y ruidosas,
- Permiten que cada forma se entienda en función de una descomposición, siendo cada entidad de esa descomposición una forma simple apropiada.
- Pueden simplificar los datos de la imagen, preservar las características esenciales y eliminar aspectos irrelevantes.

Por lo tanto, la morfología matemática se puede usar, entre otros, con los siguientes objetivos:

1. Preprocesamiento de imágenes (supresión de ruido, simplificación de formas),
2. Destacar la estructura de los objetos (extraer el esqueleto, marcado de objetos, envolvente convexa, ampliación, reducción),
3. Descripción cualitativa de objetos (área, perímetro, etc).

3.6.1 TRANSFORMACIONES MORFOLÓGICAS.

El procesamiento morfológico de imágenes constituye una clase de transformaciones en la cual la forma y la estructura de los objetos (patrones espaciales) es modificada para dilucidar su naturaleza morfológica [JLC02].

Dos de las transformaciones básicas en la morfología matemática que estudiaremos en éste trabajo son la **dilatación** y la **erosión**. En la primera de éstas, un objeto crece uniformemente en su extensión espacial, mientras que en la segunda, el objeto reduce su extensión espacial uniformemente. Estas dos operaciones están basadas en operaciones de tipo ¿intercepta, o no intercepta? (*hit or miss*) en las cuales se considera un *templete* o *mascara* de tamaño impar, el cual recorre todas las posiciones definidas por los píxeles de la imagen.

Si el patrón del *templete* o *mascara* coincide (*hit*) con el estado de los píxeles de la imagen cubiertos por este, entonces en la imagen de salida el píxel en correspondencia espacial con el centro del *templete* se pone en un estado predeterminado; por otro lado, si el *templete* no coincide (*miss*) con el estado de los píxeles en la imagen, el píxel de salida se fija con el estado opuesto predeterminado.

TESIS CON
FALLA DE ORIGEN



Fig 3.3 Miss América con filtro de erosión.



Miss América con filtro de dilatación.

Las operaciones de erosión y dilatación filtran el objeto y su complemento respectivamente. En este contexto es posible afirmar que la erosión representa un filtraje al interior y la dilatación un filtraje al exterior.

En general el propósito de todos los filtros morfológicos o de realce de imágenes es el de resaltar ciertos rasgos de la imagen, para que posteriormente sean analizados o desplegados (en una pantalla). Este proceso por sí mismo, no incrementa la información inherente de la imagen, simplemente enfatiza ciertas características específicas, por lo que en general, los algoritmos que se emplean dependen de la aplicación en cuestión.

En éste trabajo, se utilizará el filtro de realce de imagen o *unsharp mask* y los filtros morfológicos de erosión y dilatación para resaltar las características de las secuencias que

utilizamos, con el fin de que al hacer la estimación y segmentación de movimiento de las mismas, los resultados sean cualitativamente mejorados.

La razón de emplear estos filtros en particular es para estudiar, por una parte, si los efectos de la simplificación en las formas de las imágenes facilitan la detección de objetos en la escena y por otro lado, para evaluar los resultados si sólo se acentúan los bordes y el contraste de la misma.

TESIS CON
FALLA DE ORIGEN

Capítulo 4

Justificación del uso de métodos pel-recursivos y de segmentación de objetos en movimiento con miras a la compresión de vídeo.

TESIS CON
FALLA DE ORIGEN

4.1 ESTIMACIÓN DE MOVIMIENTO.

En este trabajo, el análisis de movimiento lo hemos dividido en dos partes: estimación de movimiento y segmentación. En la primera parte se genera un *campo de vectores de movimiento*, es decir se le asigna magnitud y dirección al desplazamiento o a la velocidad de los elementos de la imagen, basándonos en los cambios de los patrones de intensidad de la secuencia. En la segunda parte separamos la imagen en distintas regiones, basándonos en la información producida en la primera etapa.

4.1.1 DIFERENCIA ENTRE MOVIMIENTO 2D Y MOVIMIENTO APARENTE.

Dos conceptos importantes que empleamos para el análisis de movimiento son:

El movimiento 2D o *movimiento proyectado* se refiere a la proyección (perspectiva u ortográfica) del movimiento 3D en un plano de imagen [HORN86]. Entendiendo por plano de imagen, un plano bidimensional en el que se proyectan las escenas tridimensionales de las que esta formada una secuencia de imágenes reales.

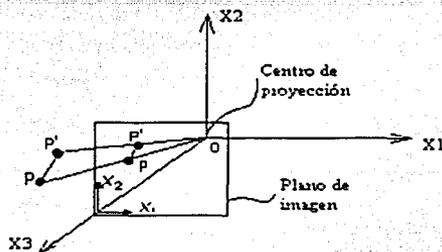


Fig 4.1. Movimiento tridimensional VS bidimensional.

El *movimiento aparente* es el movimiento observado en el plano imagen a partir de la variación en los niveles de gris de la secuencia. Este es el movimiento que vamos a estimar.

Así entonces, podemos medir dos cantidades cinemáticas, el **desplazamiento** mediante los conceptos de *campo de desplazamiento 2D* y *campo de correspondencia*, y la **velocidad** de acuerdo con la *velocidad 2D* y el *flujo óptico*.

Los campos de desplazamiento y de velocidad 2D son las proyecciones de los respectivos campos 3D en el plano imagen, mientras que el campo de correspondencia y el flujo óptico son las funciones de desplazamiento y velocidad, constituyendo así el desplazamiento y velocidad *aparentes* [DGP96].

TESIS CON
FALLA DE ORIGEN

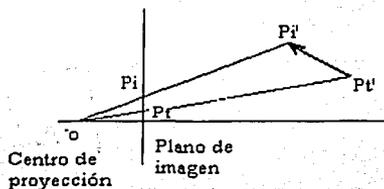


Fig 4.2. Movimiento proyectado.

El problema de la estimación de movimiento aparente lo podemos plantear en dos formas:

- a. La estimación de los vectores de correspondencia utilizando dos tramas

$$d(x, y, t; k\Delta t) = (d_x(x, y, t; k\Delta t), d_y(x, y, t; k\Delta t)) \quad 4.1$$

entre los tiempos $t - k\Delta t$ y t , y para todo punto (x, y, t) en un espacio tridimensional, donde k es un entero y Δt es el intervalo entre tramas.

- b. La estimación de los vectores de flujo óptico

$$v(x, y, t; k\Delta t) = (v_x(x, y, t; k\Delta t), v_y(x, y, t; k\Delta t)) \quad 4.2$$

en el mismo espacio tridimensional.

En este trabajo abordaremos este problema de la primera forma, por lo que para cada uno de los píxeles de la secuencia queremos encontrar la correspondencia

$$I(x, y, t) = I(x - d_x(x, y, t), y - d_y(x, y, t), t - k\Delta t) = I(z - d(x, y, t), t - k\Delta t) \quad 4.3$$

donde

$I(x, y, t)$ es la intensidad de cada píxel de la secuencia.

$z=(x, y)$ es la posición de cada píxel.

$d(x, y, t)=(d_x(x, y, t), d_y(x, y, t))$ es el vector de desplazamiento de cada píxel durante el intervalo $(t - k\Delta t, t)$.

En general, el campo de correspondencia es diferente del campo de desplazamiento 2D porque:

- Carece de suficiente gradiente de imagen espacial: Debe tener suficiente variación de niveles de grises dentro de la región de movimiento para que el movimiento actual sea observable.



Fig 4.3. Todo el movimiento proyectado no genera flujo óptico.

- Cambios en la iluminación externa: Un flujo óptico observable puede no siempre corresponder al movimiento actual.



Fig 4.4. Todo el flujo óptico no corresponde al movimiento proyectado.

4.1.2 PROBLEMAS EN LA ESTIMACIÓN DE MOVIMIENTO.

El problema de estimación de flujo óptico o correspondencia basado sólo en dos tramas es intrínseco, es un problema “mal planteado” en la ausencia de algunas suposiciones adicionales acerca de la naturaleza del movimiento. Un problema es llamado “*ill-posed*” (mal planteado) si la solución única no existe, y/o la(s) solución(es) continuamente no depende(n) de los datos. La estimación de movimiento 2D presenta los problemas de:

- Existencia de una solución: La no correspondencia puede ser establecida para puntos de segundo plano cubiertos/descubiertos.
- Unicidad de la solución: Ya que se presenta el problema de “apertura”.
- Continuidad de la solución: La estimación de movimiento es altamente sensible a la presencia de ruido observable en las imágenes de vídeo.

PROBLEMA DE OCLUSIÓN.

La oclusión se refiere a la cobertura/descubrimiento de una superficie debido a la rotación y traslación 3D de un objeto el cual ocupa solo parte del campo de vista.

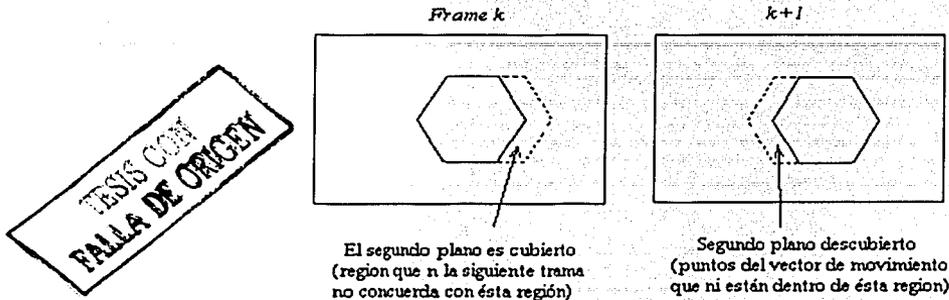


Fig 4.5. Problema de segundo plano cubierto/descubierto.

PROBLEMA DE ABERTURA.

El problema de abertura surge del hecho de que la solución del problema de estimación de movimiento 2D no es única. Si los vectores de movimiento en cada píxel son considerados como variables independientes, el número de ecuaciones es igual al número de píxeles de la imagen, en donde para cada píxel se tienen dos componentes.

Si analizamos teóricamente este problema, tenemos que sólo se puede determinar el movimiento que es ortogonal al gradiente espacial de la imagen, llamado el flujo normal, en cualquier píxel.

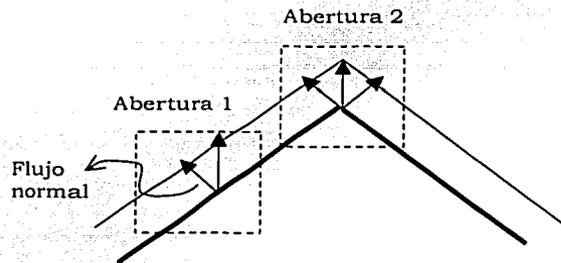


Fig 4.6. Problema de abertura

Si estimamos el movimiento basado en una ventana local (abertura 1) entonces no es posible determinar si la imagen se mueve hacia arriba o perpendicularmente al borde del objeto (flujo normal). Sin embargo, si observamos la abertura 2, es posible estimar el

movimiento correcto ya que la imagen tiene el gradiente en dos direcciones perpendiculares en la apertura. Con ello es posible superar el problema de apertura para la estimación basada en un bloque de píxeles que contiene suficientes variaciones en los niveles de gris.

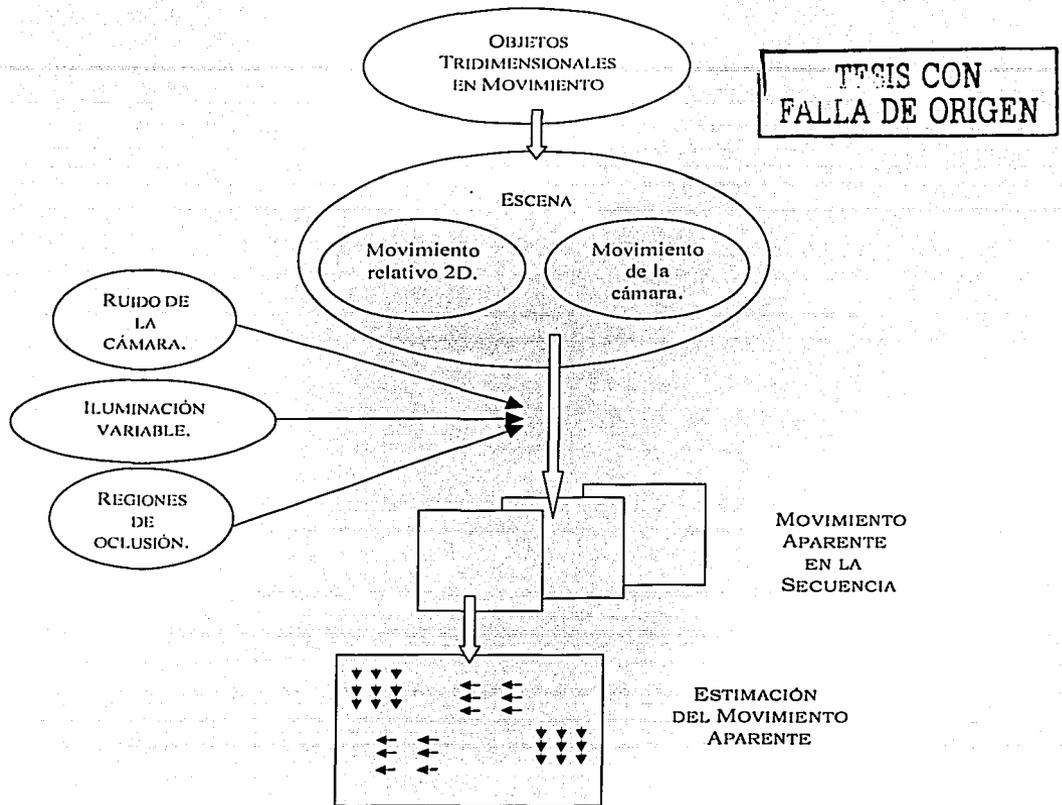


Fig 4.7. Problemática de la estimación de movimiento aparente.

Ya que la solución del problema de estimación de movimiento no es trivial, podríamos obtener diferentes soluciones obtenidas con el uso de información adicional sobre las características deseadas o supuestas para el campo de movimiento, es decir *restricciones de suavizado* como: métodos de regularización determinísticos, o enfoques estocásticos de optimización global.

4.2 MÉTODOS DE ESTIMACIÓN DE MOVIMIENTO APARENTE.

Existe una gran cantidad de métodos de estimación de movimiento aparente de una secuencia de imágenes. Dependiendo de su aplicación encontramos diferentes clasificaciones en la literatura, por lo que un método puede pertenecer a más de una categoría.

Critero de Clasificación	Clases de Métodos	Métodos Específicos
Modelo de Movimiento.	<ul style="list-style-type: none"> Paramétricos. Cuasiparamétricos. No paramétricos. 	
Estrategia de medición del movimiento.	Puesta de correspondencia.	<ul style="list-style-type: none"> Bloques de píxeles. Rasgos de imagen.
	Diferenciales.	<ul style="list-style-type: none"> Basados en la ecuación de flujo óptico. Pel-recursivos.
	Por transformadas.	
	Bayesianos.	

Tabla 4.1. Clasificación de los métodos de estimación de movimiento aparente en secuencias de imágenes.

4.3 CLASIFICACIÓN SEGÚN LOS MODELOS DE MOVIMIENTO.

Debido a su naturaleza *ill-posed*, todos los algoritmos de estimación de movimiento involucran suposiciones adicionales acerca de la estructura del movimiento a procesar. En muchos casos las suposiciones no son expresadas explícitamente, más bien son presentadas como un término de regularización en una función objetivo, o descritas en primer lugar como una suposición que permite la computación de los vectores de movimiento local [SELA93]. Cada una de estas suposiciones implícitamente representa un *modelo de movimiento* basado en el campo de movimiento, el cual describe la proyección del movimiento 3D de las escenas en el plano imagen [ANAN93].

Así, matemáticamente, la inclusión de un modelo de movimiento nos lleva a una transformación T que describe el comportamiento cinemático de cada píxel o de un grupo de píxeles dentro de una región, entre dos tramas sucesivas

$$I(z, t + 1) = T(I(z, t), \theta) \quad 4.4$$

donde θ es el *vector de parámetros* del modelo de movimiento Θ_M , que caracteriza a cada píxel o región. Con este conjunto de parámetros se pueden describir movimientos como son translación, rotación, divergencias e incluso deformaciones del objeto.

Para definir un modelo de movimiento tenemos que considerar tres factores:

- El sistema de proyección óptico de los objetos en el plano imagen (perspectivo u ortográfico). Perspectivo en nuestro caso.
- El tipo de movimiento 3D de los objetos (rígido o no rígido).
- El modelo geométrico de la estructura de los objetos (conformados por superficies planas o de otro tipo).

En las secuencias analizadas el movimiento que se presentó es no rígido y su proyección es perspectiva, además de que por tratarse de figuras humanas las superficies presentes son no planas.

Estos factores combinados de diferentes formas producen diferentes modelos que se caracterizan principalmente por dos propiedades: *complejidad* y *compacidad*, entre las cuales existe un compromiso.

4.3.1 MODELOS LINEALES DE MOVIMIENTO.

Modelo completo de movimiento 3D rígido planar bajo proyección perspectiva (8 parámetros).

$$\begin{bmatrix} d_x \\ d_y \end{bmatrix} = \begin{bmatrix} a_1 + a_2x + a_3y + a_7xy + a_8x^2 \\ a_4 + a_5x + a_6y + a_8xy + a_7x^2 \end{bmatrix} \quad 4.5$$

donde

$d=(d_x, d_y)$ es el vector de desplazamiento en el plano imagen

(x, y) son las coordenadas originales de cada píxel

$\theta=(a_1, \dots, a_8)$ son los parámetros del modelo, definidos en [ADIV85].

TESIS CON FALLA DE ORIGEN

Modelo lineal o affine (Θ_{MLIN} , 6 parámetros). Solo se consideran los términos lineales del modelo completo.

$$\begin{bmatrix} d_x \\ d_y \end{bmatrix} = \begin{bmatrix} a_1 \\ a_4 \end{bmatrix} + \begin{bmatrix} a_2 & a_3 \\ a_5 & a_6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad 4.6$$

Modelo lineal simplificado (Θ_{MLS} , 4 parámetros). Reducción del modelo completo debido a las suposiciones siguientes:

- El movimiento de rotación sólo se presenta alrededor del eje perpendicular al plano ubicado en el centro de gravedad de cada región considerada.
- Todos los planos que conforman cada objeto o región de la escena son paralelos al plano de imagen.

$$\begin{bmatrix} d_x \\ d_y \end{bmatrix} = \begin{bmatrix} t_x \\ t_y \end{bmatrix} + \begin{bmatrix} k & -\theta \\ \theta & k \end{bmatrix} \begin{bmatrix} x - x_g \\ y - y_g \end{bmatrix} \quad 4.7$$

donde

(t_x, t_y) es el vector de translación del centro de gravedad de la región (x_g, y_g) .
 k es el coeficiente de divergencia (translación a lo largo del eje óptico).
 θ es el ángulo de rotación alrededor del eje perpendicular al plano de imagen.

Modelo constante o translacional (Θ_{Mc} , 2 parámetros). Desde un punto de vista rígido no es paramétrico, es el más simple.

$$\begin{bmatrix} d_x \\ d_y \end{bmatrix} = \begin{bmatrix} t_x \\ t_y \end{bmatrix} \quad 4.8$$

MODELOS PARAMÉTRICOS.

Tratan de describir la proyección del movimiento tridimensional de una superficie en el plano imagen. Dependen de una representación de la superficie 3D. Son modelos aplicables cuando el movimiento de un píxel dentro de una región consiste en un número pequeño de parámetros, entonces todos los píxeles dentro de esta región pueden contribuir a esta estimación produciendo resultados altamente exactos. Estos modelos son aplicados bajo circunstancias limitadas [SELA93].

MODELOS QUASIPARAMÉTRICOS.

Son una subclase de los modelos paramétricos que involucran la representación del movimiento de un píxel como una combinación de una componente paramétrica válida para la región entera y una componente local como una variable independiente (profundidad) que cambia de píxel a píxel.

MODELOS NO PARAMÉTRICOS.

Estos implican el uso de algún tipo de restricciones de suavizado o uniformidad que pueden ser impuestas sobre la variación local dentro del campo de flujo. El campo resultante no puede ser descrito en una forma paramétrica simple porque el vector de flujo debe ser estimado para cada píxel, y generalmente es limitado en exactitud. Sin embargo el modelo es aplicable bajo una gran variedad de circunstancias.

La principal desventaja de los modelos paramétricos es que sólo son aplicables en el caso de movimiento rígido 3D. Las restricciones no paramétricas pueden ser clasificadas como modelos de suavizado determinísticos o estocásticos.

4.4 CLASIFICACIÓN SEGÚN LA ESTRATEGIA DE MEDICIÓN DEL MOVIMIENTO.

4.4.1 MÉTODOS POR PUESTA EN CORRESPONDENCIA.

Estos métodos buscan alguna medida de similitud entre dos grandes clases de estructuras: *bloques de píxeles* de forma y tamaño variable, y *descriptores de imagen* previamente extraídos de la secuencia, como pueden ser puntos, contornos, esquinas, regiones.

MÉTODOS POR PUESTA EN CORRESPONDENCIA DE BLOQUES (Block Matching).

Fueron desarrollados en los años 70's, se encuentran entre las aproximaciones más populares por su poca complejidad computacional y su simplicidad de implantación en circuitería (H.261 y MPEG 1-2) [MUSM85].

Estos modelos suponen que cada imagen se compone de bloques de píxeles que se mueven de una trama a otra, así la estimación consiste en asociar un vector de velocidad o de desplazamiento a cada uno de dichos bloques, empleando como herramienta alguna medida de correlación.

Inicialmente, suponían un movimiento puramente translacional, pero se han desarrollado nuevas estrategias que toman en cuenta movimientos más complejos.

MÉTODOS POR PUESTA DE CORRESPONDENCIA DE RASGOS DE IMAGEN (Feature-based).

Estos métodos necesitan una etapa previa de extracción de rasgos de la imagen que se suponen altamente identificables. Aplicando modelos de movimiento se obtiene un sistema de ecuaciones, regularmente no lineales cuya solución puede plantearse en dos etapas: primero se establece la correspondencia de los rasgos entre tramas sucesivas, para luego con el desplazamiento observado resolver el sistema de ecuaciones, obteniendo así los parámetros correspondientes a los objetos de la escena. Las soluciones se pueden alcanzar basándose en correspondencias entre puntos, entre líneas rectas y entre líneas curvas [HILDS7]. [AGGAS8], [HUAN92].

Estos métodos los podemos encontrar en aplicaciones de visión en robots, navegación, control de tráfico y en compresión de vídeo orientada a objetos [AIZA95], [TEKA95].

MÉTODOS DIFERENCIALES.

Son los métodos basados en la ecuación de flujo óptico. Emplean la *restricción de flujo* con la cual se supone una invariancia de la luminancia o intensidad de las imágenes a lo largo de toda la secuencia. Matemáticamente

$$\frac{dI(z,t)}{dt} = 0$$

4.9

TESIS CON
FALLA DE ORIGEN

37

Entonces las variaciones de intensidad que existan se atribuirán al movimiento. Se supone un movimiento translacional puro y no se consideran las regiones de recubrimiento y descubrimiento, de forma tal que, conociendo el *campo denso de vectores de desplazamiento* sea válida la ecuación 4.3.

Este último campo se puede obtener a partir de:

- Información local obtenida al derivar la función de intensidad de la secuencia con respecto al espacio y al tiempo.
- Restricciones adicionales con las que se define la estructura del movimiento.

De esta manera, la *ecuación de flujo óptico (EFO)* o *ecuación de restricción de movimiento aparente* que se obtiene de la suposición original de invariancia de intensidad y suposiciones adicionales de *continuidad y diferencialidad* de $I(z, t)$, en derivadas espacio-temporales de la luminancia con el vector de velocidad v

$$\frac{\partial I(z, t)}{\partial x} v_x(z, t) + \frac{\partial I(z, t)}{\partial y} v_y(z, t) + \frac{\partial I(z, t)}{\partial t} = \nabla I(z, t) \cdot v + \frac{\partial I(z, t)}{\partial t} = 0 \quad 4.10$$

Hay que notar que los vectores de velocidad y desplazamiento se encuentran relacionados mediante $v = dd/dt$. Para el caso particular de una secuencia de imágenes digitales, se supone que la velocidad durante cada intervalo de tiempo Δt se mantiene constante y que

$$d_x = v_x \Delta t \quad \text{y} \quad d_y = v_y \Delta t \quad 4.11$$

con lo que se puede concluir que la estimación del flujo óptico y de los vectores de correspondencia son problemas equivalentes [BAAZ91b].

A su vez, los métodos diferenciales se pueden dividir en dos grandes grupos:

1. Los que utilizan *explícitamente* la ecuación de flujo óptico.
2. Los que la utilizan *implícitamente*, métodos *pel-recursivos*.

1. MÉTODOS QUE USAN LA ECUACIÓN DE FLUJO ÓPTICO (EFO).

La EFO resulta insuficiente para determinar de manera única las dos componentes del campo de movimiento [HORNS6]. Con ella sólo puede conocerse la componente del campo v en la dirección paralela al gradiente de la intensidad (en la dirección normal a los contornos) ya que en la dirección paralela a los contornos el producto punto de la intensidad y la velocidad es nulo (problema de abertura).

Por esta razón se usan restricciones adicionales acerca de la estructura del movimiento [BARR94]. Generalmente se trata de restricciones de suavizado espacio-temporal cuya variedad origina métodos diferentes.

En general, los métodos que introducen suavizado global producen estimaciones imprecisas en las fronteras de movimiento, mientras que la introducción de suavizado adaptativo (que

representa consideraciones locales explícitas) permite la presencia de discontinuidades en el campo [TEKA95].

2. MÉTODOS PEL-RECURSIVOS.

Estiman un campo denso de vectores de desplazamiento a través de una actualización recursiva y/o iterativa del vector de desplazamiento en cada píxel de la secuencia, basada en la minimización del error de reconstrucción de la imagen predicha. Como son estimadores del tipo predictor-corrector, una predicción inicial del desplazamiento correspondiente a cada píxel se actualiza o corrige de acuerdo a la información de los gradientes espacio-temporales existentes en cada punto de la imagen, y al criterio de optimización.

La predicción inicial del desplazamiento es la estimación final realizada para algún otro píxel previamente procesado. Fueron ampliamente desarrollados en los años 80's [NETR80], [HUAN81], [HUAN83], [PAQU83], [SABR84], [WALK84], [BIEM87], [MOOR87], [WALK87], [TZIR88], [BAAZ91b].

Este es el método que utilizamos en este trabajo para obtener la estimación de movimiento en los campos x y y de las tramas utilizadas. Hablaremos más a fondo de éstos métodos en la sección 4.5.

MÉTODOS BAYESIANOS.

Pueden verse como una extensión general de los métodos diferenciales. Utilizan el criterio *máximo a posteriori (MAP)* para optimizar una función de costo global. Dicha función resulta de la modelización markoviana que facilita la introducción de distintas fuentes de información dentro de un solo esquema, así como la especificación de las interacciones locales complejas entre: la información de intensidad de imagen, vectores y discontinuidades de movimiento y regiones de occlusión.

Para incluir restricciones de suavizado direccionales, se introducen modelos que consideran, por ejemplo, campos de líneas de discontinuidad de movimiento [KORN92], [HEIT93], [BRAI95], [TEKA95].

4.5 ALGORITMOS PEL-RECURSIVOS DE ESTIMACIÓN DE MOVIMIENTO.

Inicialmente se desarrollaron para la codificación de señales de televisión con movimientos lentos, bajo un esquema *DPCM con compensación de movimiento (DPCM-MC)* [NETR79]. Se buscaba que fueran fáciles de implantar en circuitería, por lo que su lógica de operación original era bastante simple y pobre en su desempeño. Posteriormente, se desarrollaron enfoques más complejos para mejorar tanto la estimación del movimiento como la reconstrucción de las secuencias comprimidas.

En nuestro caso, al trabajar con secuencias de videoconferencia donde los movimientos no son muy grandes y se considera que la fuente de iluminación es constante, los algoritmos pel-recursivos resultan de gran utilidad.

4.5.1 DIFERENCIA DE TRAMA DESPLAZADA (DFD).

Los primeros métodos de codificación por compensación de movimiento (pel-recursivos y de puesta en correspondencia entre bloques) basaron su funcionamiento en las siguientes suposiciones [MUSM85]:

- Iluminación constante en la escena.
- Movimiento translacional puro.
- No existencia de regiones de oclusión.

Tomando en cuenta dichas suposiciones resulta válido el uso de la expresión 4.3, como habíamos explicado en la sección 4.1.1.

Ahora pues, definimos la *Diferencia de Trama Desplazada* (DFD) como

$$DFD(z, d(x, y, t)) = I(z, t) - I(z - d(x, y, t), t - k\Delta t) \quad 4.12$$

que es función de dos variables: posición z y desplazamiento d correspondiente a la posición.

La DFD se puede definir como un error de reconstrucción e indica si es posible reconstruir cada píxel de la trama actual $I(z, t)$ a partir de la información existente en la trama anterior $I(z, t - k\Delta t)$.

En la práctica, la DFD, difícilmente llega a ser nula por las siguientes razones:

- i. Existe ruido.
- ii. No se cumple con la hipótesis de partida, es decir, existe oclusión o la iluminación de la escena no es constante.
- iii. Hay errores numéricos introducidos por truncamiento.

Por consiguiente, los algoritmos pel-recursivos buscan minimizar el valor absoluto o el cuadrado de la DFD , o la parte izquierda de la EFO para estimar el campo de movimiento de 2D.

4.6 ALGORITMO PEL-RECURSIVO BÁSICO.

En esta sección daremos una descripción genérica de los algoritmos pel-recursivos, de sus restricciones de suavizado y del criterio de minimización de DFD^2 . Al terminar, hablaremos del algoritmo de estimación de movimiento que utilizamos y de las diferencias que tiene con respecto al algoritmo básico.

Los algoritmos pel-recursivos deben su nombre a la minimización recursiva de la imagen (píxel a píxel, siguiendo la dirección del barrido de una señal de televisión: de izquierda a derecha y de arriba a abajo) que además puede ser iterativa (iteraciones múltiples para un mismo píxel).

La estimación es precedida por una etapa de detección de cambios, donde la diferencia de tramas en cada píxel es comparada contra un umbral. La estimación es ejecutada sólo en los píxeles pertenecientes a la región cambiada.

En 1979, *Netravali y Robbins* propusieron el primer algoritmo pel-recursivo para la estimación del campo de desplazamiento \hat{d} [NETR79], en el que a partir de una estimación inicial del desplazamiento \hat{d}^{i-1} (predicción) se obtiene una mejor estimación \hat{d}^i

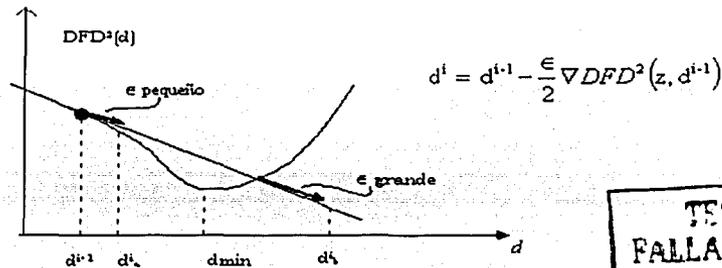
$$\hat{d}^i = \hat{d}^{i-1} + u \quad 4.13$$

donde u es el término de actualización (corrección) de la i -ésima recursión.

Una restricción de suavizado local en la estimación se impone cuando la estimación final \hat{d}^i de un píxel se convierte en la predicción \hat{d}^{i+1} para el siguiente píxel a procesar. Esta técnica se basa en la minimización recursiva de DFD^2 :

$$\hat{d} = \arg \min_d (DFD^2(z, d)) \quad 4.14$$

La minimización se realiza con el método de *descenso por pasos (steepest descent)* el cual actualiza la estimación presente de la posición del mínimo en la dirección negativa del gradiente, llamada dirección de *máximo descenso*. Cabe mencionar que es un método de *optimización local*, con lo que se asegura encontrar un \hat{d}^i que corresponda a un mínimo local de la DFD^2 .



TESIS CON FALLA DE ORIGEN

Fig 4.8. Método de descenso por pasos.

La DFD debe converger a cero cuando la estimación \hat{d} se acerca más al desplazamiento real de la imagen d . Pero normalmente la DFD no es cero, es decir, no se tiene una solución exacta. Así que se introduce un umbral T_{DFD} preestablecido, para considerar soluciones aceptables:

$$|DFD(z, \hat{d}^i)| < T_{DFD} \quad 4.15$$

Así, la expresión recursiva obtenida por el método de descenso por pasos queda

$$d^i = d^{i-1} - \frac{\epsilon}{2} \nabla DFD^2(z, d^{i-1}) \quad 4.16$$

donde

\hat{d}^i es la estimación del vector de desplazamiento para la i -ésima iteración para el píxel presente.

\hat{d}^{i-1} es la estimación inicial (predicción) del vector de desplazamiento.

ϵ es el coeficiente de convergencia fijo.

∇_d es el operador gradiente con respecto al desplazamiento.

Si desarrollando el gradiente:

$$\hat{d}^i = \hat{d}^{i-1} - \epsilon DFD(z, \hat{d}^{i-1}) \nabla I(z - \hat{d}^{i-1}, t - k\Delta t) \quad 4.17$$

donde ∇ es el gradiente de la intensidad de imagen.

En cada iteración se le agrega a la estimación anterior una cantidad vectorial paralela a la dirección del gradiente espacial de la intensidad de imagen con una magnitud proporcional al error de compensación de movimiento. En el algoritmo original, se calcula un nuevo \hat{d}^i , y el viejo \hat{d}^i se convierte en \hat{d}^{i-1} , y así sucesivamente.

La presencia de bordes con distintas orientaciones en una escena es lo que permite la convergencia de \hat{d}^i a un d que satisfaga el criterio de minimización de DFD^2 . El método funciona porque bajo condiciones ideales, al obtener un vector de desplazamiento "correcto" se espera que no cambie drásticamente de un píxel a otro si ambos píxeles forman parte de una región con el mismo movimiento. En las fronteras de movimiento, donde ocurren los cambios considerables en el desplazamiento, es donde son necesarios nuevos periodos de convergencia.

El coeficiente ϵ se elige de manera heurística³ y establece un compromiso entre velocidad y precisión en la convergencia. Si analizamos ϵ podemos decir que:

- ϵ GRANDE produce convergencia rápida, estimación poco precisa u oscilaciones alrededor del mínimo local.
- ϵ pequeños producen una estimación con mayor precisión, pero un mayor número de iteraciones.

Se fija un número máximo de iteraciones por píxel para obtener un vector de desplazamiento que cumpla con el criterio de minimización. El número de iteraciones es variable pero no asegura la convergencia.

³ Heurístico: s. Que induce a buscar la solución.

4.6.1 MÉTODO BIEMOND.

Este algoritmo está basado en la estimación de Wiener y a diferencia del algoritmo pel-recursivo básico, proporciona una estimación lineal por mínimos cuadrados del término de actualización del vector de desplazamiento

$$u = d - \hat{d}^{i-1} \quad 4.18$$

empleando N observaciones para obtener una estimación más confiable del vector de movimiento; d representa el desplazamiento real de cada píxel, y \hat{d}^{i-1} la $(i-1)$ -ésima estimación de éste.

A partir de la definición de la DFD se supone que el término de actualización y el error de linealización de la DFD son procesos estocásticos. La linealización se hace mediante la expansión por series de Taylor de la función de intensidad resultando:

$$DFD(z, \hat{d}^{i-1}) = -\nabla^T I(z - d, t - k\Delta t) \cdot (d - \hat{d}^{i-1}) + v(z, \hat{d}^{i-1}) \quad 4.19$$

El término de error $v(z, \hat{d}^{i-1})$, correspondiente a los términos de orden superior de la linealización de la DFD , no se desprecia sino que se considera como parte del proceso estocástico (ruido). Y el vector $(d - \hat{d}^{i-1})$ es la salida del proceso estocástico.

Las N observaciones representan N puntos z_j , $j=1, \dots, N$ en alguna vecindad de cada píxel analizado. Si este píxel pertenece a alguna región en movimiento, todos aquellos píxeles pertenecientes a la vecindad preestablecida se usan para encontrar una estimación del desplazamiento más estable. Aplicando la ecuación anterior a cada una de las N observaciones se tiene:

$$\begin{pmatrix} DFD(z_1, \hat{d}^{i-1}) \\ \vdots \\ DFD(z_N, \hat{d}^{i-1}) \end{pmatrix} = - \begin{pmatrix} I'_x & I'_y \\ \vdots & \vdots \\ I'_x & I'_y \end{pmatrix} \cdot (d - \hat{d}^{i-1}) + \begin{pmatrix} v(z_1, \hat{d}^{i-1}) \\ \vdots \\ v(z_N, \hat{d}^{i-1}) \end{pmatrix} \quad 4.20$$

donde

$$I'_x = I_x(z_j - \hat{d}^{i-1}, t - k\Delta t)$$

$$I'_y = I_y(z_j - \hat{d}^{i-1}, t - k\Delta t)$$

TESIS CON
FALLA DE ORIGEN

son las componentes del gradiente ∇I en cada píxel z_j desplazado. Ya que es una ecuación matricial, la podemos escribir de la siguiente forma

$$e = Gu + v \quad 4.21$$

donde

e es el vector columna de $N \times 1$ formado por las $DFD(z_j, \hat{d}^{i-1})$.

$u = d - \hat{d}^{i-1}$ es el vector columna $N \times 1$ (término de actualización)

$v(z, \hat{d}^{i-1})$ es un vector columna de errores de truncado.

G es la matriz de $N \times 2$ de gradientes.

Con el modelo anterior se pretende hallar una estimación lineal de mínimos cuadrados \hat{u} de u dadas e y G . Ahora el problema de estimación consiste en encontrar un estimador lineal L

$$\hat{u} = Le \quad 4.22$$

tal que $E\{u - \hat{u}\}^2$ se minimice cumpliendo con la ecuación anterior. Este estimador es conocido como *estimador de Wiener* [HAYK89].

El estimador lineal L se obtiene bajo la suposición de ortogonalidad entre el ruido v y el término de actualización u (intercorrelación nula). La expresión de estimación de \hat{u} queda [BIEM87]:

$$\hat{u} = [G^T R_v^{-1} G + R_u^{-1}]^{-1} G^T R_v^{-1} \cdot e \quad 4.23$$

donde R_u y R_v son las matrices de covarianza de u y v respectivamente. La actualización del vector de desplazamiento se encuentra dada por la expresión:

$$\hat{d} = \hat{d}^{i-1} + \hat{u} \quad 4.24$$

Para simplificar la expresión anterior, se introducen las hipótesis de que las matrices de covarianza tienen media nula y que están mutuamente no correlacionadas, esto es $R_u = \sigma_u^2 I_2$ y $R_v = \sigma_v^2 I_N$ donde σ_u^2 y σ_v^2 representan las variancias del término de actualización y del error de linealización, respectivamente, e I_m representa la matriz identidad de tamaño m .

De ésta manera obtenemos una expresión más sencilla para el cálculo de la actualización recursiva del desplazamiento $\hat{d} = \hat{d}^i$

TESIS CON FALLA DE ORIGEN

$$\hat{d}^i = \hat{d}^{i-1} - \begin{pmatrix} \sum_{j=1}^N (I_x^j)^2 + \mu & \sum_{j=1}^N I_x^j I_y^j \\ \sum_{j=1}^N I_x^j I_y^j & \sum_{j=1}^N (I_y^j)^2 + \mu \end{pmatrix}^{-1} \begin{pmatrix} \sum_{j=1}^N I_x^j \cdot DFD(z_j, \hat{d}^{i-1}) \\ \sum_{j=1}^N I_y^j \cdot DFD(z_j, \hat{d}^{i-1}) \end{pmatrix} \quad 4.25$$

introduciendo un término de amortiguamiento $\mu = \sigma_v^2 / \sigma_u^2$.

Para la vecindad N algunas configuraciones posibles son:



El análisis de la forma final del algoritmo muestra que de él se pueden derivar algunos otros algoritmos pel-recursivos:

1. Si $N = 1$, $\mu = 100$, se obtiene el algoritmo simplificado de Cafforio y Rocca.
2. Si $N = 1$, $\mu = 0$, y se multiplica el término de actualización por $\frac{1}{2}$, se obtiene el algoritmo de Walker y Rao.

4.7 LÓGICA DE LOS ALGORITMOS PEL-RECURSIVOS DE ESTIMACIÓN DE MOVIMIENTO.

Resumiendo lo anterior, aunque pueden plantearse varios algoritmos [NETR79], [WALK84], [BIEM87], [BAAZ91a], todos incluyen los siguientes pasos:

1. *Selección de un valor inicial para la estimación del vector de desplazamiento* ($\hat{d}^{i-1} = \hat{d}^0$).
2. *Prueba de detección de movimiento*. Sirve para asociar un desplazamiento nulo a los píxeles inmóviles. Utiliza la *diferencia interimagen o temporal FD*

$$FD(z, t) = I(z, t) - I(z, t - k\Delta t) \quad 4.26$$

La FD se calcula puntualmente o en una vecindad causal para cada píxel, y se compara con el umbral T_{FD} . Si $|FD(z, t)| < T_{FD}$, $\hat{d}^i = 0$, y el píxel en cuestión es compensado.

3. *Prueba de Recursividad*. Si para el desplazamiento inicial \hat{d}^{i-1} , se cumple que $|DFD(z, \hat{d}^{i-1})| < T_{DFD}$, se concluye que dicha estimación es suficientemente buena, y por tanto, no se necesita actualizar: $\hat{d}^i = \hat{d}^{i-1}$. El píxel es compensado.
4. *Prueba para reestablecimiento de \hat{d}^{i-1}* . Si la estimación inicial es muy mala, conviene recomenzar el proceso de estimación con un vector nulo. Es decir, mayor número de iteraciones, pero con mejor convergencia.
5. *Prueba de iteratividad*. Si las dos primeras pruebas no se cumplen, debe emplearse el algoritmo pel-recursivo para actualizar el vector de desplazamiento, hasta que:
 - a. Se cumpla la condición de convergencia o,
 - b. se llegue al tope máximo en el número de iteraciones it_{max} .
6. *Píxeles no compensados*. Aquellos para los que no se encuentre la convergencia después de todas las pruebas anteriores.⁴

⁴ Con [WALK84], se introduce el concepto de *pixel compensado* para el cual se obtiene una estimación del desplazamiento \hat{d} que cumpla la condición de convergencia ($|DFD(z, \hat{d})| < T_{DFD}$).

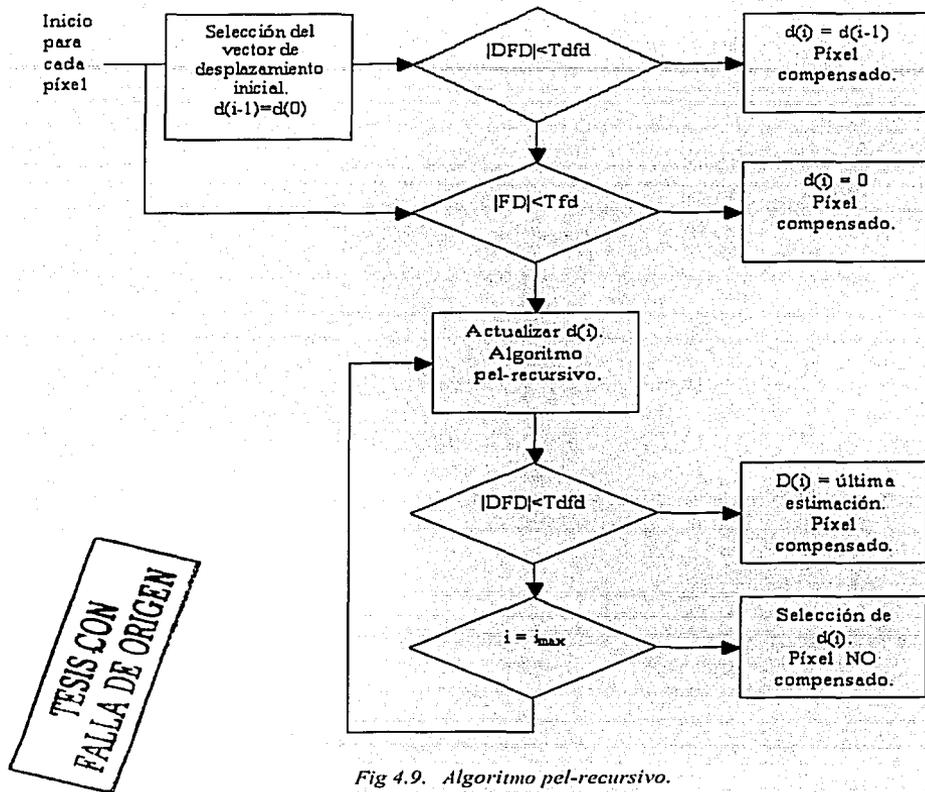


Fig 4.9. Algoritmo pel-recursivo.

4.8 SEGMENTACIÓN BASADA EN MOVIMIENTO.

Constituye la tercera etapa en este trabajo, después del preprocesamiento y la estimación del campo vectorial. A partir de la segmentación haremos el análisis de resultados.

La *segmentación basada en movimiento* (SBM) se refiere a etiquetar píxeles asociados a regiones con movimiento independiente, de objetos 3D en las secuencias [TEKA95].

La segmentación basada en movimiento requiere:

- i. *Mejoramiento de la estimación de flujo óptico.* La *ambigüedad de estimación-segmentación*, plantea que debemos tener una estimación precisa del campo, de no ser así se introducen errores en la segmentación. Para que ésta sea correcta necesitamos una descripción adecuada de los vectores de desplazamiento.

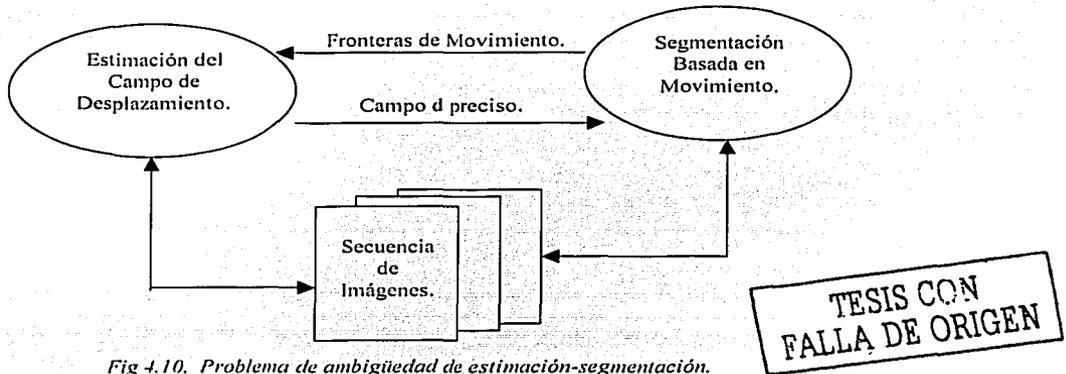


Fig 4.10. Problema de ambigüedad de estimación-segmentación.

Dicho problema se puede abordar de tres formas:

- a. Realizando primeramente la estimación y luego la segmentación con métodos separados.
- b. Invertiendo el orden de las funciones.
- c. Ejecutar ambas operaciones simultáneamente (en el mismo algoritmo).

Aquí, abordaremos el problema de acuerdo con la primera forma. Calculamos la SBM a partir del campo de desplazamiento previamente estimado con el método pel-recursivo Biemond.

- ii. *Estimación de movimiento 3D y estructura de los objetos en una escena.* La segmentación es requerida para modelar los vectores de flujo asociados a cada objeto con movimiento 3D independiente.

- iii. *Descripción de alto nivel de las variaciones temporales y/o el contenido de la secuencia de imágenes.* La información de alto nivel es la segmentación (orientada a regiones u objetos) de los vectores de movimiento (bajo nivel) asociados a cada píxel de la imagen.

En particular, en los métodos de codificación de secuencias de imágenes a tasas bajas y muy bajas, como en nuestro caso, el modelado de la imagen por regiones y su segmentación permiten una representación compacta, que asegura un mínimo de información a transmitir y una buena calidad de reconstrucción.

La SBM también es conocida como *segmentación espacio-temporal* y se diferencia de una segmentación espacial porque el desplazamiento de una región puede ser homogéneo pero al encontrarse gradientes espaciales en ella su movimiento ya no es espacialmente homogéneo y viceversa.

La base de una SBM es la definición de un *criterio de homogeneidad* en cuanto al movimiento, dicho criterio considera la similitud entre los vectores de desplazamiento, además de información adicional.

Para las secuencias que utilizamos, al tratarse de un objeto con movimiento real, se producirá un campo de desplazamiento no homogéneo pues su movimiento no es puramente translacional.

4.9 MÉTODOS DE SBM.

Si nos basamos en el rango de aplicaciones, una primera clasificación es en métodos de *aplicación genérica* y métodos de *aplicación específica*.

Una segunda clasificación emplea como criterio de clasificación la jerarquía clásica de la segmentación espacial: *orientada a contornos* y *orientada a regiones*, pero donde ahora se tienen contornos y regiones espacio-temporales.

Lo anterior lo podemos resumir en la siguiente tabla:

Criterio de Clasificación.	Clases de Métodos.	Método Específicos.
Rango de Aplicaciones.	<ul style="list-style-type: none"> • Aplicación Genérica. • Aplicación Específica. 	
Características Estructurales Buscadas.	<ul style="list-style-type: none"> • Orientados a Contornos. 	<ul style="list-style-type: none"> • Contornos en movimiento. • Fronteras en movimiento.
	<ul style="list-style-type: none"> • Orientados a Regiones. 	<ul style="list-style-type: none"> • Directos (sin estimación d previa). • Indirectos (campo d previo). • Estimación y Segmentación simultáneas.

Tabla 4.2. *Clasificación de los métodos de Segmentación Basada en Movimiento en secuencias de imágenes.*

La segmentación que producen los métodos orientados a contornos es insuficiente para la interpretación, pues es difícil identificar las diferentes regiones con movimiento similar dentro de una imagen [FRAN91].

Los métodos de SBM orientados a regiones los podemos subdividir en *métodos directos*, que lleva la segmentación por medio de los gradientes espacio-temporales de la secuencia, y en *métodos indirectos*, que la hacen partiendo de un campo de movimiento previamente estimado. Y constituyendo otra categoría se tienen a aquellos que realizan la estimación y segmentación *simultáneamente*.

Los métodos orientados a regiones son los que producen resultados con interpretación más útil. El método empleado en este trabajo es de este tipo, con una estimación previa del campo de movimiento por lo que es un método orientado a regiones indirecto.

Ahora, describiremos brevemente los métodos directos e indirectos.

4.9.1 MÉTODOS DIRECTOS DE SBM.

Llevan acabo la segmentación dentro de regiones de movimiento independientes basándose en intensidad de imagen espacio-temporal e información de gradiente.

Este método es el contraste entre estimación del campo de flujo óptico entre dos tramas y la segmentación de imagen basada en el campo de flujo óptico estimado.

El método de este tipo más simple que se conoce es el llamado *detector de movimiento*, el cual se basa en el cálculo de la FD sobre una vecindad comparado con un umbral, de tal forma que se binariza la imagen generando regiones *móviles* (objetos en movimiento y regiones de oclusión) y regiones *inmóviles* [BIER86].

La principal desventaja de este método es la asignación errónea de píxeles sin movimiento a las regiones móviles y viceversa, atribuible al ruido presente en la secuencia; por lo que las regiones móviles obtenidas son de mayor dimensión que las regiones reales, además de que quedan regiones móviles pequeñas aisladas. La corrección sería posible si se plantea un postprocesamiento adecuado (filtrado de mediana o morfológico). Sirve como un primer paso para esquemas de segmentación más elaborados [HOTT88].

Los métodos directos más conocidos son los propuestos por Hotter y Thoma [HOTT88] y Dichl, definidos como algoritmos *jerárquicamente estructurados de arriba hacia abajo*, pues comienzan con un número pequeño de regiones que conforme el algoritmo avanza, éstas se van dividiendo en regiones más pequeñas, dependiendo de qué tan bien se ajuste el modelo de movimiento propuesto a cada región o subregión.

Un enfoque directo estadístico es el de Bouthemy y Santillana [FRAN91], se basa en pruebas de verosimilitud siguiendo un enfoque jerárquico de modelos de movimiento.

4.9.2 MÉTODOS INDIRECTOS DE SBM.

Para estos métodos se cuenta con un campo de movimiento calculado a priori y con un modelo paramétrico asociado con dicho campo. Buscan encontrar un número de regiones N_R que describan adecuadamente el movimiento de cada uno de los objetos de la escena, y determinar el conjunto de parámetros Θ_R del modelo para cada región, asignando así cada vector de movimiento a la región cuyos parámetros lo representen mejor. Y dado que no se conocen ni N_R , ni Θ_R regularmente el cálculo es de forma iterativa.

Los métodos indirectos más representativos son el basado en la transformada de Hough y el bayesiano basado en campos de Markov, que es el que aplicamos en éste caso y que explicaremos más ampliamente en el siguiente capítulo.

4.9.3 MÉTODOS BAYESIANOS.

Éstos métodos buscan por medio del *máximo de probabilidad a posteriori* (MAP) el campo de etiquetas de segmentación dado el campo de desplazamiento. Los campos aleatorios de Markov permiten incorporar múltiples datos previos en forma de iteraciones espacio-temporales locales para obtener las características deseadas en la segmentación. Ya con el modelo el problema es de optimización global que se resuelve mediante cálculos locales.

Murria y Buxton [MURR87] fueron los primeros que propusieron un método de segmentación MAP donde los datos de flujo óptico son modelados suponiendo el movimiento 3D de superficies planas, y el campo de segmentación modelado por una distribución de Gibbs. Ya que los parámetros del modelo que corresponden a cada nivel no se conocen a priori, la segmentación alterna entre la estimación de los parámetros del modelo y la asignación de los niveles de segmentación hasta alcanzar la convergencia.

En [FRAN91] se proporciona un segundo enfoque que se emplea como la primera fase de un sistema de interpretación cualitativa de movimiento en secuencias de imágenes. Éste aunque se basa en el tratamiento estadístico hecho por Boutherm y Santillana, lo hace como una modernización markoviana. Las características de homogeneidad son incluidas en la función de energía a minimizar. Su esquema de relajación es del tipo determinístico, variante de los métodos ICM (*Iterated Conditional Modes*), y emplea una tabla de *etiquetas de inestabilidad*.

Un tercer enfoque markoviano aparece en [CHAN93], incluye información de niveles de grises de la imagen, suponiendo que el campo de desplazamiento y la intensidad de la secuencia son estadísticamente independientes.

4.9.4 MÉTODOS DE ESTIMACIÓN Y SEGMENTACIÓN SIMULTÁNEAS.

Debido a que requieren la introducción de una gran cantidad de información emplean como esquema de modelado los MRF (*campos aleatorios de markov*). [HEIT93] es uno de estos métodos, entrega un campo de velocidades junto con un campo de líneas de discontinuidad de movimiento (SBM orientada a contornos). Y [TEKA95] es otro, éste proporciona junto con el campo de velocidad, un campo de etiquetas de segmentación y un vector de parámetros por cada región obtenida. Mediante la utilización de variaciones de los ICM se optimizan reduciendo complejidad computacional.

4.10 ASPECTOS A CONSIDERAR EN LA ELECCIÓN DE MÉTODO DE SBM.

Aunque un aspecto muy importante a considerar es la aplicación, considerando el desempeño de cada uno, podemos mencionar que:

1. Todos emplean modelos paramétricos de movimiento.
2. Los métodos directos tienen la ventaja de no necesitar el campo de desplazamiento, aunque explotan al máximo el contenido de un cierto tipo de secuencias (videoconferencia) limitando su aplicación. Constituyen esquemas rígidos por lo que no es evidente que se les pueda introducir otro tipo de información.
3. En los métodos indirectos existe inestabilidad temporal, apareciendo, desapareciendo o dividiendo muchas regiones de una trama a otra y segmentando excesivamente, sobretodo en zonas con poca textura.
4. Los métodos indirectos dependen de la precisión del campo de movimiento previamente estimado, aunque con procesos iterativos de estimación y segmentación pueden mejorarse.
5. Los métodos bayesianos son flexibles pues permiten incorporar en un mismo modelo información de diversos tipos y pueden ser aplicados a una variedad de secuencias.
6. Los métodos simultáneos, aunque elegantes y con buenos resultados en la estimación final del campo de movimiento, son computacionalmente muy complejos.

TESIS CON
FALLA DE ORIGEN

52

Capítulo 5

Método empleado y resultados obtenidos.

5.1 PLANTEAMIENTO GENERAL DEL PROBLEMA.

El método abordado en este trabajo para la SBM es bayesiano indirecto basado en un modelo paramétrico, que toma como base el campo de desplazamiento generado mediante el método pel-recursivo basado en Biemond.

Una segmentación basada en movimiento SBM puede plantearse como un problema de estimación Bayesiana global que permita la extracción de la información de la segmentación en forma de un campo de etiquetas que asocia a cada pixel de la imagen a una región en movimiento.

Además, se hará la comparación entre los resultados de este método y los efectos empleando un preprocesamiento con transformaciones morfológicas y con un filtro máscara de enfoque.

5.1.1 CAMPOS DE MARKOV Y DISTRIBUCIONES DE GIBBS.

Sea $e = \{e_s, s \in S\}$ un conjunto de variables aleatorias, en nuestro caso *campo de etiquetas*, definidas en un plano bidimensional $S = \{s_1, s_2, \dots, s_N\}$. Donde cada $s \in S$ se conoce como *sitio*. Además, sean L el conjunto de todos los valores posibles de las variables e_s , donde e consiste en una observación cualquiera del campo $e: e = (e_{s_1}, \dots, e_{s_N})$, y Ψ el conjunto de todas las configuraciones posibles para $e: \Psi = \{e = (e_{s_1}, \dots, e_{s_N})\}$.

Consideraremos un *sistema de vecindad* V definido en $S: \{V = v_s, s \in S\}$, que satisface las siguientes propiedades

- I. s no pertenece a v_s , es decir, un sitio no pertenece a su propio sistema de vecinos.
- II. $s_i \in v_{s_j} \leftrightarrow s_j \in v_{s_i}$, lo que significa que si s_i es el vecino de s_j , entonces s_j debe ser vecino de s_i y viceversa.

Así, e es un campo aleatorio de Markov para un sistema de vecindad V si

- I. $\forall e \in \Psi, p(e = e) > 0.$
- II. $\forall s \in S, p(e_s | e_{s_j}, \forall s_i \neq s_j) = p(e_s | e_{s_j}, \forall s_j \in v_{s_i}).$

Donde Π , significa que la etiqueta de un sitio no depende más que del estado de las etiquetas de sus vecinos; éste es el contexto local del enfoque markoviano.

Con el apoyo del teorema de Hammersley y Clifford se establece la equivalencia entre un campo de Markov y un campo cuya densidad de probabilidad es una *distribución de Gibbs*. Entonces la densidad de probabilidad de e puede escribirse como

TESIS CON
 FALLA DE ORIGEN

$$p(\mathbf{e} = e) = \frac{1}{Z} e^{-U(\mathbf{e})/T} \quad 5.1$$

donde:

T es el parámetro temperatura.

Z es la constante de normalización llamada *función de partición*.

$$Z = \sum_{\mathbf{e} \in \Psi} e^{-U(\mathbf{e})/T} \quad 5.2$$

La función de energía, $U(\mathbf{e})$, se puede descomponer en una suma de funciones locales V_c .

$$U(\mathbf{e}) = \sum_{\mathbf{c} \in C} V_c(\mathbf{e}) \quad 5.3$$

Donde C representa al conjunto de plantillas c de S correspondientes al sistema de vecindad v . A su vez una plantilla es un subconjunto de S donde todos los sitios son vecinos mutuos.

Las *funciones de potencial*, $V_c(\mathbf{e})$, dependen únicamente de los valores e_s , tales que $s \in c$. Estas funciones expresan la interacción entre píxeles vecinos, así que son mediante las cuales podemos introducir las propiedades locales deseadas para el campo de etiquetas e de acuerdo a las características del problema en particular.

5.1.2 ESTIMACIÓN BAYESIANA GLOBAL USANDO MRF.

En forma general el algoritmo de un proceso de estimación bayesiana va como sigue:

Se tiene diversas *fuentes* de información que se usan como *observaciones* en el proceso de estimación: desplazamiento aparente estimado, intensidad de las imágenes (niveles de gris), gradientes espaciales fuertes, píxeles no compensados por el método de estimación de movimiento.

Se hará la combinación de las observaciones de acuerdo a *modelos estructurales locales*, en los que se introduce el conocimiento a priori que se tiene sobre las variables *desconocidas* para estimarlas. En el caso de la SBM, dichas variables son las *etiquetas de segmentación* asociadas a cada píxel, pero puede existir otro tipo de información que resulte útil conocer: *discontinuidades* o *fronteras de movimiento* y *regiones de oclusión* de la imagen.

Modelando las observaciones y las etiquetas en dos campos aleatorios \mathbf{o} y \mathbf{e} , que cumplan con las propiedades necesarias para considerarse campos de Markov, el criterio de *máximo*

a posteriori (MAP) busca estimar el campo e que maximice la densidad de probabilidad condicional a posteriori

$$p(e = e | o = o) \quad 5.4$$

es decir, deseamos encontrar

$$e = \arg \max_e p(e = e | o = o) \quad 5.5$$

Aplicando el teorema de Bayes se puede escribir

$$p(e = e | o = o) = \frac{p(o = o | e = e)p(e = e)}{p(o = o)} \quad 5.6$$

Dado que la maximización no depende del denominador, el criterio MAP se reduce a maximizar la densidad de probabilidad conjunta de las variables observadas y buscadas

$$p(o = o | e = e) p(e = e) = p(e = e | o = o) \quad 5.7$$

5.1.3 APLICACIÓN.

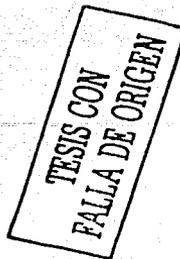
Aplicando la teoría al caso particular planteado en este caso, el problema de la SBM se expresa como

$$(z, l, \theta) = \arg \max_{z, l, \theta} p(z, l, \theta | d_x, d_y, i, p, g) \quad 5.8$$

donde

z es el campo de *etiquetas de segmentación*. Siendo éste el objetivo de nuestra búsqueda, definido para todos los píxeles de la imagen S , tiene asociado un modelo de movimiento Θ_{s_l} . Por lo tanto al encontrar z encontramos un vector de parámetros Θ para cada región.

l es el campo de *líneas de discontinuidad de movimiento*. Campo auxiliar que se introduce con el fin de optimizar la segmentación marcando la posición de todas las discontinuidades en el campo de movimiento. Los sitios s_l de l se localizan entre cada par de sitios en el plano imagen, tanto vertical como horizontalmente. S_l designa el conjunto de todos los s_l . Los valores posibles para los elementos del campo de líneas son 1 ó 0 dependiendo de si hay o no discontinuidad en la respectiva posición. El estado real de cada sitio del campo de líneas es una variable desconocida, que se estima junto con el campo z . Este campo es usado en otras aplicaciones de MRF [GEMA84], [KONR92], [HEIT93], [BRA195], [GARC95a]. Se le puede considerar un resultado secundario.



o es el *campo de oclusión*, definido sobre S , se puede introducir para marcar las regiones del mismo nombre. Los diferentes valores que pueden tomar los elementos son recubrimiento, descubrimiento, no oclusión [DUBO93].

d_x y d_y son los campos observados de las *componentes horizontales y verticales del desplazamiento*. En el caso de ser estimado con un método pel-recursivo existe un vector d para cada sitio de la imagen $s \in S$.

i es el campo de *intensidad de la imagen* (niveles de grises), por lo tanto definido en S . Útil cuando la información de la estimación del desplazamiento es imprecisa como en las fronteras de movimiento.

p es el campo binario de *píxeles no compensados* por el estimador pel-recursivo, definido sobre S . Toma los valores 0 cuando es compensado y 1 cuando no lo es. Se usa para reducir errores en el algoritmo de SBM, las regiones ruidosas o fronteras de movimiento debido al campo d mal estimado.

g es el campo de *bordes espaciales horizontales y verticales* de la imagen, que proporciona información binaria acerca de las discontinuidades de intensidad (toma los valores 1 si hay gradiente fuerte, y 0 en caso contrario). Definido al igual que l sobre S , se emplea para determinar la existencia de discontinuidades de movimiento.

Algunos métodos que parten de este modelo de *campos de Markov acoplados* son los trabajos publicados:

- [MURR87], donde se incluyen z , l , d_x y d_y .
- [CHAN93], donde se consideran z , i , d_x y d_y , y se asume independencia estadística entre las observaciones.

En el trabajo presente se excluye el campo de oclusión, o

$$(z, l) = \arg \max_{z, l} p(z, l | d_x, d_y, i, p, g) \quad 5.9$$

aplicando el teorema de Bayes se puede descomponer:

$$(z, l) = \arg \max_{z, l} p(d_x, d_y, i | z, l, p, g) p(z, l, p, g) p(l | p, g) \quad 5.10$$

ya que el denominador no depende de las etiquetas. Maximizar la ecuación anterior equivale entonces a maximizar la densidad de probabilidad conjunta de observaciones y etiquetas

$$\max p(z, l, d_x, d_y, i, p, g) \quad 5.11$$

Además supondremos independencia estadística entre d e i y omitiremos la dependencia con respecto a l y g . Por otro lado, se establece que el modelo a priori de etiquetas z , sólo

TESIS CON
 FALLA DE ORIGEN

depende de l , y que el modelo a priori del campo de líneas sólo depende de g . Adecuándonos a nuestro problema en particular, entonces tenemos

$$(z, \hat{l}) = \arg \max_{z, l} p(d_x, d_y | z, p) p(i | z, p) p(z | l) p(l | g) \quad 5.12$$

Si se modelan las probabilidades correspondientes a las observaciones (las dos primeras) suponiendo campos aleatorios gaussianos condicionalmente independientes, y las correspondientes a etiquetas (las dos segundas) por medio de una distribución de Gibbs [GEMA84], [CHEL93] se demuestra que maximizar la densidad de probabilidad anterior equivale a minimizar la función de energía

$$U(z, l, d_x, d_y, i, p, g) \quad 5.13$$

TESIS CON
FALLA DE ORIGEN

con la forma específica

$$U(z, l, d_x, d_y, i, p, g) = \alpha U_d(d_x, d_y, z, p) + \beta U_i(i, z, p) + \gamma U_z(z, l) + \kappa U_l(l, g) \quad 5.14$$

donde cada término de energía U se obtiene de la densidad de probabilidad correspondiente en la ecuación anterior, y α, β, γ y κ son constantes de ponderación. Así que finalmente

$$(z, \hat{l}) = \arg \min_{z, l} U(z, l, d_x, d_y, i, p, g) \quad 5.15$$

5.1.4 MODELADO DE LAS FUNCIONES DE ENERGÍA.

Los términos U_d y U_i modelan la relación entre las observaciones y las etiquetas de segmentación, es decir, la densidad de probabilidad condicional $p(o|e)$, mientras que U_z y U_l representan el modelo a priori de las etiquetas, la densidad de probabilidad $p(e)$.

MODELOS DE LAS OBSERVACIONES.

CAMPO DE DESPLAZAMIENTO Y PÍXELES NO COMPENSADOS.

La energía $U(d_x, d_y, z, p)$ es un indicador de qué tan bien el modelo de movimiento Θ_M asociado refleja el campo real estimado, ya que modela las observaciones del campo de movimiento, dado el campo de etiquetas de segmentación y el de píxeles no compensados.

El Θ_M considerado en la elaboración de este trabajo es el modelo lineal simplificado de 4 parámetros Θ_{MLS} .

Despreciando por un momento la influencia del campo p . De manera similar a [MURR87], suponiendo que el Θ_{MLS} es el adecuado, la distancia euclidiana entre los vectores de desplazamiento estimados d y los sintetizados por Θ_{MLS}, d_Θ está dada por

$$\delta_{d, d_\Theta} = \|d - d_\Theta\| \quad 5.16$$

se puede modelar como ruido gaussiano con media nula y variancia σ^2 fija, e independientemente en cada uno de los sitios s . La función de densidad de probabilidad $p(\mathbf{d}_x, \mathbf{d}_y | z)$ se puede expresar entonces

$$p(\mathbf{d}_x, \mathbf{d}_y | z) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\sum (\delta_{d_x, d_y})_s^2 / 2\sigma^2\right) \quad 5.17$$

Maximizar $p(\mathbf{d}_x, \mathbf{d}_y | z)$ equivale a minimizar el argumento de la función exponencial. Ya que el segundo término dentro de la sumatoria es constante, la función de energía a minimizar puede expresarse como

$$U_d(\mathbf{d}_x, \mathbf{d}_y, z) = \lambda \sum_{s=1}^N (\delta_{d_x, d_y})_s^2 \quad 5.18$$

Así su valor depende del ruido presente en el campo de desplazamiento, de los errores en la segmentación y de los errores de modelado. Por lo que la energía será menor entre menor sea el error δ_{d_x, d_y} .

Según el criterio de reconstrucción, el campo p proporciona una medida de la calidad de la estimación del desplazamiento, por lo tanto debe de reducirse su contribución a la energía total a fin de disminuir los efectos indeseables en la segmentación, para ello hay que darle más valor a otro tipo de información, como los niveles de gris. Por lo que queda

$$U_d(\mathbf{d}_x, \mathbf{d}_y, z) = \lambda \sum_{s=1}^N (1 - p_s) (\delta_{d_x, d_y})_s^2 \quad 5.19$$

INFORMACIÓN DE NIVELES DE GRIS Y PÍXELES NO COMPENSADOS.

Cuando la información de movimiento es inexacta, es decir, en las regiones ruidosas o de oclusión, es adecuado hacer prevalecer el criterio de coherencia espacial local en una vecindad. Entendiendo por vecindad aquellos píxeles pertenecientes a una misma región espacial y que tienen valores de intensidad similar. Modelando la diferencia entre la intensidad i en cada píxel s y la intensidad media sobre una ventana pequeña v centrada en s , como ruido blanco gaussiano [PAPP92], [CHAN93], se encuentra

**TESIS CON
FALLA DE ORIGEN**

$$U_d(i, z, p) = \lambda_2 \sum_{s=1}^N p_s (i_s - i_M(z))^2 \quad 5.20$$

Aquí la interacción con el campo de etiquetas z introduce el valor de la intensidad media $i_M(z)$ sobre v , al considerar para el cálculo sólo aquellos vecinos que compartan la misma etiqueta de segmentación del píxel central s .

MODELO A PRIORI DE LA SEGMENTACIÓN.

ETIQUETAS DE SEGMENTACIÓN Y LÍNEAS DE DISCONTINUIDAD DE MOVIMIENTO.

U_z representa el modelo a priori del campo de etiquetas de segmentación z , a través de su interacción con el campo de líneas de discontinuidad de movimiento l , representado por una distribución de Gibbs

$$p(z|l) = \frac{1}{Z} e^{-U_z(z,l)}$$

TESIS CON
FALLA DE ORIGEN

5.21

En cuya definición se introducen las condiciones que nos llevan a alcanzar los resultados que correspondan con la realidad física. Dichas condiciones deben favorecer la *continuidad* en la segmentación, es decir, la existencia de regiones conexas, claro esta, siempre y cuando las etiquetas de sitios vecinos pertenezcan a la misma región de movimiento, considerando el hecho de que las líneas de discontinuidad de movimiento marcan las fronteras entre regiones con diferente movimiento. La función de energía que cumple con estas condiciones es

$$U_z(z, l) = \sum_{s=1}^N \sum_{s_j \in v_s} (1 - 2\delta(z_s - z_{s_j})) (l(z_s, z_{s_j})) \quad 5.22$$

El sentido de $U_z(z, l)$ queda claro cuando uno se percató de que existe un costo, asociado a cada plantilla, que se incrementa si el campo de etiquetas se aleja localmente del modelo a priori propuesto:

- El término $(1 - 2\delta(z_s - z_{s_j}))$ favorece la continuidad en z , pues U_z es mínima (valor de energía negativo) cuando z_s y z_{s_j} son iguales, y no se presentan líneas de discontinuidad entre los píxeles ($l(z_s, z_{s_j})=0$).
- Si el elemento de línea vale 1, el costo asociado a la plantilla es cero, para no introducir penalización por la existencia de etiquetas de segmentación distintas de los píxeles vecinos, debido a que esto es lo que debe de ocurrir en las fronteras de movimiento.
- El caso de incompatibilidades: aquel en el que los píxeles vecinos tienen etiquetas de segmentación diferentes, sin línea de discontinuidad entre ellos, representa el peor de los casos.

LÍNEAS DE DISCONTINUIDAD DE MOVIMIENTO Y GRADIENTES ESPACIALES.

El objetivo del modelo de energía $U_l(l, g)$ asociado al campo de líneas es introducir o no la penalización que apoyará a la minimización de la energía U_z . Así, dado el campo de bordes, basándose en una modelación por distribución de Gibbs se tiene

$$p(l|g) = \frac{1}{Z} e^{-U_l(l,g)} \quad 5.23$$

Así entonces, hay que tomar dos consideraciones. La primera, el modelo a priori de l debe estar condicionado, en este caso, por la información de gradientes espaciales, entonces una escena que da origen a una discontinuidad de movimiento en ciertos píxeles mostrará bordes espaciales en esas mismas posiciones. Por lo tanto, se supone en el modelo que la introducción de un elemento de línea debe coincidir con un gradiente espacial. Y la segunda, U_l debe favorecer la presencia de ciertas estructuras (líneas, rectas, ángulos y esquinas), es decir, debe modelar la *geometría* de las fronteras de movimiento, penalizando la creación de otras estructuras (cruces, líneas dobles, triples, píxeles y esquinas). Con lo que la expresión adoptada resulta igual a la usada en [GARC95]

$$U_l(l, g) = \sum_{c_i \in C_l} (l_{s_i} (1 - g_{s_i}) + V'_c) \quad 5.24$$

donde

C_l es el conjunto de todas las plantillas para el campo l .

V'_c es el potencial asociado a cada plantilla.

g_{s_i} es valor de g en el sitio s_i .

l_{s_i} es valor de l en el sitio s_i .

TESIS CON
FALLA DE ORIGEN

ESQUEMA DE OPTIMIZACIÓN

Para la minimización global de $U(z, l, dx, dy, i, p, g)$ se tienen dos estrategias de relajación:

1. *Estocástico*: Asegura la convergencia hacia un mínimo global de energía, pero es muy costoso en tiempo de cálculo.
2. *Determinístico*: Reduce el tiempo de cálculo pero no asegura la convergencia al mínimo global, a menos que se tenga una buena inicialización.

Se utilizó el método determinista conocido como ICM (Iterated Conditional Modes) que consiste en minimizar la probabilidad local $p(o_s / e_s) p(e_s / e_{s_j}, s_j \in v_s)$ equivalente a minimizar la energía local $\Delta U_s(e_s)$ en cada sitio s ; pero con una modificación que consiste en la minimización de la energía $U(z, l, dx, dy, i, p, g)$ se realiza en dos partes:

1. Optimizar el campo z a través de la minimización ICM del término:
$$U_1 = \alpha U_d(d_x, d_y, z, p) + \beta U_i(i, z, p) + \gamma U_z(z, l)$$
2. Optimización del campo l a través de la minimización ICM del término.
$$U_2 = \gamma U_z(z, l) + \kappa U_l(l, g)$$

donde $U_z(z, l)$ es el término de enlace entre las dos etapas del proceso de optimización.

La inicialización de los campos i , d y p se obtiene a partir del algoritmo pel-recursivo de estimación de movimiento basado en Biemond. El campo z_0 se genera con un detector de movimiento para la primera iteración; el resto de las iteraciones utilizan los campos z y l de la iteración anterior. El campo l_0 se calcula a partir de las fronteras entre las regiones de la segmentación inicial z_0 . Finalmente, el campo g se obtiene con un método de umbral que determina si existen o no discontinuidades de intensidad.

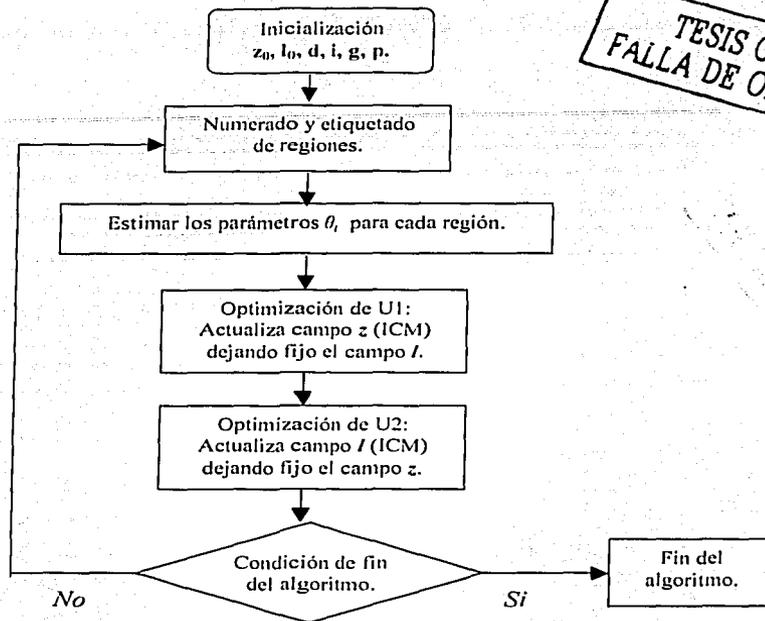


Fig 5.1 Diagrama de flujo del algoritmo de Segmentación Basada en Movimiento utilizado.

Con ayuda del histograma de la imagen, el algoritmo calcula el número de regiones conectadas y las etiqueta como regiones con movimiento distinto. Después de cada relajación se vuelve a enumerar y etiquetar, ya que la relajación puede generar o fusionar regiones. La aparición de nuevas regiones R se basa en el siguiente criterio:

- Si $R > 30$ píxeles: Se crea una nueva región y se etiqueta convenientemente.
- Si $R < 30$ píxeles: Se les asigna una etiqueta provisional para que en relajaciones posteriores adopten la etiqueta de la región vecina que minimice la energía en esos sitios.

Al terminar el cálculo de las regiones válidas, se calcula el vector de parámetros θ_i para cada región R_i con el modelo θ_{MLS} . Los sitios píxel s se visitan aleatoriamente, usando una tabla de inestabilidad, para evitar efectos direccionales o estructuras parásitas en la imagen de segmentación.

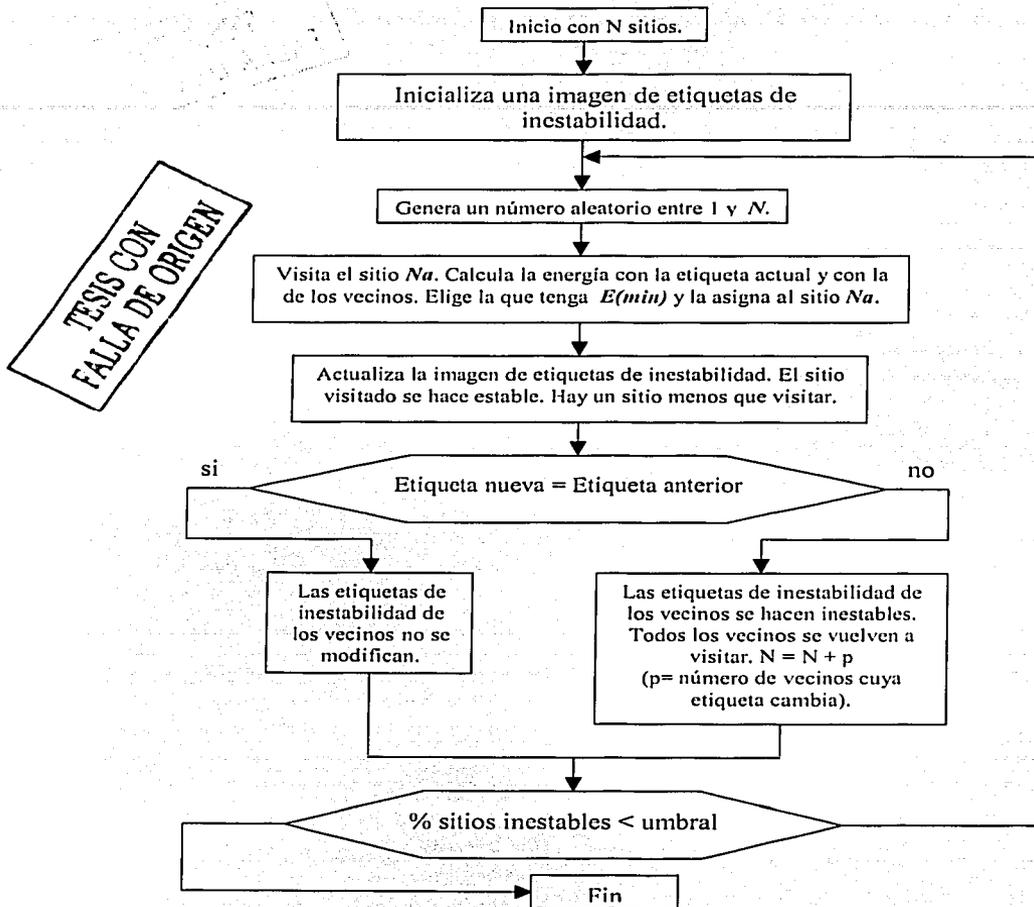


Fig 5.2 Diagrama de flujo del método de relajación.

El algoritmo finaliza cuando:

1. Se llega al número máximo de iteraciones descadas.
2. Se alcanza la convergencia, de acuerdo al número máximo de etiquetas de segmentación de líneas que cambian entre una iteración y otra.

5.2 SECUENCIAS DE PRUEBA.

5.2.1 SECUENCIA MISS AMÉRICA.

Es una secuencia real de videotelefonía de dimensiones (288x360, 8 bits/píxel), que muestra la cabeza y los hombros de una persona en movimiento sobre un fondo fijo que tiene poco contraste con su cabello. Contiene diferentes tipos de movimiento a lo largo de la secuencia; en ocasiones se presenta movimiento global, en otras el movimiento se concentra en los ojos y en la boca mientras que el torso y la cabeza se mantienen casi inmóviles.



TESIS CON
FALLA DE ORIGEN

Fig 5.3 Secuencia Miss América.

5.2.2 SECUENCIA CLAIRE.

Esta es una secuencia real de televisión de dimensiones (288 x 352, 8 bits/píxel), que muestra la cabeza y parte superior del torso de una mujer en movimiento sobre un fondo claro que contrasta con su cabello y ropa. En ésta secuencia, se presenta movimiento local y global, dominando el segundo sobre el primero.



Fig 5.4 Secuencia Claire.

5.3 SELECCIÓN DE LOS PARÁMETROS DEL MODELO DE ESTIMACIÓN.

El criterio de selección de los parámetros del algoritmo de basa en los resultados de trabajos anteriores [DGP96].

Selección del vector inicial de desplazamiento.

El vector inicial de desplazamiento d^0 es igual al d^1 previamente estimado para algún vecino espacial, de forma que produzca la DFD, más pequeña en valor absoluto para el píxel analizado.

Tamaño de la vecindad de observación N .

En nuestra experimentación se encontró que la vecindad de observación N para el método BD (Biemond) N , que produjo una estimación más estable, homogénea y robusta al ruido fue de $N=9$. El efecto de N es el de suavizar las características del campo de tal forma que:

- Si N es chico, entonces el suavizado local es débil y por lo tanto, el beneficio de la vecindad es bajo.
- Si N es GRANDE entonces el suavizado es excesivo, sobretodo en las fronteras de movimiento.

Selección del método de implantación del gradiente.

El método diferencial de estimación de movimiento es una modificación del operador para detección de bordes de Prewitt utilizado en trabajos anteriores [DGP96], [HARA92] y [BAAZ91a].

Para el cálculo de las derivadas en x y en y , respectivamente, se utilizaron las matrices:

$$\frac{1}{6} \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix} \quad \frac{1}{6} \begin{pmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

Número de iteraciones por píxel.

El valor que se utiliza como máximo es $it_{max}=1$, para mantener una buena coherencia del campo y una buena reconstrucción utilizando el método BD, que fue descrito en este trabajo en el punto 4.6.1.

Al incrementar este valor, la entropía H se incrementa rápidamente de igual manera en busca de una mejor reconstrucción y la DFD se minimiza, disminuyendo el error de reconstrucción y aumentando la entropía del campo. De tal forma que para mantener un equilibrio entre coherencia y reconstrucción se utiliza $it_{max}=1$.

Tamaño de la vecindad causal N_{vec} .

Los métodos de codificación predictiva asociados a los algoritmos pel-recursivos obligan a la elección de una vecindad causal de tamaño variable para la elección de d'' .

El tamaño de la vecindad espacial influye en:

- El desempeño de los métodos,
- la calidad del campo y en
- la reconstrucción de secuencias.

Al aumentar N_{vec} , también aumenta la entropía⁵ H y la Relación Señal a Ruido Pico (PSNR), debido a una mayor libertad en la condición inicial para converger a distintos mínimos locales de la DFD (mejor reconstrucción) pero introduce errores en la estimación del movimiento aparente.

Interpolación de la intensidad de las imágenes.

El campo de desplazamiento se representa en números flotantes para mantener una buena precisión numérica. Como es necesario calcular la DFD y V en la posición desplazada ($x-d_{xi}$, $y-d_{yi}$) se utilizan operaciones de interpolación de la intensidad de la imagen. Para la DFD se usa interpolación bilineal y V se calcula con redondeo del vector de desplazamiento [NETR79].

Selección de los umbrales.

Los umbrales T_{FD} y T_{DFD} son fijos y se proponen heurísticamente entre 2 y 4. Su valor depende del ruido y de la cantidad de movimiento presentes en cada secuencia [DGP96].

Píxeles no compensados.

A estos píxeles se les asocia cualquier valor de desplazamiento pues no se encontró ninguno que haya satisfecho el criterio de convergencia después del número máximo de iteraciones permitidas.

⁵ Entropía: Información promedio generada por una fuente. Proporciona el número mínimo de bits requerido para codificar su salida.

5.4 RESULTADOS DEL ALGORITMO DE ESTIMACIÓN DE MOVIMIENTO.

En la Figura 5.5 se muestra una visualización de los campos de desplazamiento estimados generados por el algoritmo Biernond BD en las tramas 3 y 4 de la secuencia *Miss América*.

Los campos generados contienen dos regiones de movimiento importante, uno en el rostro y otro en el cuello de la mujer. El movimiento en los ojos, boca y oreja izquierda es real, mientras que las regiones del cuello, cabello y hombro derecho de la mujer son resultado del movimiento aparente generado por las sombras de la escena.

El campo generado con preprocesamiento utilizando un filtro de erosión contiene campos de movimiento más concentrados en las zonas de los ojos, boca y cuello, pero aparecen unos vectores parásitos en el fondo. El movimiento aparente de la escena ha quedado casi eliminado excepto en la región del cabello y hombro derecho de la mujer.

El campo generado con el filtro máscara de enfoque es muy similar al que se tiene sin ningún preprocesamiento. Se observan las mismas regiones que en el original pero de una manera más acentuada. Se tiene una mejor delimitación de la silueta de la mujer en torso, rostro y cabello, además que los vectores de movimiento en las regiones de movimiento real de la escena son de mayor magnitud.

El campo generado con el filtro de dilatación presenta zonas donde los vectores de estimación de movimiento no coinciden con el movimiento real de la escena; aunque se alcanza a percibir la silueta de la mujer, no existen zonas definidas de movimiento.

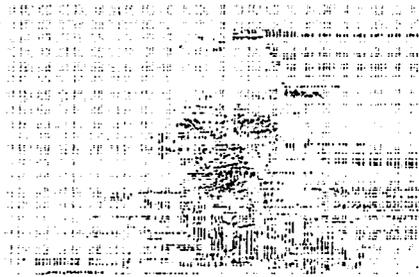
En general todos los campos de desplazamiento estimado, a excepción del que se obtiene con el filtro de dilatación, se pueden considerar buenos, pues reflejan de forma clara el movimiento real de la escena. Además, en el campo generado a partir de un preprocesamiento con el filtro de erosión el movimiento aparente es descartado, concentrándose solamente en los ojos, boca y en menor magnitud, en el cuello de la mujer.

En adelante, se comentarán los resultados de la estimación de movimiento en conjunto con la información de entrada al algoritmo de segmentación.

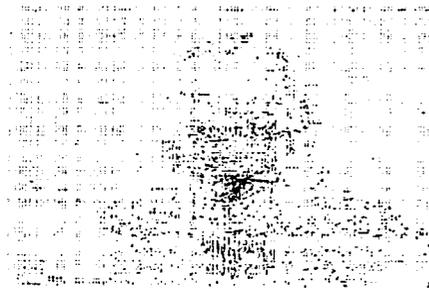
CAMPOS RESULTANTES DEL ALGORITMO DE ESTIMACIÓN DE MOVIMIENTO.



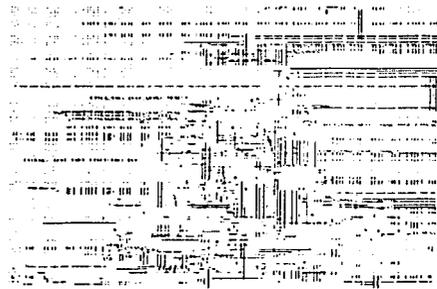
a) Secuencia original.



b) Secuencia con filtro de erosión.



c) Secuencia con filtro máscara de enfoque.



d) Secuencia con filtro de dilatación.

Fig 5.5 Campos de desplazamiento en direcciones x y y con la secuencia original y las secuencias preprocesadas. Para mejorar la visualización la escala de los vectores de movimiento se ha amplificado por 4.

TESIS CON
FALLA DE ORIGEN

5.5 SELECCIÓN DE LOS PARÁMETROS DEL MODELO DE SEGMENTACIÓN.

Los parámetros de la ponderación del modelo markoviano α , β , γ y κ , determinan la importancia que se le otorga a cada término de energía y la interacción local entre observaciones y etiquetas; por lo tanto, de estos parámetros depende el desempeño global del algoritmo.

En trabajos anteriores [DGP96] se utilizaron valores de $\alpha = 10$, $\beta = 10$, $\gamma = 3$ y $\kappa = 1$, los cuales fueron elegidos heurísticamente tomando como base la secuencia sintética *Seqmilt* en la cual aparecen 4 objetos de textura variada con movimiento de translación, rotación y divergencia pura sobre un tablero de ajedrez como fondo.

5.6 RESULTADOS DE LA APLICACIÓN DEL ALGORITMO DE SEGMENTACIÓN.

Se definió la segmentación como la partición de una imagen en regiones que pueden o no tener un significado relativo a la escena respectiva. Este proceso puede llevarse a cabo con base en una serie de modelos y criterios: por regiones, por bordes o líneas, por propiedades espectrales, por propiedades espaciales o por una combinación de estas dos últimas.

El algoritmo de segmentación basada en movimiento que se utilizó y que fue descrito en el punto 4.8 y 5.3.3 necesita cierta información de entrada u observaciones que se utilizan como condiciones iniciales, a partir de las cuales se obtendrán los resultados en la primera iteración.

a) Imagen original i .

La imagen original en este caso son las tramas 2 y 4 de la secuencia Miss América y la trama 7 de la secuencia Claire.

b) Máscara de segmentación inicial z_0 .

Para la obtención de ésta máscara se utilizó un detector de movimiento. El detector de movimiento calcula la diferencia (FD) entre la trama actual (t) y la anterior ($t-1$) en una vecindad N . De tal forma que si el valor absoluto de la FD es mayor a un umbral el píxel central de la ventana será 250 y de lo contrario será 0. El umbral se elige heurísticamente. En este caso el umbral se escogió de 30 para Miss América y 20 para Claire.

La elección del umbral es complicado y exige cierto conocimiento previo del movimiento de la secuencia. En particular para las tramas 3 y 4 de *Miss América* el movimiento es local y se concentra en la zona de los ojos, boca y barbilla, cuello, hombro izquierdo y cabello. Por lo tanto, se debe elegir una máscara en la cual se represente correctamente éste movimiento.

c) Campo de líneas de discontinuidad de movimiento l.

En éste caso se supone que los campos de líneas horizontales y verticales son dependientes entre sí, por lo que por facilidad se hacen iguales.

Para este campo se puede utilizar el campo de gradientes espaciales o las fronteras de la segmentación inicial. En este caso se usó la segunda opción porque el campo de líneas converge mas rápido en la segmentación.

d) Campo de gradientes espaciales g.

Este campo es el de los bordes espaciales. Para generarlo se usó el siguiente procedimiento: se pasó la imagen a través de un umbral, contra el que se comparan los niveles de gris de la imagen original, el resultado es entonces binario. A la imagen resultante se le encuentran las fronteras.

e) Campo de píxeles no compensados p.

Este campo es generado con el algoritmo de estimación de movimiento. Contiene los píxeles a los cuales se les asigno el desplazamiento que produjo la DFD más baja, pues no se encontró ninguno que haya satisfecho el criterio de convergencia al final de las iteraciones máximas permitidas.

f) Campo de desplazamiento estimado d.

Este campo es generado con el algoritmo pel-recursivo de estimación de movimiento, descrito en los puntos 4.6 y 4.7. Debe contener los campos de desplazamiento dx y dy para la estimación de desplazamiento aparente.

A continuación se presentan las observaciones y los resultados obtenidos con el algoritmo de segmentación.

5.6.1 SECUENCIA MISS AMÉRICA TRAMAS 3 Y 4.

Los resultados de la segmentación basada en movimiento después de 5 iteraciones muestran que el proceso en general ha sido bueno. Las regiones de segmentación presentan bordes redondeados y homogéneos además que representan el movimiento de forma aceptable en las regiones correspondientes a los ojos, oreja derecha, barbilla y cuello. Sin embargo, las regiones del cabello y hombro izquierdo, aunque se encuentran representadas, no son lo suficientemente grandes como para representar el movimiento que existe. Son demasiado pequeñas, en especial la del hombro izquierdo.

Al superponer las fronteras de la segmentación con la imagen original, Figura 5.7.b, tenemos que las regiones corresponden en buena medida al movimiento real de la secuencia. La única región que no se ajusta del todo a los contornos de movimiento es la región de la boca en su borde izquierdo.

Con respecto a los campos de líneas de discontinuidad de movimiento, Figura 5.7.c y 5.7.d, se observa que éstas corresponden a las fronteras del campo de segmentación obtenido.

Se tiene que en general el algoritmo de segmentación basada en movimiento funciona correctamente. Sin embargo, los resultados obtenidos no son tan buenos como se esperaba. Por lo tanto se propone hacer un preprocesamiento a las imágenes originales con el propósito de mejorar los resultados que se tienen.

Se proponen en principio dos tipos de preprocesamiento:

1. De tipo morfológico utilizando filtros de erosión y de dilatación.
2. Utilizando un filtro máscara de enfoque.

Los resultados de la segmentación con ambos preprocesamientos se comentan a continuación:

INFORMACIÓN DE ENTRADA AL ALGORITMO DE SEGMENTACIÓN.



a) Imagen original I .



b) Máscara de segmentación inicial z_0 .



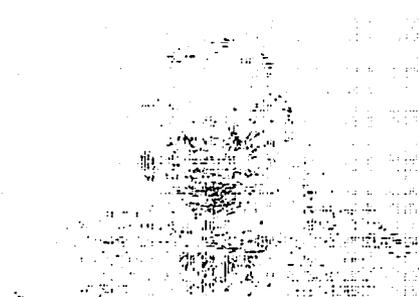
c) Campo de líneas L .



d) Campo de gradientes espaciales g .



e) Campo de píxeles no compensados p .



f) Campo de desplazamiento estimado d .

TESIS CON
FALLA DE ORIGEN

Fig 5.6 Información de entrada al algoritmo de segmentación basada en movimiento con la secuencia Miss América tramas 3 y 4. Esta información no tiene preprocesamiento de ningún tipo.

INFORMACIÓN PRODUCIDA POR EL ALGORITMO DE SEGMENTACIÓN.



a) Campo de etiquetas de segmentación 3.



b) Superposición de la trama 4 y las fronteras de segmentación.



c) Campo de líneas verticales final.



d) campo de líneas horizontales final.

Fig 5.7 Información producida por el algoritmo de segmentación basada en movimiento para las tramas 3 y 4 de la secuencia Miss América. Se observa la segmentación de la imagen en todas las zonas de movimiento real y aparente; sin embargo existen zonas, como la boca, cabello y hombro izquierdo, que no se cubren por completo o que no corresponden con el movimiento real de la escena.

TESIS CON
FALLA DE ORIGEN

INFORMACIÓN DE ENTRADA AL ALGORITMO DE SEGMENTACIÓN UTILIZANDO UN PREPROCESAMIENTO CON UN FILTRO DE EROSIÓN.



a) Imagen original i .



b) Máscara de segmentación inicial z_0 .



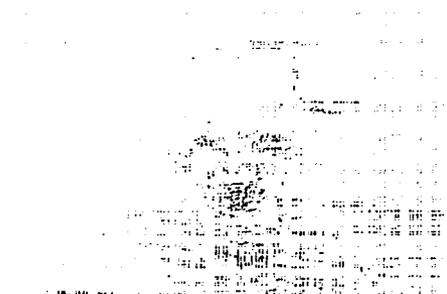
c) Campo de líneas l .



d) Campo de gradientes espaciales g .



e) Campo de píxeles no compensados p .

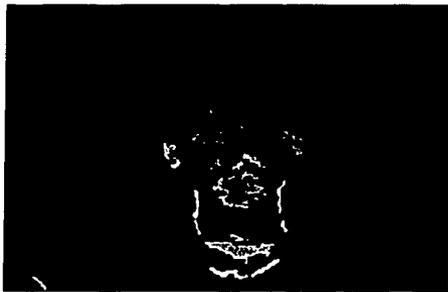


f) Campo de desplazamiento estimado d .

CENTRO DE INVESTIGACIONES Y DESARROLLOS TECNOLÓGICOS
 INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSOS
 DIRECCIÓN DE SISTEMAS DE INFORMACIÓN

Fig 5.8 Información de entrada al algoritmo de SBM con la secuencia Miss América tramas 3 y 4 después de aplicarles un preprocesamiento con un filtro de erosión.

*INFORMACIÓN PRODUCIDA POR EL ALGORITMO DE SEGMENTACIÓN
UTILIZANDO UN PREPROCESAMIENTO CON UN FILTRO DE EROSIÓN.*



a) Campo de etiquetas de segmentación z.



b) Superposición de la trama 4 y las fronteras de segmentación.



c) Campo de líneas verticales final.



d) Campo de líneas horizontales final.

Fig 5.9 Información producida por el algoritmo de segmentación basada en movimiento para las tramas 3 y 4 de Miss América. Se puede ver que la segmentación produce regiones sólo en ojos, nariz, boca y cuello, sin considerar el movimiento aparente del cabello y hombro izquierdo.

TESIS CON
FALLA DE ORIGEN

SECUENCIA PREPROCESADA CON UN FILTRO DE EROSIÓN.

Estos resultados se obtienen siguiendo los mismos parámetros que se establecieron en el caso anterior.

En la Figura 5.9.b se tiene una máscara de segmentación inicial donde se observa una mejor definición en regiones como los ojos, cuello y boca; sin embargo, regiones como la oreja derecha, cabello y hombro izquierdo no se encuentran bien definidas o son demasiado pequeñas como para poder asegurar que representan correctamente el movimiento en esa región de la imagen.

La máscara afecta directamente al campo de líneas, Figura 5.9.c, pues éste es generado con las fronteras de la máscara de segmentación inicial; mientras que el campo de gradientes espaciales, Figura 5.9.d, se mantiene más o menos similar al anterior en la región del rostro, pero se pierden los bordes de regiones como los hombros y parte de la cabeza.

El campo de píxeles no compensados, Figura 5.9.e, al parecer mejora, sobre todo en la región del chaleco, cuello, rostro y parte del cabello de la mujer. Aunque se presentan zonas rectangulares en las cuales no se pudieron compensar los píxeles con algún modelo de movimiento.

Los resultados del algoritmo de segmentación son de cierta forma muy pobres pues daría la impresión de que sólo se mueven los ojos y el área del cuello y de la boca, Figura 5.10.a. Se presenta deformación en la oreja derecha y desaparecen las regiones del hombro izquierdo y gran parte del cabello.

Al superponer la imagen original y las fronteras de la segmentación como se observa en la Figura 5.10.b se definen bien ciertos bordes como el del cuello, brazo derecho y sobre todo se define correctamente el área de la boca y barbilla; lo que no sucede con la oreja izquierda.

SECUENCIA PREPROCESADA CON UN FILTRO MÁSCARA DE ENFOQUE (*UNSHARP MASK*).

En cuanto a las observaciones de la secuencia con un preprocesamiento usando un filtro de máscara de enfoque (*unsharp mask*), se tiene que la imagen original se encuentra en cierta forma realzada en sus contrastes definiéndose más claramente los bordes. En general produce la impresión de ser una imagen más clara que la original, ver Figura 5.11.a.

La máscara de segmentación inicial, definida con los mismos parámetros que las anteriores, tiene más elementos, Figura 5.11.b. Se alcanza a definir, de cierta manera, el chaleco de la mujer y su frontera con el cuello. Además las zonas de mayor interés (ojos, boca, barbilla, cabello, cuello y hombro izquierdo) son más grandes y mejor definidas en comparación con las anteriores.

Este efecto se repite en el campo de líneas de discontinuidad de movimiento y en el de gradientes espaciales, Figuras 5.11.c y d, en el que se presentan más detalles de la imagen que en el mismo campo de la imagen original. Se tiene, sin embargo, la presencia de

regiones ruidosas sobretodo en la parte superior de la imagen, pero no son lo suficientemente grandes como para afectar los resultados de la segmentación.

El campo de los píxeles no compensados es muy similar al que se obtuvo con la imagen original, Figura 5.11.e.

Los resultados de la segmentación con este preprocesamiento, como se puede observar en la Figura 5.12.a, son superiores a los obtenidos con anterioridad. Las regiones de segmentación en regiones como el cabello, hombro izquierdo, cuello y rostro son mas grandes y casi no tienen filos en sus bordes.

Regiones como el cabello y hombro izquierdo son visiblemente mejoradas pues aparte de ser más grandes, se ajustan en forma y tamaño a las reales, como se puede ver en la Figura 5.12.b. La región de la boca también es mayor y abarca barbilla, ojos y pómulos, es decir, la región de mayor movimiento en el rostro. Sin embargo, se presentan regiones aisladas en la zona del chaleco que el algoritmo de segmentación no eliminó o que no pudo unir con otras regiones vecinas.

El campo de líneas, al igual que en los casos anteriores, concuerda con los bordes del campo de segmentación final en dirección vertical y horizontal, Figuras 5.12.c y 5.12.d.

SECUENCIA PREPROCESADA CON UN FILTRO DE DILATACIÓN.

En la Figura 5.13.b se observa que la máscara de segmentación inicial obtenida es muy similar a la que se tiene usando un preprocesamiento con un filtro de erosión; pero en este caso, las regiones son mas abundantes y mas grandes. Se tiene una buena definición de las regiones de movimiento: ojos, boca, cuello y cabello. La región del hombro izquierdo también está presente, aunque no es muy grande. Existe una diferencia muy marcada en la región de la nariz que es resultado del preprocesamiento, el cual resalta en buena medida la parte de la punta de la nariz.

El campo de líneas y de gradientes se mantiene casi con las mismas características que los anteriores, Figuras 5.13.c y 5.13.d; mientras que el campo de píxeles no compensados, Figura 5.13.e, presenta una zona rectangular en la boca de tamaño considerable en la cual no se encontró un modelo de movimiento adecuado.

La segmentación producida por el algoritmo es bastante buena y concuerda con el movimiento de la escena, Figura 5.14.a. Genera regiones en ojos, oreja derecha, parte de la oreja izquierda, mentón y cuello. Sin embargo, al superponerla con la imagen original, Figura 5.14.b, se tiene que dichas regiones no coinciden del todo con las partes del rostro de la mujer sobre todo en la oreja derecha, oreja izquierda y boca. Además de que la región del hombro izquierdo no describe lo suficientemente bien el movimiento de la secuencia.

Los campos de líneas conservan el comportamiento de las anteriores, siguiendo las fronteras de la segmentación final obtenida, Figuras 5.14.c y 5.14.d.

INFORMACIÓN DE ENTRADA AL ALGORITMO DE SEGMENTACIÓN CON
PREPROCESAMIENTO DE FILTRO MÁSCARA DE ENFOQUE (UNSHARP MASK).



a) Imagen original i .



b) Máscara de segmentación inicial z_0 .



c) Campo de líneas L .



d) Campo de gradientes espaciales g .



e) Campo de píxeles no compensados p .



f) Campo de desplazamiento estimado d .

Fig 5.10 Información de entrada al algoritmo de SBM con la secuencia Miss América tramas 3 y 4 después de aplicarle un preprocesamiento con un filtro máscara de enfoque.

TESIS CON
FALLA DE ORIGEN

*INFORMACIÓN PRODUCIDA POR EL ALGORITMO DE SEGMENTACIÓN
UTILIZANDO UN PREPROCESAMIENTO CON UN FILTRO MÁSCARA DE ENFOQUE
(UNSHARP MASK).*



a) Campo de etiquetas de segmentación z.



b) Superposición de la trama 4 y las fronteras de segmentación.



c) Campo de líneas verticales final.



d) Campo de líneas horizontales final.

Fig 5.11 Información producida por el algoritmo de segmentación basada en movimiento. Como se puede ver, existen más regiones segmentadas que abarcan casi la totalidad del rostro y partes del chaleco. Estos resultados cubren la totalidad de movimientos reales de la escena, incluyendo las zonas del cabello y hombro izquierdo.

TESIS CON
FALLA DE ORIGEN

INFORMACIÓN DE ENTRADA AL ALGORITMO DE SEGMENTACIÓN UTILIZANDO UN PREPROCESAMIENTO CON UN FILTRO DE DILATACIÓN.



a) Imagen original I .



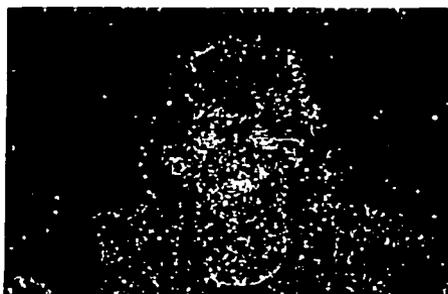
b) Máscara de segmentación inicial z_0 .



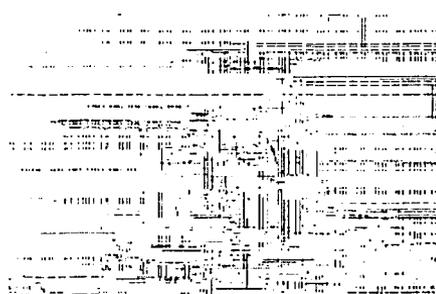
c) Campo de líneas l .



d) Campo de gradientes espaciales g .



e) Campo de píxeles no compensados p .



f) Campo de desplazamiento estimado d .

TESIS DE GRADO
 FALLA DE ORIGEN

Fig 5.12 Información de entrada al algoritmo de SBM con la secuencia Miss América tramas 3 y 4 después de procesarlas con un filtro de dilatación.

*INFORMACIÓN PRODUCIDA POR EL ALGORITMO DE SEGMENTACIÓN
UTILIZANDO UN PREPROCESAMIENTO CON UN FILTRO DE DILATACIÓN.*



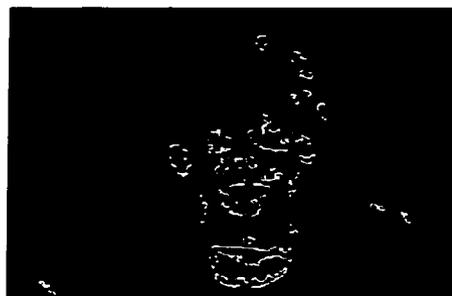
a) Campo de etiquetas de segmentación z .



b) Superposición de la trama z y las fronteras de segmentación.



c) Campo de líneas verticales final.



d) Campo de líneas horizontales final.

Fig 5.13 Información producida por el algoritmo de segmentación basada en movimiento. A diferencia de los resultados obtenidos con el filtro de erosión, éstos resultados incluyen regiones de segmentación en cabello y hombro izquierdo aunque no cubren por completo la zona de la boca.

TESIS CON
FALLA DE ORIGEN

5.6.2 SECUENCIA CLAIRE TRAMAS 6 Y 7.

A continuación se muestran los resultados de la aplicación del algoritmo de segmentación en las tramas 6 y 7 de la secuencia *Claire*.

En este caso, el movimiento de la mujer es global pues está moviendo un poco la cabeza de abajo hacia arriba y a la izquierda como se puede observar en la máscara de segmentación inicial de la Figura 5.15.b. En éste campo se aprecia claramente que el movimiento de la mujer incluye la cabeza y parte del torso, especialmente los hombros, cuello y pecho. Estos tres últimos elementos con movimiento resultan de cambios en la iluminación de la escena, creando sombras.

Los campos de gradiente espacial y de líneas de discontinuidad de movimiento, se generaron con el mismo procedimiento que se usó para la secuencia anterior. En éste caso, se puede observar en las Figuras 5.15.c y 5.15.d, se ponen de manifiesto ciertas regiones donde la iluminación de la escena no es homogénea, sobre todo en el campo de gradientes espaciales donde el fondo de la escena aparece de cierta forma ruidoso.

El campo de los píxeles no compensados, Figura 5.15.e, parece definir la silueta de la cabeza de la mujer pero sin concentrarse en ningún punto excepto en la zona comprendida entre la barbilla de la mujer y su labio inferior. Este campo se encuentra muy relacionado con el campo de los vectores de desplazamiento (Figura 5.15.f) donde, a pesar del ruido de fondo, se distingue la forma de la cabeza y las zonas de movimiento correspondientes a sombras.

Los resultados de la segmentación son bastante buenos. Se puede observar en la Figura 5.16.a, se tiene una región grande y homogénea que cubre casi toda la cabeza y el cuello; dos regiones pequeñas, una dentro de la cara y otra en el cuello; y otras tres regiones más que corresponden a un movimiento aparente en el cuello de la blusa, borde de la solapa del saco, y hombro izquierdo.

Al superponer la imagen original y los bordes de la segmentación final, Figura 5.16.b, se tiene que en efecto la segmentación abarca casi toda la cabeza de la mujer exceptuando la parte más externa del cabello cuyo movimiento relativo es menor al del resto de la cabeza.

SECUENCIA PREPROCESADA CON UN FILTRO DE EROSIÓN

Como se puede observar en la Figura 5.17.b, la máscara de segmentación inicial que se obtiene después de filtrar la imagen con un filtro de erosión es similar a la anterior, en cuanto a que el movimiento se concentra en la zona de la cabeza y parte de los hombros. Sin embargo, esta máscara no es del todo uniforme, pues el borde en la parte izquierda es muy irregular y no alcanza a seguir correctamente la forma de la cabeza en esta zona.

Los campos de líneas de discontinuidad y de gradientes mantienen características similares al anterior, Figuras 5.17.c y d, aunque en el segundo, la presencia de ruido ha disminuido parcialmente.

Los píxeles no compensados han disminuido en el fondo de la escena, no así en la mujer, donde se hallan muy concentrados en ciertas regiones que incluyen la zona derecha del cabello, ojo izquierdo, boca, cuello de la blusa y hombro izquierdo, ver Figura 5.17.e.

El resultado de la segmentación para un preprocesamiento con filtro de erosión no es muy bueno, Figura 5.18.a. Aunque abarca en gran parte las zonas de movimiento real y aparente de la escena, las fronteras de las regiones de segmentación son bastante irregulares. Además, se presenta el fenómeno de "corrimento" de una región, que está relacionado con que la máscara de segmentación inicial no tenga bordes regulares y con el campo de vectores de desplazamiento que, para esa zona, presenta una estimación de movimiento falsa.

En general, para un movimiento global, el preprocesamiento con filtro de erosión no parece de mucha utilidad.

SECUENCIA CON PREPROCESAMIENTO FILTRO MÁSCARA DE ENFOQUE (*UNSHARP MASK*):

Al aplicar una máscara de enfoque, al igual que en el caso anterior, la imagen parece más nítida y con los bordes resaltados, Figura 5.19.a. La máscara de segmentación inicial, a diferencia de las dos anteriores, es más homogénea pues casi no presenta huecos a su interior, es más grande, sus bordes son más continuos (no tienen quiebres) y con los mismos parámetros obtenemos más información en cuanto a bordes en el torso de la mujer. Sin embargo, el precio que se paga es la presencia de regiones de movimiento aparente en el fondo de la escena, sobretodo en la parte superior y a la derecha de la misma, Figura 5.19.b.

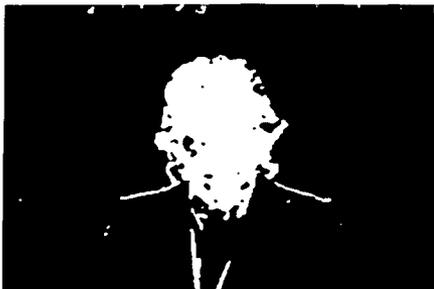
Esta presencia de movimiento aparente en el fondo de la escena, aparece reflejada en una u otra forma en todas las imágenes de entrada al algoritmo de segmentación. En el campo de líneas, Figura 5.19.c, se presenta como consecuencia natural de su generación a partir de la máscara z_0 . En el campo de gradientes, Figura 5.19.d, se manifiesta como zonas ruidosas en el fondo de la escena, también debidas en parte a la iluminación no uniforme de la escena como ya se había comentado. El campo de píxeles no compensados, es muy similar al de la secuencia sin preprocesamiento, aunque contiene más elementos en la escena, Figura 5.19.e.

La segmentación final para este filtro es buena en cuanto a que considera una única zona de movimiento global que abarca casi la totalidad de la cabeza de la mujer, aunque los bordes de la segmentación no son tan regulares como se esperaba, Figura 5.20.a. Se tienen además pequeñas regiones de movimiento en el hombro y brazo izquierdo, en el borde de la solapa del saco y en el brazo derecho. Esta segmentación incluye las regiones de movimiento aparente en la parte derecha y superior del fondo de la escena.

INFORMACIÓN DE ENTRADA AL ALGORITMO DE SEGMENTACIÓN.



a) Imagen original i .



b) Máscara de segmentación inicial z_0 .



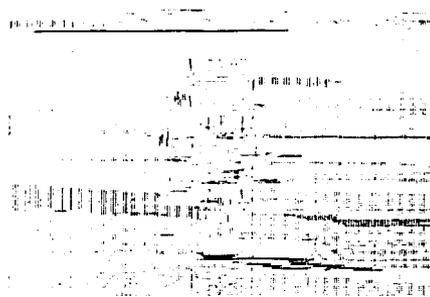
c) Campo de líneas l .



d) Campo de gradientes espaciales g .



e) Campo de píxeles no compensados p .

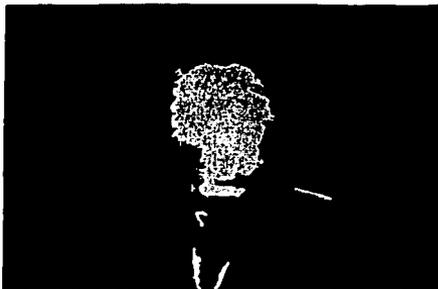


f) Campo de desplazamiento estimado d .

TESIS CON
FALLA DE ORIGEN

Fig 5.14 Información de entrada al algoritmo de SBM con la secuencia Claire tramas 6 y 7 sin utilizar ningún filtro para su preprocesamiento. En la máscara de segmentación se observa que el movimiento en la cabeza es global.

INFORMACIÓN PRODUCIDA POR EL ALGORITMO DE SEGMENTACIÓN.



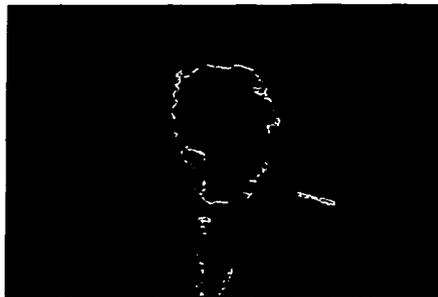
a) Campo de etiquetas de segmentación z .



b) Superposición de la trama T y las fronteras de segmentación.



c) Campo de líneas verticales final.



d) Campo de líneas horizontales final.

Fig 5.15 Información producida por el algoritmo de segmentación basada en movimiento. Aquí el algoritmo segmenta la región de la cabeza, incluyendo regiones de sombras en la mejilla derecha y parte izquierda del cuello.

TESIS CON
FALLA DE ORIGEN

INFORMACIÓN DE ENTRADA AL ALGORITMO DE SEGMENTACIÓN UTILIZANDO UN PREPROCESAMIENTO CON UN FILTRO DE EROSIÓN.



a) Imagen original i .



b) Máscara de segmentación inicial z_0 .



c) Campo de líneas l .



d) Campo de gradientes espaciales g .



e) Campo de píxeles no compensados p .

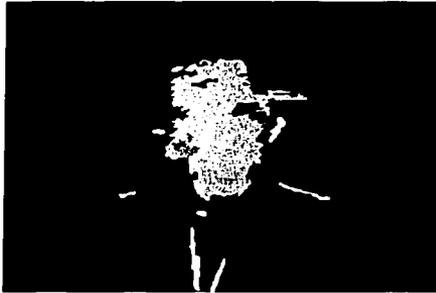


f) Campo de desplazamiento estimado d .

Fig 5.16 Información de entrada al algoritmo de SBM con la secuencia Claire tramas 6 y 7 después de preprocesarla con un filtro de erosión.

TESIS CON
FALLA DE ORIGEN

INFORMACIÓN PRODUCIDA POR EL ALGORITMO DE SEGMENTACIÓN
UTILIZANDO UN PREPROCESAMIENTO CON UN FILTRO DE EROSIÓN.



a) Campo de etiquetas de segmentación z .



b) Superposición de la trama 7 y las fronteras de segmentación.



c) Campo de líneas verticales final.



d) Campo de líneas horizontales final.

Fig 5.17 Información producida por el algoritmo de segmentación basada en movimiento de Claire con preprocesamiento con un filtro de erosión. Se observa en a) y b) la zona de "corrimiento" además de que los bordes de las regiones de segmentación no son suaves, pues presentan muchos quiebres.

TESIS CON
FALLA DE ORIGEN

SECUENCIA CON PREPROCESAMIENTO FILTRO DE DILATACIÓN.

La máscara de segmentación inicial, a diferencia de la anterior, ya no tiene el movimiento aparente del fondo de la escena; aunque los bordes de la región de la cabeza son muy irregulares, Figura 5.21.b. En el campo de gradientes, el ruido del fondo de la escena ha disminuido parcialmente, al igual que cuando se aplicó un filtro de erosión, Figura 5.21.d. Los píxeles no compensados se concentran en la parte superior del cabello, ojo derecho, cuello de la blusa y la parte entre la barbilla y el labio inferior de la mujer, Figura 5.21.e.

El resultado de la segmentación es en cierta forma similar al que se obtuvo directamente de la secuencia original con una región de movimiento que abarca casi toda la cabeza, dividida en su interior en cuatro regiones pequeñas que cubren la mejilla derecha, parte de la barba y el cuello, que se divide a su vez en dos secciones, ver Figura 5.22.a. Parte de esta segmentación se debe al movimiento aparente producido por las sombras en la mejilla y cuello de la mujer.

Cabe mencionar que en este resultado también aparece el fenómeno de "corrimiento" de una región en la misma zona que en el caso del filtro de erosión pero en menor medida.

*INFORMACIÓN DE ENTRADA AL ALGORITMO DE SEGMENTACIÓN CON
PREPROCESAMIENTO FILTRO MÁSCARA DE ENFOQUE (UNSHARP MASK).*



a) Imagen original i .



b) Máscara de segmentación inicial z_0 .



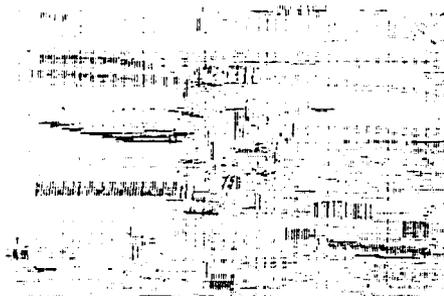
c) Campo de líneas l .



d) Campo de gradientes espaciales g .



e) Campo de píxeles no compensados p .



f) Campo de desplazamiento estimado d .

TESIS CON
FALLA DE ORIGEN

Fig 5.18 Información de entrada al algoritmo de SBM con la secuencia Claire tramas 6 y 7. Se observa que al realzar la imagen, se hacen más evidentes los cambios de iluminación de la escena.

INFORMACIÓN PRODUCIDA POR EL ALGORITMO DE SEGMENTACIÓN CON
PREPROCESAMIENTO FILTRO MÁSCARA DE ENFOQUE (UNSHARP MASK).



a) Campo de etiquetas de segmentación z .



b) Superposición de la trama 7 y las fronteras de segmentación.



c) Campo de líneas verticales final.



d) Campo de líneas horizontales final.

Fig 5.19 Información producida por el algoritmo de segmentación basada en movimiento. El algoritmo de segmentación, en este caso, no ha sido capaz de eliminar las regiones de movimiento aparente del fondo de la escena.

TESIS CON
FALLA DE ORIGEN

INFORMACIÓN DE ENTRADA AL ALGORITMO DE SEGMENTACIÓN UTILIZANDO UN PREPROCESAMIENTO CON UN FILTRO DE DILATACIÓN.



a) Imagen original i .



b) Máscara de segmentación inicial z_0 .



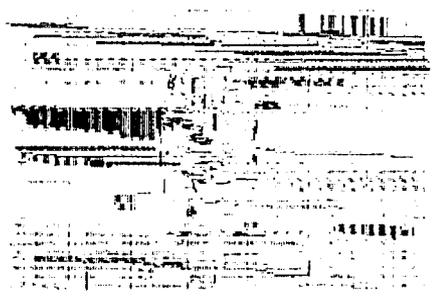
c) Campo de líneas l .



d) Campo de gradientes espaciales g .



e) Campo de píxeles no compensados p .

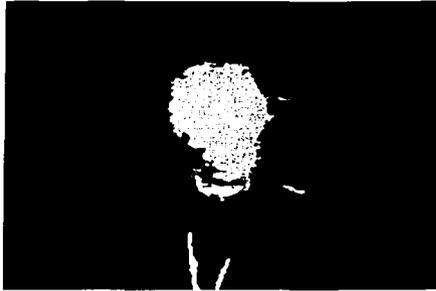


f) Campo de desplazamiento estimado d .

Fig 5.20 Información de entrada al algoritmo de SBM con la secuencia Claire tramas 6 y 7 al aplicarle un filtro de dilatación.

TESIS CON
FALLA DE ORIGEN

*INFORMACIÓN PRODUCIDA POR EL ALGORITMO DE SEGMENTACIÓN
UTILIZANDO UN PREPROCESAMIENTO CON UN FILTRO DE DILATACIÓN.*



a) Campo de etiquetas de segmentación z .



b) Superposición de la trama 7 y las fronteras de segmentación



c) Campo de líneas verticales final.



d) Campo de líneas horizontales final.

Fig 5.21 Información producida por el algoritmo de segmentación basada en movimiento. Se observa una pequeña zona de corrimiento en la parte superior izquierda de la cabeza y más regiones de segmentación en la parte inferior del rostro y parte del cuello.

TESIS CON
FALLA DE ORIGEN

5.6.3 SECUENCIA MISS AMÉRICA TRAMAS 1 Y 2.

Las tramas 1 y 2 de la secuencia Miss América presentan un movimiento global en cuanto a que la mujer mueve la cabeza de un lado hacia otro. Las condiciones de iluminación son similares y existe poco contraste entre el cabello y blusa de la mujer y el fondo de la escena.

La máscara de segmentación inicial revela que existe un movimiento de tipo translacional en este par de tramas, lo cual se pone de manifiesto en la presencia de movimiento en la parte lateral del chaleco y no en la parte superior correspondiente a los hombros, ver Figura 5.23.b.

El valor del umbral que se escogió en ésta máscara fue de 30, al igual que en las tramas 3 y 4, pero ahora las regiones de movimiento cambian. El movimiento de la secuencia se concentra en la totalidad del rostro (no solamente en regiones como los ojos o la boca), y gran parte de la cabeza, sobretudo la parte izquierda del cabello. El torso en cambio, presenta poco movimiento y sólo se distinguen algunos bordes del chaleco y blusa de la mujer.

El campo de gradientes espaciales es muy similar al que se tiene en las tramas 3 y 4 pero en este caso se observa en la parte superior una zona de ruido que corresponde a un movimiento aparente de la escena que también aparece en la máscara de segmentación inicial, Figura 5.23.d.

En el campo de píxeles no compensados se tiene la silueta de la mujer y del chaleco, pero sin que se tengan concentraciones importantes de éstos en ninguna zona. Aquí también se alcanza a distinguir la zona de ruido en la parte superior, Figura 5.23.e. Sin embargo, en el campo de los vectores de desplazamiento no hay ningún vector que describa movimiento en esa zona. Los vectores se concentran principalmente en las fronteras de la silueta de la mujer sin definir ninguna región en específico, Figura 5.23.f.

Los resultados de la segmentación no son muy buenos, ver Figura 5.24.a y 5.24.b, porque no abarcan la totalidad del rostro y existen muchas regiones pequeñas en el cabello que no son de utilidad para describir el tipo de movimiento que se está presentando en ésta escena.

SECUENCIA PREPROCESADA CON UN FILTRO DE EROSIÓN.

La máscara que resulta de usar un filtro de erosión es más detallada en cuanto a regiones como la boca y el borde entre el rostro y el cabello de la frente, Figura 5.25.b. Los bordes en las fronteras de movimiento en este caso, Figura 5.25.c, son muy suaves y sin quiebres abruptos. Sin embargo, esta máscara no está describiendo de forma correcta el movimiento de la escena en su totalidad, pues aunque en la escena se presenta un movimiento local de ojos y boca, éste tiene una magnitud menor al movimiento translacional de la cabeza de la mujer, el cual debiera ser dominante en la máscara de segmentación.

Los campos de líneas de discontinuidad y de gradientes espaciales se mantienen con las mismas características que los anteriores sólo que en el segundo la imagen es un poco menos detallada que antes pero se conservan las fronteras fundamentales de la escena, Figura 5.25.d.

Además, el campo de píxeles no compensados es un poco menos denso que el de la secuencia original aunque se puede distinguir una zona considerable en la zona de la boca, como se ve en la Figura 5.25.e. El campo de vectores de desplazamiento no es muy bueno pues existen regiones con vectores que no corresponden a un movimiento real de la escena; sin embargo, aún se alcanza a distinguir la silueta de la mujer, Figura 5.25.f.

El resultado de la segmentación después de 5 iteraciones es muy similar a la máscara de segmentación inicial aunque no en su totalidad. El algoritmo de segmentación fue capaz en este caso de acrecentar las regiones que se le proporcionaron inicialmente, hasta que se juntaron la región del cabello y la del rostro (en la parte izquierda de la mujer), ver Figura 5.26.a.

Existe otro fenómeno de "corrimiento" en la imagen a la altura del cuello de la mujer, pero en este caso se debe a dos cosas distintas. Una parte de este corrimiento está delimitando una sombra en el chaleco y la otra una región en el cuello de la mujer. En ambos casos esto es muy probable que se deba al movimiento aparente de la escena.

SECUENCIA PREPROCESADA CON UN FILTRO MÁSCARA DE ENFOQUE (*UNSHARP MASK*).

El filtro máscara de enfoque al acentuar los bordes y el contraste de la imagen hace posible detectar más regiones que con los filtros morfológicos. De esta forma, utilizando los mismos valores de umbral y ventana que en los casos anteriores obtenemos una máscara muy buena que abarca casi la totalidad de la cabeza de la mujer excepto en la región donde el contraste del cabello con el fondo de la escena es prácticamente nulo, Figura 5.27.b. Además se tienen más regiones en la zona del chaleco que ayudarán en buena medida a describir el movimiento aparente del torso de la mujer.

De la misma manera, el campo de gradientes es más detallado en el rostro y chaleco de la mujer, Figura 5.27.d. El campo de píxeles no compensados muestra una distribución homogénea de puntos en el chaleco, cuello y rostro de la mujer con regiones de concentración en los ojos y boca, pero sin que lleguen a ser de tamaño considerable, Figura 5.27.c. El campo de vectores de desplazamiento al igual que en el caso de la secuencia original, muestra vectores que delimitan la silueta de la mujer y pequeñas regiones de vectores que describen un movimiento erróneo a lo largo del cuerpo y en la región de fondo de la escena, Figura 5.27.f.

INFORMACIÓN DE ENTRADA AL ALGORITMO DE SEGMENTACIÓN.



a) Imagen original i .



b) Máscara de segmentación inicial z_0 .



c) Campo de líneas l .

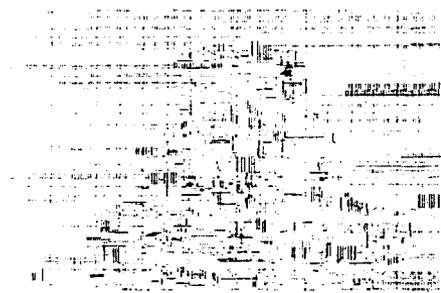


d) Campo de gradientes espaciales g .

TESIS CON
FALLA DE ORIGEN



e) Campo de píxeles no compensados p .



f) Campo de desplazamiento estimado d .

Fig 5.22 Información de entrada al algoritmo de SBM con la secuencia Miss América tramas 1 y 2.

INFORMACIÓN PRODUCIDA POR EL ALGORITMO DE SEGMENTACIÓN.



a) Campo de etiquetas de segmentación z .



b) Superposición de la trama 2 y las fronteras de segmentación.



c) Campo de líneas verticales final.



d) Campo de líneas horizontales final.

Fig 5.23 Información producida por el algoritmo de segmentación basada en movimiento. Se presenta un movimiento global de la escena. Las regiones de segmentación abarcan la parte iluminada de la mujer, sin segmentar regiones como la de la parte superior derecha de la cabeza.

TESIS CON
FALLA DE ORIGEN

INFORMACIÓN DE ENTRADA AL ALGORITMO DE SEGMENTACIÓN UTILIZANDO
UN PREPROCESAMIENTO CON UN FILTRO DE EROSIÓN.



a) Imagen original i .



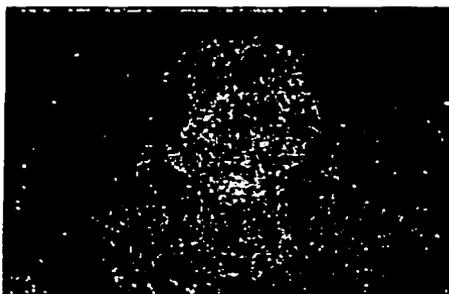
b) Máscara de segmentación inicial z_0 .



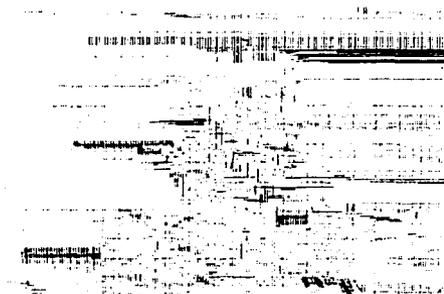
c) Campo de líneas l .



d) Campo de gradientes espaciales g .



e) Campo de píxeles no compensados p .



f) Campo de desplazamiento estimado d .

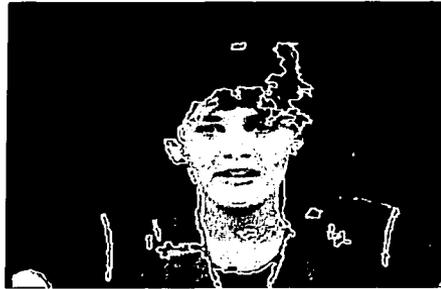
TESIS CON
FALLA DE ORIGEN

Fig 5.24 Información de entrada al algoritmo de SBM con la secuencia Miss América tramas 1 y 2 tras aplicarle un filtro de erosión.

*INFORMACIÓN PRODUCIDA POR EL ALGORITMO DE SEGMENTACIÓN
UTILIZANDO UN PREPROCESAMIENTO CON UN FILTRO DE EROSIÓN.*



a) Campo de etiquetas de segmentación z .



b) Superposición de la trama 2 y las fronteras de segmentación.



c) Campo de líneas verticales final.



d) Campo de líneas horizontales final.

Fig 5.25 Los campos resultantes del algoritmo de segmentación a la secuencia después de pasarla a través de un filtro de erosión representan pobremente el movimiento real de la escena.

TESIS CON
FALLA DE ORIGEN

Como se puede observar en las Figuras 5.28.a y b, el resultado de la segmentación es bueno porque abarca casi la totalidad del rostro de la mujer sobretodo en áreas de interés como son: ojos nariz, boca, pómulos y barbilla de la mujer. Sin embargo, la parte inferior de esta segmentación es irregular, esto es, en la parte inferior de las mejillas. Las regiones del chaleco describen correctamente las zonas donde existe movimiento aparente debido a sombras por los pliegues de la ropa. La región del cabello que corresponde a la parte iluminada, se ha unido con la que describe el movimiento del rostro y además se han creado pequeñas regiones en la parte superior de la cabeza.

SECUENCIA PREPROCESADA CON UN FILTRO DE DILATACIÓN.

La máscara de dilatación define mejor las regiones del rostro como la boca ojos y orejas y abarca en mayor proporción la parte iluminada del cabello y las regiones en la parte superior de la cabeza que la máscara de erosión, Figura 5.29.b. Aunque no abarca la totalidad de la cabeza o del rostro, ésta segmentación inicial cubre regiones importantes de la cara.

El efecto del preprocesamiento también es evidente en el campo de gradientes que es en cierta forma, el menos detallado de todos los anteriores; pero aún así, describe correctamente los bordes espaciales principales de la escena, Figura 5.29.d.

El campo p mantiene una distribución uniforme de puntos y solo se observan pequeñas regiones de concentración que no están en ninguna zona en particular, Figura 5.29.e; el campo de los vectores de desplazamiento es también bastante bueno pues describe un movimiento global de la mujer sin ruido en el fondo de la escena de magnitud importante, Figura 5.29.f.

La segmentación final ha agrandado ciertas regiones del rostro de la cara, sobretodo las de la parte de la frente, la cual ha crecido hasta unirse con la zona del cabello que se encuentra iluminada. Sin embargo, no pudo hacer lo mismo con la parte inferior del rostro que abarca la nariz y la boca, dejando fuera los pómulos, y en otra región separada a la barbilla. Las regiones de bordes en el cuello, blusa y chaleco se encuentran bien delimitadas y describen un movimiento real, Figuras 5.30.a y b.

Por lo que respecta a los campos finales de líneas verticales y horizontales, Figuras 5.30.c y d, cabe decir que en todos los casos éstos campos corresponden a las fronteras de la segmentación final en sus componentes vertical y horizontal respectivamente.

INFORMACIÓN DE ENTRADA AL ALGORITMO DE SEGMENTACIÓN CON UN
PREPROCESAMIENTO FILTRO MÁSCARA DE ENFOQUE (UNSHARP MASK).



a) Imagen original i .



b) Máscara de segmentación inicial z_0 .



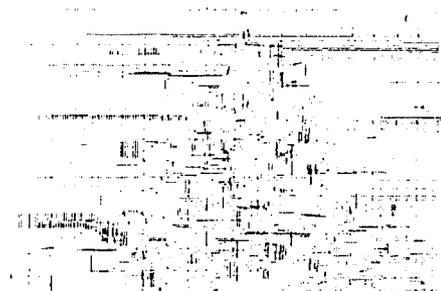
c) Campo de líneas l .



d) Campo de gradientes espaciales g .



e) Campo de píxeles no compensados p .



f) Campo de desplazamiento estimado d .

TESIS CON
... DE ORIGEN

Fig 5.26 Información de entrada al algoritmo de SBM de la secuencia Miss América tramas 1 y 2. En b) y c) se observa un incremento considerable de regiones, producto del preprocesamiento.

*INFORMACIÓN PRODUCIDA POR EL ALGORITMO DE SEGMENTACIÓN CON UN
PREPROCESAMIENTO FILTROFILTRO MASCARA DE ENFOQUE (UNSHARP MASK).*



a) Campo de etiquetas de segmentación z .



b) Superposición de la trama z y las fronteras de segmentación.



c) Campo de líneas verticales final.



d) Campo de líneas horizontales final.

Fig 5.27 La segmentación producida se centra en dos grandes regiones que abarcan el rostro y cabello y el cuello. Existen, además, regiones producto del movimiento aparente (sombras) en el chaleco de la mujer.

TESIS DON
FALLA DE ORIGEN

INFORMACIÓN DE ENTRADA AL ALGORITMO DE SEGMENTACIÓN UTILIZANDO UN PREPROCESAMIENTO CON UN FILTRO DE DILATACIÓN.



a) Imagen original i .



b) Máscara de segmentación inicial z_0 .



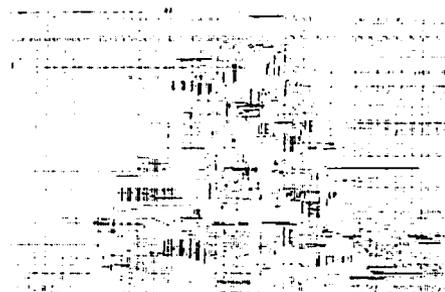
c) Campo de líneas l .



d) Campo de gradientes espaciales g .



e) Campo de píxeles no compensados p .



f) Campo de desplazamiento estimado d .

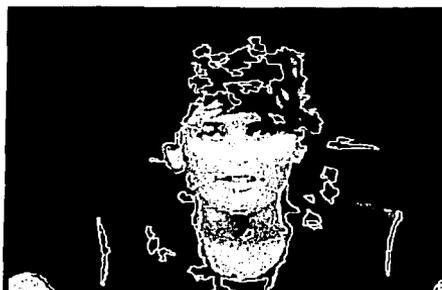
FALLA DE ORIGEN

Fig 5.28 Información de entrada al algoritmo de SBM de la secuencia Miss América tramas 1 y 2 preprocesada con un filtro de dilatación.

*INFORMACIÓN PRODUCIDA POR EL ALGORITMO DE SEGMENTACIÓN
UTILIZANDO UN PREPROCESAMIENTO CON UN FILTRO DE DILATACIÓN.*



a) Campo de etiquetas de segmentación z.



b) Superposición de la trama z y las fronteras de segmentación.



c) Campo de líneas verticales final.



d) Campo de líneas horizontales final.

Fig 5.29 Información producida por el algoritmo tras aplicarle un filtro de dilatación, presenta zonas muy irregulares en la zona superior de la cabeza.

INFORMACIÓN
DE ORIGEN

Capítulo 6

Conclusiones.

TESIS CON
FALLA DE ORIGEN

Los resultados de la segmentación están directamente relacionados con los de la estimación, de forma tal que si se tiene una estimación mala o pobre, la segmentación lo será también. Al definirse mejor un contorno o zona que presente movimiento se puede hacer una estimación más precisa, lo que nos lleva a obtener regiones de segmentación que corresponden de manera más cercana al movimiento real.

La segmentación es evaluada cualitativamente a lo largo de todo el trabajo; por lo que los resultados de la secuencia sin preprocesamiento, aunque son considerados buenos, no describen la totalidad del objeto en movimiento como se ve en la Figura 6.1, en donde existen regiones que no describen adecuadamente el movimiento de objetos significativos, como la boca, o de objetos importantes que proporcionan al espectador la impresión de que se trata de un movimiento continuo (vídeo), como los del cabello y el hombro.



Fig 6.1 Resultados de la segmentación sin preprocesamiento de la secuencia Miss América tramas 3 y 4.

El objetivo de haber incluido un preprocesamiento para resaltar los elementos de la escena, es facilitar la identificación de los objetos que la componen para obtener una estimación congruente con la estructura y movimiento de dichos elementos y a su vez mejorar la forma de las regiones de segmentación.

Para evaluar los efectos de los 3 diferentes filtros utilizaremos como referencia los resultados de la segmentación sin preprocesamiento.

En general, se analizan dos tipos de movimiento: uno local, en donde la persona presente en la escena mueve solamente los ojos y la boca mientras que el resto del cuerpo permanece casi inmóvil. El otro tipo de movimiento es aquel al que llamaremos global porque, además

de que presenta movimiento de ojos y boca, el movimiento predominante se presenta en el torso y cabeza de la persona.

En el movimiento local de una secuencia de vídeo las principales regiones donde se concentra la información (movimiento) son muy delimitadas y la información espacial de la escena es muy redundante. Esto ocasiona que se puedan tener tasas de compresión muy altas, pues la información relevante es muy poca si se compara con la información de la totalidad de la escena. Por lo tanto, lo que nos interesa en este caso es que la segmentación refleje de manera más fiel la forma de las regiones en movimiento y que el número de éstas sea el suficiente para representar adecuadamente los cambios de la escena.

En este sentido los filtros morfológicos tienen un desempeño superior al del filtro de realce, pues a diferencia del segundo, éstos filtros enfatizan su acción en los rasgos de la imagen resultando en regiones de segmentación separadas que permiten que el modelo de movimiento asociado a cada una de ellas corresponda al movimiento real de la región.

En el caso particular de los filtros de erosión y dilatación que estudiamos en este trabajo vemos que, aunque en los dos se presenta la separación de regiones, en uno de ellos hay más que en el otro. Esto es debido a las características particulares de cada uno; así entonces, consideramos que los dos tienen un desempeño adecuado, aunque para propósitos de compresión de información, el adecuado sería el de erosión ya que es el que tiene menos regiones; además de que en nuestro caso particular consideramos que para representar de forma suficiente un movimiento local no es tan necesario incluir movimientos como el del cabello.

En el movimiento global de una secuencia, la información no se concentra en algunas regiones en específico, sino que está localizada en una región grande que puede abarcar todo el movimiento presente en la escena. Con esto, ganamos simplicidad en la descripción de movimiento y a su vez suponemos que generará tasas de compresión grandes. Bajo este criterio el filtro que generó mejores resultados fue el de máscara de enfoque pues a diferencia de los morfológicos, la segmentación producida es más global y las subregiones contenidas dentro de ésta, corresponden a cambios en la iluminación que pueden ser sombras o brillos.

De lo anterior podemos concluir que el tipo de preprocesamiento que se debe aplicar depende del movimiento presente en la escena (global o local), pues en general todos los resultados obtenidos proporcionan información que corresponde en mayor o menor medida al movimiento observado. Sin embargo, la elección del filtro depende en gran parte de la aplicación final de la secuencia procesada.

Este trabajo se enfocó al estudio de secuencias donde aparece el torso y el rostro humano con movimientos lentos, por lo que se sugiere como un punto de partida para aplicaciones tales como la videoconferencia, el reconocimiento de patrones o el vídeoteléfono.

Anexo 1

Métodos de Compresión de Secuencias de Imágenes.

A.1. Fundamentos.

Dado que todos los sistemas de comunicaciones presentan en su medio de transmisión la limitante de una determinada tasa de transmisión, con la aparición de servicios que aceleran la obtención de información el deseo por incrementar esta velocidad se hace mayor y el principal medio por el que podemos hacerlo posible lo representa la compresión.

El objetivo de la operación de compresión en este caso de imágenes es reducir la cantidad de datos para hacer más eficiente su transporte y/o almacenamiento.

A.1.1. ¿QUÉ ES COMPRESIÓN?

Muchas de las señales sensoriales contienen una cantidad importante de información redundante o superflua y la eliminación de esa información permite reducir el tiempo de transmisión, así como la cantidad de espacio necesario para su almacenamiento. Es decir, la *compresión* se refiere a la eliminación de esa información redundante.

Así entonces las señales de imagen, audio y vídeo son susceptibles a la compresión debido a los siguientes factores [BHKO97]:

- ✓ Las señales de audio y vídeo digital contienen una cantidad significativa de *redundancia estática*.
 1. Esto es, muestras similares a otras, *correlación espacial*, de tal forma que una muestra puede ser predicha con bastante precisión desde otra vecina. Se trata de extraer la información esencial de la imagen, descartando la información no esencial, de manera que la imagen pueda ser posteriormente reconstruida con precisión (operación de descompresión).
 2. El segundo tipo de datos superfluos es la redundancia perceptual, información que el sistema audio-visual humano no puede escuchar ni ver, *correlación espectral*. A diferencia de la redundancia estática, la información removida basada en las limitaciones de la percepción humana no se puede recuperar. Desafortunadamente, la percepción humana es muy compleja, varía de persona a persona, y depende del contexto y la aplicación.

3. Por último para datos temporales, como el video, existe una correlación significativa entre muestras en diferentes segmentos de tiempo. Esto se refiere a la correlación temporal.
- ✓ Algunos datos tienden a tener características de alto nivel que son redundantes a través del espacio y del tiempo, es decir, los datos son de una naturaleza fractal.

Para una aplicación dada, los programas de compresión pueden explotar uno o todos los factores descritos para alcanzar la tasa de compresión deseada.

Un sistema que realiza la compresión puede ser descrito de la siguiente forma:

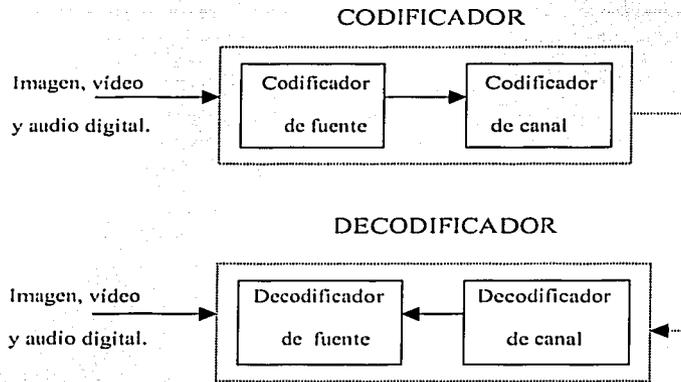


Fig a.1 Sistema general de compresión.

Tanto en el codificador como en el decodificador de fuente se involucran tres operaciones distintas aunque interrelacionadas: *Representación*, *Quantización* y *Asignación del Código de palabras*.

El núcleo del codificador es el codificador de fuente. El codificador de fuente presenta el proceso de compresión mediante la reducción de la tasa de entrada de datos a un nivel que puede ser soportada por el medio de transmisión o almacenamiento. La tasa de salida de bits del codificador es medida en bits por muestra o bits por segundo. Además de la *tasa de bits* o otro término que permite caracterizar la capacidad de compresión del sistema es la *relación de compresión*, denotada c_r , cuya definición intuitiva es:

$$c_r = \frac{\text{tamaño de la entrada al codificador fuente}}{\text{tamaño de la salida del codificador fuente}}$$

Esta definición es un tanto ambigua y depende del tipo de datos y el método de compresión empleado. En el caso del video comúnmente la idea de tamaño usada es de los bits necesarios para representar un segundo de video.

Regularmente en un sistema real el codificador de fuente es seguido por un segundo nivel de codificación: la codificación de canal, esta traduce la corriente de bits comprimidos en una señal protegida contra el ruido que se presentará durante la transmisión (control de errores).

El problema de la compresión se puede reducir a un problema de minimización de la tasa de datos, donde varias limitantes pueden ser conocidas, incluyendo las siguientes:

- × Especificación de niveles de calidad de la señal. Aplicada comúnmente en el decodificador.
- × Implementación compleja. Frecuentemente aplicada en el decodificador y en algunas ocasiones en ambos: codificador y decodificador.
- × Retraso en la comunicación. Se refiere al retraso de punto-a-punto, y es medido desde que una muestra empieza a ser codificada hasta que la misma completa la codificación.

La compresión es una operación reversible, es decir, es posible la descompresión como se ilustra:

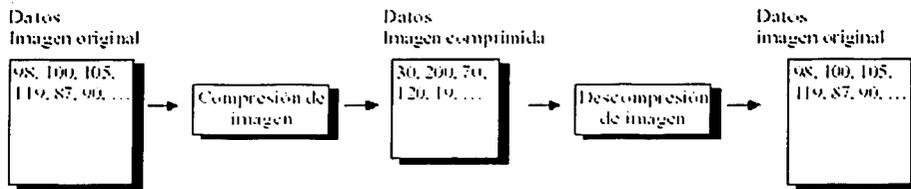


Fig a.2 Diagrama de flujo de las operaciones de compresión y descompresión de imagen.

La compresión, reduce los datos de las imágenes, mientras que la descompresión se refiere a la conversión de los datos comprimidos a su forma original. Ambas operaciones son llamadas de *codificación de imágenes*, pues emplean la codificación para representar las imágenes de una forma más concisa.

La *relación de compresión*, es la medida de qué tanto se comprime, es el cociente del tamaño de los datos de la imagen original y el tamaño de la imagen ya comprimida. Una razón de compresión grande indica una menor cantidad de datos en la imagen comprimida. Así entonces el objetivo es maximizar la razón o tasa de las operaciones de compresión, respetando claro esta la calidad de la imagen, tiempo de compresión y descompresión, esfuerzo computacional. Así entonces los algoritmos usados en los sistemas de compresión dependen directamente de la disponibilidad del ancho de banda o la capacidad de almacenamiento de la que se disponga, de las características requeridas por la aplicación, y del disponer de un *hardware*⁶ requerido para la implementación del algoritmo de compresión.

⁶ *Hardware: Elementos electrónicos y mecánicos de una computadora.*

TESIS CON
 FALLA DE ORIGEN

A.2. CONSIDERACIONES EN LA COMPRESIÓN.

La *calidad* de la imagen es la consideración más importante en la elección del algoritmo de compresión.

Los dos aspectos más importantes de la calidad de vídeo son la resolución espacial y la resolución temporal. La *resolución espacial* describe la claridad o falta de definición en los contornos de la imagen desplegada. La *resolución temporal* describe la suavidad del movimiento. Un gran número de píxeles implica una alta resolución espacial. Similarmente entre más marcos por segundo tengamos la resolución temporal es mayor.

A.2.1 ROBUSTEZ.

Ya que mediante la compresión la redundancia es removida de manera total o parcial, cada bit comprimido se vuelve más importante, es decir un error durante la transmisión o almacenamiento tiene efectos que recaerán sobre una grande región de la imagen o sobre un extenso periodo de tiempo.

La corrección de errores es usualmente agregada a lo largo de una señal codificada con una variedad de técnicas para disimular los errores, las cuales regularmente son exitosas en la reducción de los efectos de errores aleatorios aislados. Entonces la elección del algoritmo de compresión depende de los errores de transmisión propios del medio en el cual la aplicación reside.

A.2.2. COMPRESIÓN Y RETARDO DE LA PAQUETIZACIÓN.

Los avances en los algoritmos de compresión se obtienen predominantemente a través de mejoras en el análisis de la señal en cuestión. Así también, el análisis de la fase en la decodificación tiende a ser más complejo, convirtiéndose el retraso en la codificación en un parámetro importante. Un algoritmo de compresión que busca un gran número de muestras y realizar muchas operaciones complejas que usualmente lleva un retraso de codificación grande.

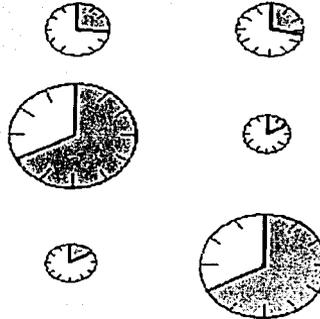
En algunas aplicaciones el retardo presente en la codificación es tolerable, pero para otras no. La televisión *broadcast* en algunas ocasiones puede permitir un retraso del orden de segundos. Teleconferencias y grupos de trabajo de multimedia pueden tolerar un pequeño retraso. Así entonces la comunicación moderna de datos a través de paquetización de información siempre introducirá retardo de codificación.

SIMETRÍA.

Las operaciones de compresión y descompresión no siempre son simétricas, pues dependen de la técnica utilizada, así entonces en función del tiempo y esfuerzo computacional involucrado en cada uno de los casos, la codificación puede ser:

- Simétrica.
- Con compresión asimétrica.
- Con descompresión asimétrica.

Tiempo de compresión Tiempo de descompresión



TIEMPO CON
FALLA DE ORIGEN

Fig a.3 La simetría de la compresión es una medida del tiempo y esfuerzo computacional requerido para una operación de compresión respecto a su correspondiente descompresión.

- a. **Codificación simétrica.** Aquella en la que el esfuerzo computacional y tiempo requeridos tanto para la compresión como para la descompresión es el mismo. Esta codificación es empleada preferentemente para el intercambio de imágenes entre dos puntos en los que además el número de veces que la imagen es comprimida y descomprimida es equivalente.

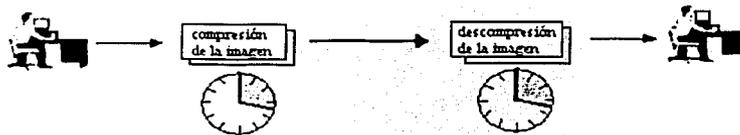


Fig 3.4 La Codificación simétrica se utiliza cuando una imagen va a ser sometida a compresión y descompresión por igual número de veces.

- b. **Codificación con compresión asimétrica.** Aquella en la que el tiempo y/o esfuerzo computacional de la compresión es mayor que el de la descompresión, haciendo esta última operación más rápida y/o eficiente computacionalmente. Con lo que el tiempo total de compresión/descompresión es reducido significativamente. Requerida cuando la imagen es comprimida una vez y a disposición de varios usuarios, esta será descomprimida varias veces, por ejemplo para la distribución de imágenes.

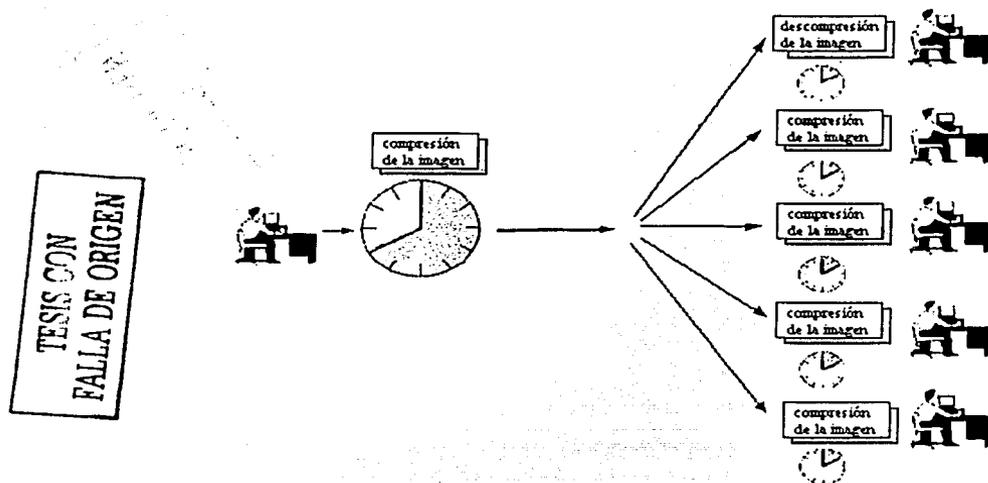


Fig a.4 La codificación con compresión asimétrica se utiliza cuando una imagen va a ser descomprimida muchas más veces que comprimida.

- c. **Codificación con descompresión asimétrica.** En ésta es la descompresión la que involucra más tiempo y/o esfuerzo computacional que la compresión. Conveniente para el archivo de imágenes, es decir para las situaciones en las que la imagen se comprime varias veces y se descomprime una o muy pocas veces.

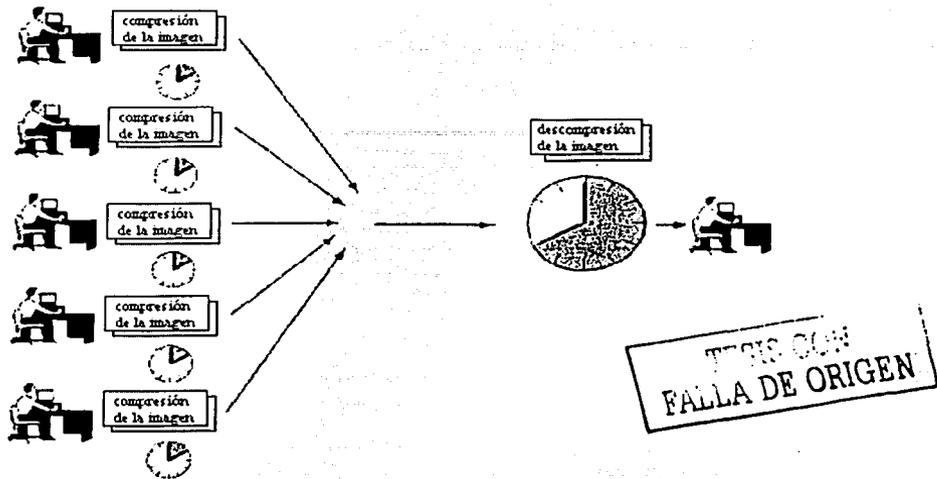


Fig a.5 La Codificación con descompresión asimétrica se usa cuando una imagen va a ser más veces comprimida que descomprimida.

A.2.3. ESCALABILIDAD.

Una señal comprimida puede ser nuevamente comprimida empleando una representación alternativa a la original. Su utilidad es deseable para crear presentaciones en diferentes resoluciones (en espacio, tiempo, amplitud, etc.) de acuerdo a las limitaciones del equipo empleado en aplicaciones particulares. La escalabilidad puede alcanzar su mayor potencia en la descompresión reduciendo la resolución y la ganancia de la compresión. Como sea la secuencia de operaciones introducen un retraso, complejidad y un deterioro en la calidad.

A.3. TAXONOMÍA DE LA COMPRESIÓN.

Las técnicas de compresión se dividen en dos grupos:

- a. **Compresión sin pérdida:** compresión conservando el contenido exacto de la información de la imagen original.
- b. **Compresión con pérdida:** conserva la calidad de la imagen hasta un nivel dado, pero no conserva el contenido total de información de la imagen original.

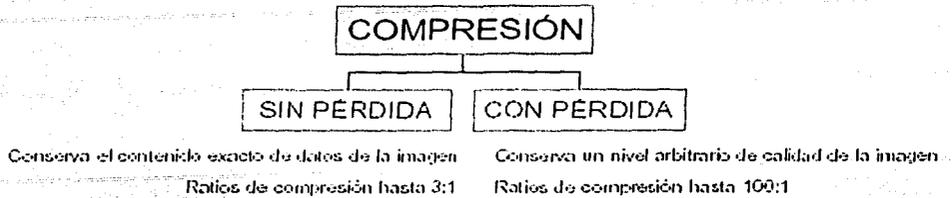


Fig a.6 División de las técnicas de compresión.

MÉTODOS DE COMPRESIÓN PARA IMAGEN, VÍDEO Y AUDIO.

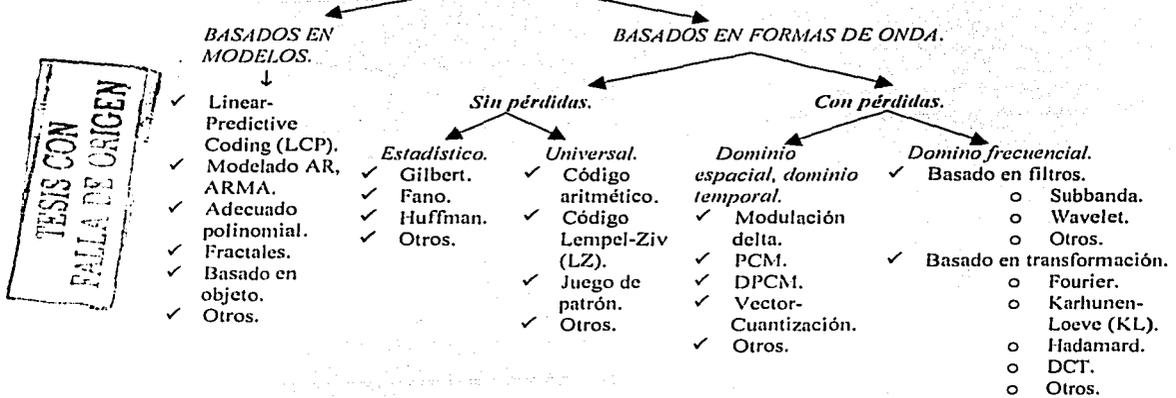


Fig a.7 Una taxonomía de la compresión para imagen, vídeo y audio.

A.3.1. COMPRESIÓN SIN PÉRDIDA VS COMPRESIÓN CON PÉRDIDA.

COMPRESIÓN SIN PÉRDIDAS.

En muchas aplicaciones, el decodificador tiene que recuperar la información sin ningún tipo de pérdida. Para la elección de un método para una aplicación específica de la compresión existen tres factores con los que hay que negociar: eficiencia de codificación, complejidad del codificador y retraso de codificación [BHK097].

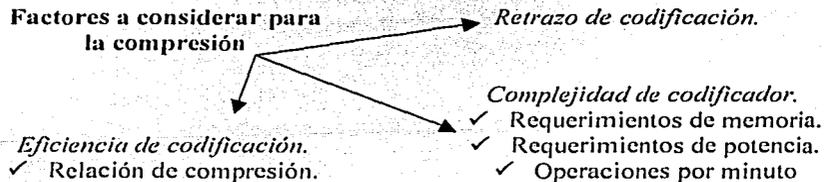


Fig a.8 Aspectos importantes a considerar en la compresión sin pérdidas.

Eficiencia de Codificación. Normalmente es medida en bits por muestra o bits por segundo (bps). La principal limitante para la eficiencia de la codificación es la entropía o cantidad de información proporcionada por la fuente. En terminos intuitivos, la entropía de una fuente X proporciona una medida de la "aleatoriedad" de X. Desde del punto de vista teórico de la compresión, las fuentes con una gran entropía son más difíciles de comprimir, por ejemplo, el ruido aleatorio [BHK097].

Complejidad de codificador. Este proceso es análogo al esfuerzo computacional necesario para implementar el codificador y decodificador. Normalmente este esfuerzo computacional es medido en términos de los requerimientos de memoria y el número de operaciones aritméticas. En algunas aplicaciones, como lo son las portátiles, la complejidad de codificación puede estar limitada por la potencia requerida del *hardware* que se pretende implementar.

Retrazo de codificación. Frecuentemente un proceso complejo de codificación conlleva un incremento de los retrasos en el codificador y decodificador. Estos retrasos pueden ser compensados lo que implicaría un incremento en el procesamiento y sería poco práctico en ambientes donde existe una potencia o equipo computacional que no puede ser mejorado. Además dependiendo de la aplicación, por ejemplo en comunicaciones interactivas, estos retrasos tienen que ser restringidos, lo que a llevado a diseñar sistemas de compresión que usan algoritmos simples.

COMPRESIÓN CON PÉRDIDAS.

La mayoría de las aplicaciones de procesamiento de datos de imagen y vídeo no requieren que los datos reconstruidos y originales sean de valor idéntico, sin embargo sí deben permitir una buena reconstrucción. Así un proceso de compresión con pérdidas es aquel cuyo resultado es una reconstrucción imperfecta pero aceptable. Este proceso de pérdida es irreversible. En la práctica, la mayoría de los procesos de compresión irreversibles degradan rápidamente la calidad de la señal cuando ellos son aplicados en varias ocasiones sobre los datos previamente descomprimidos.

La elección de un método de compresión con pérdidas específico involucra un compromiso entre cuatro aspectos.

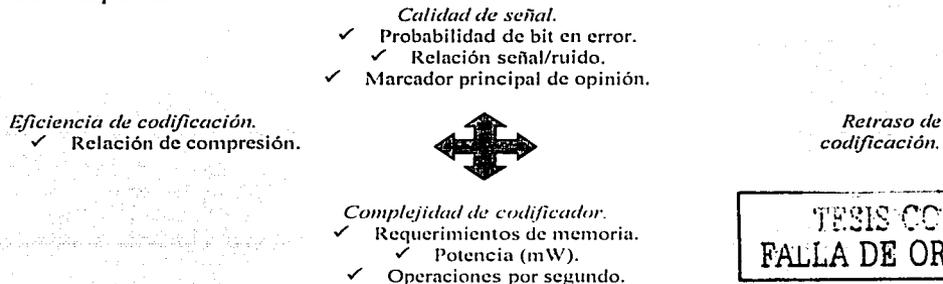


Fig a.9 Aspectos importantes a considerar en la compresión con pérdidas.

Como consecuencia del poder reducir cierta cantidad de información, la compresión con pérdidas produce relaciones mayores que las de los métodos sin pérdidas.

Calidad de señal. Este término frecuentemente es usado para caracterizar la señal en la salida del decodificador. No existe una medida de la calidad de la señal aceptada universalmente. Aunque una medida cuantitativa regularmente empleada es la relación señal a ruido, la cual puede ser expresada como:

$$SNR = 10 \log_{10} \frac{\text{energía de la señal de salida del codificador}}{\text{energía de la señal de ruido}}$$

La energía de la señal de ruido es definida como la energía medida de una señal de error que es la diferencia entre la señal de entrada al codificador y la señal de salida del decodificador. Para el caso de las imágenes y vídeo en lugar de la SNR se emplea el PSNR (peak signal-to-noise ratio), dado por la siguiente ecuación:

$$PSNR(dB) = 10 \log_{10} \frac{255^2}{MSE}$$

Donde:

$$MSE = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (y_{ij} - x_{ij})^2$$

Dado que son medidas promedio, un alto valor de SNR o PSNR no siempre significa una señal con alto nivel de calidad perceptual. Otra medida de la calidad de la señal es el marcador principal de opinión, donde la representación de un proceso de compresión es caracterizado por la calidad subjetiva de la señal decodificada. Para ello se plantea una escala de cinco puntos: *muy molesto, molesto, ligeramente molesto, perceptible pero no molesto e imperceptible*. Cualquiera de ellos puede ser utilizado para caracterizar el perjuicio en la salida del decodificador y frecuentemente se usan ambos.

La calidad de los datos de entrada afectan la relación de compresión tanto para el método de compresión con pérdidas, como sin pérdidas. Por ejemplo, la adquisición de ruido, los errores en el momento del muestreo, e incluso el proceso de conversión analógico-digital afecta la calidad de la señal y reduce la correlación espacial y temporal.

Los métodos de compresión con y sin pérdidas se ajustan dentro de un modelo general descrito en la figura siguiente.

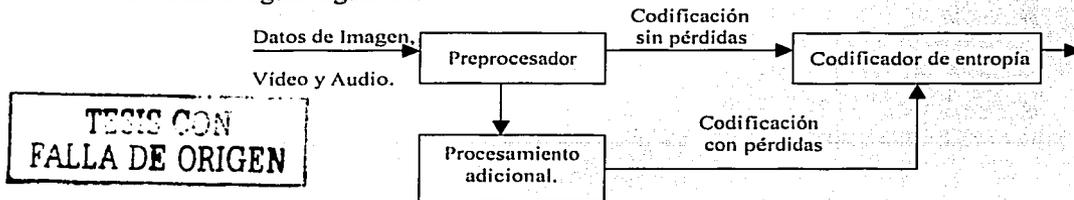


Fig a.10 Marco de codificación para la codificación con y sin pérdidas.

A.3.2. TASA VARIABLE DE BITS vs TASA CONSTANTE DE BITS.

Con la combinación de los tres aspectos que se involucran en el método de compresión sin pérdidas podemos diseñar varios tipos de codificadores. En una aproximación, uno puede imaginar un código donde los retrasos de codificación son fijos. Esto puede llevar a un codificador sin pérdidas donde la eficiencia de la codificación fluctúa de una muestra a otra, pero los símbolos de salida son entregados a una tasa constante. Por otro lado, si el retraso de codificación no es crítico, uno puede procesar múltiples muestras en un tiempo para obtener una alta eficiencia de codificación. En este caso, el tamaño de los datos de salida del codificador es fijo, pero los símbolos de salida son entregados en intervalos de tiempo irregulares. Así dependiendo de la aplicación podemos tener cualquiera de las dos opciones: datos con longitudes fijas en intervalos de tiempo irregulares o símbolos de longitud variable en intervalos de tiempo constantes. Ahora que si observamos la estructura del codificador genérico, es posible convertir un flujo de bits con longitud variable a la salida del codificador de fuente en un flujo de bits con longitud fija a la salida del codificador de canal.

De igual forma para el método de codificación con pérdidas, podemos obtener varias clases de codificadores. Un codificador como este puede ser requerido en un sistema de comunicación donde la tasa de transmisión no puede exceder de un valor específico o en aplicaciones cuya capacidad de almacenamiento es fija. Por lo que estos codificadores son conocidos también como codificadores de *tasa de bits constante*.

A.4. COMPARACIÓN DE LOS DOMINIOS DE COMPRESIÓN.

Los procesos de compresión pueden explotar redundancia en los datos a través del espacio, del tiempo, o la frecuencia. Los métodos basados en muestreo escalar son frecuentemente basados en el dominio del tiempo o espacio; en este caso no hay ahorro inherente de bits si adoptamos una aproximación en el dominio de la frecuencia. Los métodos de compresión basados en vectores encajan dentro de las tres principales categorías: basadas en el dominio espacial, temporal y frecuencial. Para datos multidimensionales, tales como el vídeo, que tiene dos componentes, espacial y temporal, es adoptada una aproximación híbrida en el dominio espacial (y frecuencial). Tales técnicas híbridas son la base de todos los estándares de compresión de imágenes y vídeo.

A.5. CUESTIONES EN LA ELECCIÓN DEL MÉTODO DE COMPRESIÓN.

Cuando se elige un método específico de compresión, se deben de considerar las siguientes cuestiones:

- *Sin pérdidas o con pérdidas.* Esto es usualmente dictado por los requerimientos de eficiencia de la codificación.
- *Eficiencia de codificación.* Hasta en los procesos de compresión con pérdidas, la eficiencia de codificación deseada puede no ser alcanzada. Este es especialmente el caso cuando hay especificaciones obligatorias sobre la calidad de la señal de salida.

- *Variabilidad en la eficiencia del código.* En algunas aplicaciones, grandes variaciones en la eficiencia de la codificación entre conjuntos de datos puede no ser aceptable.
- *Resistencia a los errores de transmisión.* Algunos métodos de compresión son más robustos a los errores de transmisión que otros.
- *Requerimientos complejos.* En muchas implementaciones, es importante sobre todo guardar una baja complejidad en el codificador-decodificador. Sin embargo, ciertas aplicaciones pueden requerir sólo baja complejidad en la decodificación.
- *Degradaciones naturales en la salida del decodificador.* Los métodos con pérdidas introducen datos artificiales en la señal decodificada.
- *Representación de datos.* En muchas aplicaciones, existe la necesidad de apoyarse en dos fases de decodificación. En la primera fase, la decodificación es realizada para encontrar una señal inteligible. En la segunda fase, la decodificación es realizada para alcanzar una alta calidad de señal. Uno puede generalizar esta idea sugiriendo que algunas aplicaciones requieren una representación jerárquica de los datos. En el contexto de compresión, podemos referirnos a los programas de compresión como métodos escalables de compresión.
- *Uso repetido del tándem codificador-decodificador.* En aplicaciones tales como la edición de vídeo existe la necesidad de aplicar en múltiples ocasiones las operaciones de codificación-decodificación.
- *Interacción con otras modalidades de datos, tales como audio y vídeo.* En sistemas donde hay que soportar varios tipos de datos, los métodos de compresión de cada uno de ellos deben de tener algunos elementos en común.
- *Interactuar con otros sistemas.* En un entorno de mercado masivo, existen muchos tipos de datos y múltiples sistemas de compresión, es en este ambiente en donde se puede hacer necesario un transcodificador para llevar desde un método de compresión a otra forma los datos que se necesite.

A.6. TECNOLOGÍAS DE COMPRESIÓN.

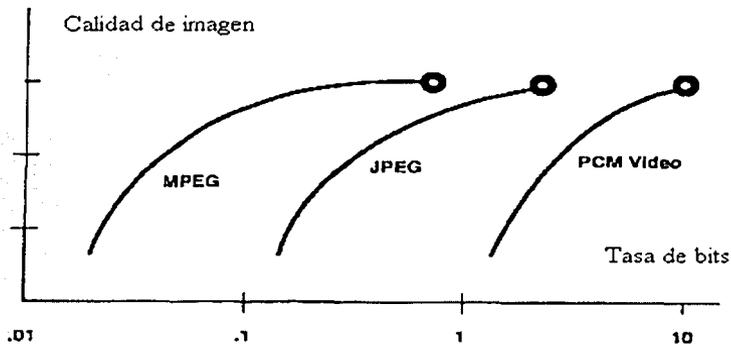
Estas técnicas de compresión se han clasificado de acuerdo con los grupos de píxeles sobre los que operan.

ELEMENTOS CODIFICADOS	CODIFICACIÓN	EJEMPLO
Elemento de imagen	PCM, DPCM	Wirephoto
Bloque de elementos	Transformada espacial	JPEG, subband
Translación de bloques	Compensación de movimiento	MPEG-1,-2
Regiones	Codificadores basados en objetos	MPEG-4 (1998)
Contenido nombrado	Elementos de escena 3-D	

Tabla a.1 Diferentes técnicas de compresión.

La codificación utilizando unos pocos píxeles es considerablemente más simple que aquella que requiere un procesamiento espacial-temporal complejo de grupos de píxeles.

Con el paso de los años, la compresión ha alcanzado mayores eficiencias, es decir a mejorado su relación de calidad de la imagen para una de tasa de bits determinada, pero con ellos los algoritmos se han hecho más complejos.



TESIS CON
FALLA DE ORIGEN

Fig a.11 Medida aproximada de la calidad de la imagen contra la tasa de bits/píxeles de varios algoritmos de compresión.

Observando en la imagen anterior podemos notar que mientras menos complejo es el algoritmo más rápida será la degradación de la calidad de la imagen con sólo pequeños cambios en la tasa de compresión. Entonces, podemos concluir que dado que ambos, tanto el algoritmo de compresión y la tasa de compresión se relacionan directamente con la calidad de la imagen, estos se determinarán de acuerdo a las necesidades de la aplicación deseada.

TESIS CON
FALLA DE ORIGEN

118

BIBLIOGRAFÍA

- [ADIV85]. ADIV, G. *Determining three-dimensional motion and structure from optical flow generated by several moving objects*. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. PAMI-7, No. 4 julio de 1985.
- [AGG88]. AGGARWAL, J. K. y NANDHAKUMAR, N. *On the computation of motion from sequences of images*. Proceeding of the IEEE. Vol. 76, No. 8, agosto de 1988.
- [AIZA95]. AIZAWA, K. y HUANG, T. S. *Model-based image coding: advanced video coding techniques for every low bit-rate applications*. Proceeding of the IEEE. Vol. 83, No. 2, febrero de 1995.
- [ANAN93]. ANANDAN, P. Et al. *Hierarchical model-based motion estimation*. En "Motion Analysis and Image Sequence Processing". M. I. Sezan y R. L. Lagendijk, eds. Norwell, MA, Kluwer, 1993.
- [ANIL91]. JAIN, Anil K. *Fundamentals of digital image processing*. Ed Prentice Hall. USA 1989.
- [BAAZ91a]. BAAZIZ N. y LABIT, C. *Multigrid motion estimation on pyramidal representation for image séquence coding*. Reporte técnico 572, IRISA-Rennes, febrero de 1991.
- [BAAZ91b]. BAAZIZ N. *Approches d'estimation et de compensation de mouvement multiresolutions pour le codage de sequences d'images*. Tesis Doctoral, Universidad de Rennes I, Francia, octubre de 1991.
- [BARR94]. BARRON, J. L., FLETE D. J. y BEAUCHEMIN, S. S. *Systems and Experiment. Performance of Optical Flowtechniques*. International Journal of Computer Vision. Vol. 12. No. 1. 1994. pp. 43-77.
- [BHKO97] Image and Video Compression Standards. BHASKARAN, Vasudev y KONSTANTINIDES, Konstantinos. Ed. Kluwer Academic Publeshers. EUA 1997.
- [BIEM87]. BIEMOND, J., BOEKKEE, Looijenga y PLOMPEN, R. *A pel-recursive Wiener-based displacement estimation algorithm*. Signal processing. Vol. 13, No. 4, diciembre de 1987, pp. 399-412.
- [BIERS86]. BIERLING, M. y THOMA, R. *Motion compensating field interpolation using a hierarchically structured displacement estimator*. Signal Processing. Vol. 11, 1986. pp. 387-404.

- [BRA195]. BRAILEAN, J. C. y KATSAGGELOS, A. K. *A recursive nonstationary MAP displacement vector field estimation algorithm*. IEEE Transactions on Image Processing. Vol. 4, No. 4, abril de 1995, pp. 416-429.
- [CHAN93]. CHANG, M. M., TEKALP, A. M. y SEZAN, M. I. *Motion-field segmentation using an adaptative MAP criterion*. Proceeding of the IEEE International Conference on ASSP, 1993.
- [CHEL93]. CHELLAPA, R. y JAIN A. K., eds. *Markov random fields. Theory and Applications*. San Diego, Academic Press, 1993.
- [CKR96]. CASTLEMAN, Kenneth R. *Digital Image Processing*. Ed. Prentice Hall. USA. 1996.
- [DGP96] D. Gatica-Pérez, *Estimación de movimiento y segmentación basada en movimiento en secuencias de imágenes*. Tesis de Maestría, Universidad Nacional Autónoma de México, Distrito Federal, México, junio 1996.
- [DUBO93]. DUBOIS, E. y konrad, j. *Estimation of 2-D motion fields from image sequence with application to motion compensated processing*. En "Motion analysis and image sequence processing". M. I. Sezan y R. L Lagendijk, eds. Norwell, MA., Kluwer Academic Publishers. 1993.
- [EVER56]. EVERITT W. L. *Communication Engineering*. Ed. Mc Graw Hill. USA. 1956.
- [FEA91]. FELIX, Alejandro Estrada, et al. *Lecciones de física*. Ed. Compañía Editorial Continental. México, 1991.
- [FRAN91]. FRANCOIS, E. *Interpretation qualitative du mouvement a partir d'une sequence d'images*. Tesis Doctoral, Universidad de Rennes I, Francia, junio de 1991.
- [GARC95]. GARCÍA, Víctor. *Une approche de compression orientee-objets par suivi de segmentation basee mouvement pour le codage de sequences d'images numeriques*. Tesis doctoral, Universidad de Rennes I , Francia, mayo de 1995.
- [GAT196]. GATICA, Daniel. *Estimación de movimiento y segmentación basada en movimiento en secuencias de imágenes*. Reporte técnico. DEPF-UNAM, 1996.
- [GEMA84]. GEMAN, S. y GEMAN, D. *Stochastic relaxation, Gibbs distributions, and bayesian restoration of images*. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. PAMI-6, No. 6, noviembre de 1984. pp. 721-741.
- [GRC02]. GONZALEZ, Rafael C. *Digital Image Processing*. Ed. Prentice Hall. USA 2002.

- [GREEN83]. GREEN, William B. *Digital Image Processing*. Ed. Van Nostrand Reinhold Company Inc. USA 1983.
- [HAYK89]. HAYKIN, S. *Digital filters*. Macmillan Publishing Co., 1989.
- [HEIT93]. HEITZ, F. y BOUTHEMY, P. *Multimodal estimation of discontinuous optical flow using Markov Random Field*. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 15, No. 12, diciembre de 1993. pp. 1217-1232.
- [HHG97]. HARRY, G. *Digital Video: an introduction to MPEG-2*. HASKELL. Ed. Kluwer Academic Publishers. USA. 1997.
- [HILD87]. HILDRETH, E. *Computations underlying the measurement of visual motion*. En "Image Understanding 1985-1986". W. Richards y S. Ullman, Eds. Ablex Publishing Co, 1987.
- [HORN86]. HORN, B. K. P. *Robot vision*. The MIT Press- McGraw Hill Co. Cambridge, Ma. Séptima reimpresión, 1991.
- [HOTT88]. HOTTER, M. y THOMA, R. *Image segmentation based on object oriented mapping parameter estimation*. Signal Processing. Vol. 15. 1988. pp. 315- 334.
- [HUAN81]. HUANG, T. S., ed. *Image Sequence Analysis*. Berlin, Springer-Verlag, 1981.
- [HUAN83]. HUANG, T. S., ed. *Dynamic Scene Analysis and Image Sequence Processing*. Berlin, Springer-Verlag, 1983.
- [HUAN92]. HUANG, T. S. *Visual Motion Analysis*. En "Encyclopedia of Artificial Intelligence". S. Shapiro, ed. Segunda Edición. Nueva York, John Wiley and Sons, 1992.
- [IBLA93] IBRAHIM, Sezan y LAGENDIJK Reginald. *Motion Análisis and Image Séquence Processing*. Ed. Kluwer Academic Publeshers. EUA 1993.
- [JLC02]. LIRA, Jorge Chavez. *Introducción al tratamiento digital de imágenes*. Ed. Conjunta, Instituto Politécnico Nacional, Fondo de Cultura Económico y Universidad Nacional Autónoma de México. México, 2002.
- [KORN92]. KORNARD J. y dubois E. *Bayesian ESTIMATION of motion vector fields*. IEEE Transactions on Pattern Análisis and Machine Intelligence. Vol. 14, No. 9, septiembre de 1992. pp. 910-927.
- [MOOR87]. MOORHEAD, R. J., RAJALA, S. A. y COOK, L. W. *Image sequence compression using a pel-recursive motion-compensated technique*. IEEE Journal on Selected Areas in Communications. Vol. SAC-5, No. 7, agosto de 1987. pp. 1100-1114.

- [MURR87]. MURRAY, D. W. y BUXTON, B. F. *Scene segmentation from visual motion using global optimization*. IEEE Transaction on Pattern Analysis and Machine Intelligence. Vol. PAMI-9, No. 2, marzo de 1987. pp. 220-228.
- [MUSM85]. MUSMANN, H. G., PIRSH, P. y GRALLERT, H. J. *Advances in picturecoding*. Proceeding of the IEEE. Vol. 73, No. 4, abril de 1985. pp. 572-597.
- [NAGE86]. NAGEL, H. H. y ENKELMANN, W. *An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences*. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. PAMI-8, No. 5, septiembre de 1986. pp. 565-593.
- [NETR79]. NETRAVALI, A. N. y ROBBINS, J. D. *Motion-compensated television coding : part I*. The Bell System Technical Journal. Vol. 58, No. 3, marzo sw 1979. pp. 631-670.
- [NETR80b]. NETRAVALI, A. N. y ROBBINS, J. D. *Motion-compensated coding : some new results*. The Bell System Technical Journal. Vol. 59, No. 9, noviembre de 1980. pp. 1735-1745.
- [NICO91]. NICOLAS, H. y LABIT, C. *Global motion identification for image séquence aalysis and coding*. Proceeding of the ICASSP 91, IEEE. pp. 2825-2828.
- [NICO92b]. NICOLAS, H. *Hierarchie de modeles de mouvement et methodes d'estimation asocies*. Aplicacion au codage de sequences d'images. Tesis Doctoral, Universidad de Rennes I. Francia, septiembre de 1992.
- [PAPP92]. PAPPAS, T. N. *An adaptative clustering algorithm for image segmentation*. IEEE Transactions on Signal Processing. Vol. SP-40, abril de 1992, pp. 901-914.
- [PAQU83]. PAQUIN, R. y DUBOIS, E. *A spatio-temporal gradient method for estimating the displacement field in time-varying imagery*. Computer Vision, Graphics and Image Processing. Vol. 21, 1983. pp. 205-221.
- [PRATT91]. PRATT, William. *Digital image processing*. Ed. John Willey & Sons, Inc. EUA 1991.
- [SABR84]. SABRI, S. *Movement compensated interframe prediction for NTSC color TV signals*. IEEE Transactions on Communications. Vol. COM-32, No. 8, a gosto de 1984. pp. 954-968.
- [SELA93] SEZAN, M. Ibrahim y LAGENDIJK, Reginald. *Motion Analisis and Image Séquence Processing*. Ed. Kluwer Academic Publishers. EUA 1993.
- [TEKA95] TEKALP, A. Murat. *Digital Video Processing*. Ed. Prentice Hall PTR. EUA 1995.

- [TZIR88]. TZIRITAS, G. *Estimation recursive-iterative du mouvement apparent dans une sequence d'images*. *Traitement du Signal*. Vol. 5, No. 4, 1988. pp. 305-310.
- [WALK84]. WALKER, D. R. y RAO, K. R. *Improved pel-recursive motion compensation*. *IEEE Transactions on Communications*. Vol. COM-32, No. 10, octubre de 1984. pp. 1128-1134.
- [WALK87]. WALKER, D. R. y RAO, K. R. *Motion-compensated coder*. *IEEE Transactions on Communications*. Vol. COM-35, No. 11, noviembre de 1987. pp. 1171-1178.
- [WAT94]. WATKINSON, John. *An Introduction to Digital Video*. Ed. Butterworth-Heinemann Ltd. Great Britain. 1994.
- [ZIEM81]. ZIEMER R. E. *Principios de comunicaciones. Sistemas, modulación y ruido*. Ed. Trillas. México. 1981.
- Secuencia Claire. The Ohio State Univrsity. Signal analysis and machine perception laboratory. Department of electrical engineering.
<http://sampl.eng.ohio-state.edu/~sampl/database.htm>
- http://216.239.51.100/search?q=cache:JVdomyxdSBwC:wicc2000.info.unlp.edu.ar/wicc2000/area6/tra6_4.doc+M%C3%A9todo+Biemond&hl=es&lr=lang_de&lang_es&ie=UTF-8
- http://www.utem.cl/cyt/comunicacion/teoria1_2.htm
- <http://www.library.cornell.edu/preservation/tutorial-spanish/intro/intro-01.html>
- http://www.upv.es/satelite/trabajos/Grupo8_99.00/perfut.html