

52
2ej



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

Facultad de Ingeniería

Estimación de movimiento en secuencias de imágenes por medio de Transformadas Polinomiales

T E S I S
Que para obtener el título de:
INGENIERO EN COMPUTACION
p r e s e n t a

José Luis Silván Cárdenas

Director de Tesis:

Dr. Boris Escalante Ramíres



México, D. F.

1999

TESIS CON
FALLA DE ORIGEN

269673



Universidad Nacional
Autónoma de México



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.

FALTAN PAGINAS

De la:

i

A la:

iii

Agradecimientos

Esta tesis es el resultado de varios meses de trabajo en la división de estudios de Posgrado de la Facultad de Ingeniería, pero su satisfactoria finalización no habría sido posible sin la ayuda de tantas y tantas personas que directa o indirectamente influyeron en la forma que tomó finalmente.

Mi agradecimiento ante todo a Dios, por la vida y las infinitas razones que me dio para vivirla. Agradezco profundamente a mis padres por creer en mi; por su cariño y apoyo; por todos sus años de espera y sacrificio. A mis tíos, en especial a mi tía Elsa y a mi tío José Luis por su valiosa ayuda, por brindarme su casa y el calor de una familia. A mis hermanos Landy, Reyna, Daniel, Jorge, Magdalena y Blanca por su comprensión y amistad.

A todos mis maestros por contribuir efectivamente en mi formación académica. Al ingeniero Carlos Castillo Tejero, quien siendo coordinador del Programa de Alto Rendimiento Académico (PARA) de la facultad, me brindó apoyo en los asuntos académicos y la guía de sus consejos. Al propio PARA, por todo lo que representa en mi formación. A todos mis compañeros, sinceramente, GRACIAS.

Merecen especial mención el MI. Manuel García Flores, quien fue mi tutor académico durante la mayor parte de la cerrera, por sus consejos que más que de maestro, fueron de un amigo y de un ser humano; el Dr. Boris Escalante Ramírez, mi maestro y director de tesis, por compartirme sus experiencias, apoyar ampliamente el proyecto y por la propuesta del trabajo cuyos resultados aquí presento. Al Dr. Francisco García Ugalde por sus importantes observaciones sobre los resultados obtenidos. A todos ellos mi agradecimiento.

Contenido

Agradecimientos	iv
Contenido	v
Lista de Figuras	vii
Lista de Tablas	viii
1. Introducción	9
1.1 Papel de la percepción humana en el diseño de sistemas automáticos de visión.....	9
1.2 Visión automática basada en modelos biológicos	10
1.3 Análisis de movimiento en imágenes: motivación	12
1.4 Objetivo	14
2. Antecedentes	15
2.1 Introducción.....	15
2.2 Descripción del problema	15
2.3 Estimación basada en el gradiente	18
2.3.1 Restricción fundamental	18
2.3.2 Regularización del flujo.....	20
2.4 Estimación basada en correlación	21
2.5 Estimación basada en la energía espaciotemporal	24
2.6 Resumen	25
3. Transformadas Polinomiales	26
3.1 Introducción.....	26
3.2 Definición	27
3.3 Transformada de Hermite	30
3.3.1 Aproximación Discreta.....	31
3.4 Transformada generalizada de Laguerre.....	34
3.4.1 Aproximación Discreta.....	36
3.5 Representación de imágenes	38
3.6 Resumen	40

4. Representación 2-D de imágenes dinámicas	41
4.1 Introducción	41
4.2 Coordenadas intrínsecas	41
4.3 Aproximación local a 2-D	43
4.3.1 Función de ángulo	46
4.3.2 Criterio de detección.....	48
4.4 Formulación discreta	49
4.5 Estimación del flujo óptico local	50
4.5.1 Resultados experimentales.....	52
5. Representación 1-D de imágenes dinámicas	55
5.1 Introducción	55
5.2 Problema de apertura	55
5.3 Aproximación local a 1-D	57
5.4 Relación con la ecuación de movimiento de Horn-Schunck	58
5.5 Estimación del flujo normal.....	59
5.5.1 Resultados experimentales	61
6. Integración	64
6.1 Introducción.....	64
6.2 Relación entre los modelos 2-D y 1-D.....	64
6.3 Suavizado de campo	66
6.4 Esquema jerárquico	68
6.5 Descripción del método	69
6.5.1 Resultados experimentales	72
7. Conclusiones.....	80
Referencias	83

Lista de Figuras

2.1 Relación entre el movimiento de una estructura orientada y un punto	19
2.2 El problema de apertura en los métodos basados en correlación	22
3.1 Análisis y síntesis de la transformada polinomial	29
3.2 Filtros de la aproximación discreta de la TLG ($N = 6$)	37
4.1 Transformación de coordenadas	43
4.2 Rotación del sistema tridimensional	43
4.3 Histogramas de frecuencias de la energía y del coeficiente de contraste típicos	48
4.4 Estimación del flujo local basada en la energía direccional	51
4.5 (a) Imagen sintética TEXTURE. (b) Flujo óptico real. (c) Flujo estimado	53
4.6 Tramas 1, 4 y 7 de la secuencia INTERVIEW y flujo óptico estimado	54
5.1 Distribución de la energía direccional de un borde	56
5.2 Problema de apertura	56
5.3 Estimación del flujo normal basada en la energía direccional	61
5.4 (a) Tramas 1, 4 y 7 de la secuencia SPHERE. (b) Flujo normal	62
5.5 (a) Fragmento de la secuencia INTERVIEW. (b) Estimación usando	63
6.1 Algoritmo de estimación de movimiento basado en las aproximaciones 1-D y 2-D	71
6.2 (a) Imagen de textura (b) Flujo real	74
6.3 Estimación para la secuencia SPHERES. (a) Estimación inicial	75
6.4 Estimación inicial para la secuencia INTERVIEW	76
6.5 Estimación después de propagar velocidades	77
6.6 Tramas 1 y 3 de la secuencia MISSAME y estimación después de	78
6.7 Tramas 3 y 5 de la secuencia CALTRAIN y estimación después	79

Lista de Tablas

2.1 Métodos basados en el gradiente.....	21
2.2 Métodos basados en Correlación.....	23
3.1 Filtros de análisis de la THD de orden 2, 4 y 6.....	33
4.1 Orden de las funciones de ángulo no nulas para $p = 0, 1$ y 2	47

Capítulo 1

Introducción

1.1 Papel de la percepción humana en el diseño de sistemas automáticos de visión

El tópico de visión automática (o visión por computadora, como es referido por algunos autores), ha producido durante los últimos años una gran cantidad de textos con variadas propuestas para el tratamiento de problemas sobre la percepción visual. Entre las técnicas más ampliamente discutidas en la literatura popular se encuentran aquellas que difícilmente son capaces de llevar a cabo procesos similares a los que lleva a cabo el sistema de visión humano. Desde luego se puede argumentar *¿por qué debería uno esperar que los sistemas de visión automáticos lleven a cabo procesos similares a los del sistema visual humano?*; mas aún, *¿por qué debe limitarse la visión automática a los límites encontrados en el ojo?*

Sin embargo, pese a estos y otros argumentos en favor de los sistemas de visión automáticos que no están restringidos por los procesos de la visión humana, una buena parte de la comunidad de investigadores en el área se han creado firmes convicciones de que hay muy buenas razones sobre porqué la visión por computadora general (en oposición a los sistemas diseñados para ciertas situaciones muy específicas) debe limitarse precisamente por los límites de la visión humana. Quizá la más concurrida de estas razones es que la mayoría de las tareas en las cuales se intenta emplear la visión automática son similares a las tareas del sistema visual, del cual se sabe las lleva acabo en forma eficiente. Por ejemplo, se ha demostrado que los sistemas de muestreo espacial tienen una capacidad de transferencia de información cercana a la óptima cuando emplean una rejilla de muestreo hexagonal seguido de un filtro paso-bajas aplicado a la imagen, similar al modo de operación de la retina [Huck, 1983].

En este trabajo argumentamos que si es posible aprender cómo el sistema visual lleva a cabo su análisis e interpretación, esto puede proporcionar información acerca de mecanismos eficientes para la solución de los problemas concernientes a la visión automática, no solo por el hecho de que los procesos visuales que uno desearía automatizar involucran análisis e

interpretación de imagen similares a los que lleva a cabo el sistema visual humano, sino también porque la mayoría de dichos procesos se llevan a cabo sobre la misma gama de escenarios observados por los seres humanos.

Otro papel importante que juega la percepción humana en el diseño de sistemas de visión, especialmente en aquellos que incluyen despliegues (como los sistemas de imagen), es el hecho de que los datos de salida son finalmente evaluados por un observador, de ahí que las características del sistema visual humano proporcionen los parámetros críticos en el diseño de dichos sistemas. Un claro ejemplo es la composición de video en televisión. En los sistemas estándares de televisión se presentan cuadros sucesivos a una frecuencia temporal de 50 o 60 Hz sin percibir irregularidad en el movimiento de los objetos de la escena, pero a frecuencias menores es posible percibir un parpadeo en las regiones de alta *luminancia*. Otro ejemplo se tiene en procesamiento digital de imágenes, donde a menudo se considera el hecho de que el ojo humano se adapta mejor a la degradación por emborronamiento que a la degradación por ruido impulsivo (ruido *sal y pimienta*) para mejorar la calidad de una imagen.

El razonamiento cualitativo del ser humano también puede aprovecharse para definir las capacidades del sistema. En muchas situaciones, una descripción cualitativa es suficiente para proveer información útil de manera estable y robusta. Por ejemplo, algunos estudios teóricos indican que la geometría de las velocidades aparentes en una escena, por sí sola, contiene información significativa [Koenderink & Van Doorn, 1975]. Los beneficios de algunos aspectos del razonamiento cualitativo y el modelado han sido apuntados por Thompson para el análisis de escenas dinámicas en [Thompson & Kearney, 1986], donde se remarca que una reconstrucción cuantitativa tridimensional no siempre es necesaria, y que la caracterización del movimiento en cierta gama de clases resulta bastante relevante.

En suma, si existe un sistema especializado que ha sido perfeccionado a lo largo de miles de años, que sabemos lleva a cabo las tareas requeridas en el campo de la visión automática en forma eficiente; que es adaptable a las condiciones de la escena y del cual cada vez se conoce más sobre los mecanismos de procesamiento que lleva a cabo, ¿porqué no deberíamos basar nuestros modelos de análisis en un modelo biológico adecuado de dicho sistema y sobre todo cuando la evaluación final de los resultados las lleva a cabo él mismo?

1.2 Visión automática basada en modelos biológicos

Una vez conscientes de la importancia del papel del sistema visual en el desarrollo de sistemas automáticos de visión y sistemas de procesamiento de imagen, es necesario una revisión razonada de los trabajos relacionados con los mecanismos básicos de la visión para intentar modelar los procesos que llevan a cabo.

El grado en que los aspectos de la visión humana puedan ser incorporados en los sistemas automáticos de visión y al procesamiento de imágenes dependerá de la profundidad de nuestro entendimiento del sistema visual. Obviamente, dada la inconveniencia de explorar extensamente las redes neuronales y el ojo de un ser humano, la mayoría de los trabajos sobre

el funcionamiento del sistema visual se han desarrollado sobre el sistema visual de una amplia variedad de animales menores a lo largo de muchos años y es claro que durante la evolución los bloques constituyentes principales del sistema de visión se fueron especializando de acuerdo a la especie, por ello, en lo que a la visión automática concierne, no es conveniente profundizar demasiado en los sistemas visuales de especies distintas a los primates. Se ha encontrado que sólo esta especie contiene algunos de los más importantes grupos de componentes neuronales como los del sistema de visión humana [Spillman et al., 1987]. Para la visión automática estas investigaciones tienen un valor significativo en el modelado de las funciones neuronales básicas.

Desde luego, uno debe esperar que mientras más preciso sea el modelo de visión, mejores deben ser las estimaciones de los parámetros visuales (orientación, velocidad, etc.) obtenidas a partir de las propiedades del modelo. Y lo deseable es que el incremento, tanto en complejidad como en los costos de los métodos derivados del modelo, se justifiquen por la magnitud del beneficio.

En el campo de la codificación de imágenes, Schreiber fue uno de los primeros en reconocer las implicaciones de los modelos de visión humana [Schreiber, 1967]. Stockham comenzó el diseño de las técnicas de procesamiento de imágenes basados en modelos biológicos, desarrollando una técnica de realce de contraste basado en la propiedad de que la percepción del brillo de escenas naturales está principalmente determinada por las propiedades de reflexión de la escena iluminada, y no por la intensidad de la fuente de iluminación [Stockham Jr., 1972]. Dado que la *luminancia* es proporcional al producto de la *reflectancia* y la *iluminación*, se emplea una operación logarítmica seguida de un filtro lineal con el fin de incrementar la componente de reflectancia y disminuir la componente de iluminación de una imagen, asumiendo que la iluminación varía poco sobre la imagen.

Un modelo más avanzado y preciso de brillantez, desarrollado por Blommaert y Martens [Blommaert & Martens, 1990], se basa en dos propiedades de la percepción del brillo conocidas como invariabilidad escalar (el brillo es aproximadamente independiente de la distancia de visión) e independencia de luminancia (el brillo es aproximadamente independiente de la luminancia de la fuente de luz). El modelo sostiene que la luz incidente en la retina se procesa por campos receptivos de varios tamaños dando como resultado un mecanismo *multicanal* del sistema visual [Koenderink & Van Doorn, 1978]; [Koenderink & Van Doorn, 1982]. Tanto los neurobiólogos como psicofísicos coinciden en que los procesos de visión en sus etapas primarias pueden ser descritos por un conjunto de canales operando en paralelo. El brillo en cada posición dentro del campo visual es obtenido integrando las respuestas de todas las celdas con diferentes tamaños de campos receptivos en esa posición.

Por otro lado, se ha demostrado que las derivadas de Gaussianas son un buen modelo de los campos receptivos con respecto al procesamiento espacial de estímulos visuales. La integración de estos operadores en un modelo matemático se conoce como transformada de Hermite [Martens, 1990a]. Algunos trabajos recientes han mostrado que esta transformada

puede ser usada en la construcción eficiente de aplicaciones tales como reducción de ruido [Escalante-Ramírez & Martens, 1992], codificación de imagen [Escalante & Venegas, 1995] y segmentación de imagen [Moctezuma & Escalante, 1997], entre otras.

En lo que a la estimación de movimiento se refiere, las investigaciones psicofísicas en percepción visual indican que el sistema visual podría usar muchas fuentes de bajo nivel de información. Sobre este punto de vista Yuille y Grzywacz propusieron una teoría de percepción visual, estipulando la existencia de cooperación *multimodal* entre diferentes fuentes de medición de movimiento [Yuille & Grzywacz, 1992]; sin embargo los aspectos computacionales no fueron estudiados. En algunas técnicas de estimación de movimiento basadas en el análisis espectral se ha considerado la conformidad con resultados psicofísicos para el diseño de filtros tridimensionales orientados teniendo como modelo de los campos receptivos filtros de Gabor *espaciotemporales* [Watson & Ahumada, 1985]; [Adelson & Bergen, 1985] y [Fleet & Jepsen, 1985]. Resultados de investigaciones más recientes indican, sin embargo, la existencia de ciertos operadores temporales similares a los operadores derivadas de Gaussianas en los campos receptivos [den Brinker & Roufs, 1992]. La integración de estos operadores en un modelo matemático se conoce como la transformada generalizada de Laguerre. Tanto la transformada de Hermite como la transformada de Laguerre son casos especiales de la transformada polinomial presentada en el Capítulo 3.

1.3 Análisis de movimiento en imágenes: motivación

Un factor importante en cualquier situación observable práctica es el cambio de la imagen con el tiempo. Esto se debe ya sea al movimiento de un objeto dentro de la escena, al movimiento relativo entre el sistema observador y la escena, y/o a variaciones de la iluminación en la escena. En este último caso no existe movimiento de objetos dentro de la escena, pero sí hay cambios en los patrones de luminancia percibidos. Una simplificación que a menudo se hace cuando se analiza movimiento en imágenes es suponer condiciones de iluminación invariables durante el período de análisis. A la proyección de cada punto visible en la escena se asocia un vector de velocidad, y al campo de velocidades se le conoce como *flujo óptico*. Gibson fue uno de los primeros, quien en el contexto de visión por computadora, articuló la noción de un flujo óptico y estableció su importancia en la percepción espacial hablando del flujo de superficies proyectadas, a cada punto de la cual asoció una magnitud de velocidad y una dirección [Gibson, 1950].

El análisis de movimiento en imágenes puede jugar varios papeles fundamentales en las funciones de visión automática [Mítiche & Bouthemy, 1996], por ejemplo en:

Detección de Movimiento en Escenas. El movimiento en la imagen puede ser indicativo del movimiento en la escena observada. En sistemas de visión estáticos, la detección de movimiento en la escena se logra directamente de la detección de movimiento en la imagen. En el caso de sistemas de visión en movimiento se requieren pasos adicionales un poco más elaborados o una fase de compensación de movimiento de la cámara.

Segmentación de Objetos. El movimiento en imágenes puede ser utilizado para identificar regiones de la imagen correspondientes a diferentes objetos en movimiento y para distinguirlos del fondo. La hipótesis es que el campo de velocidades varía muy suavemente en los puntos interiores al contorno de oclusión.

Rastreo de Objetos. El movimiento en imágenes puede ser usado para el seguimiento de objetos en una escena sobre la base de su comportamiento dinámico actual.

Estimación de Profundidad y Movimiento Real de Objetos Físico. La distancia relativa de un objeto a la cámara y su velocidad física pueden ser cuantitativamente relacionadas con la posición en la imagen y el flujo óptico. La información tridimensional inferida puede usarse para estimar la forma en el espacio, la posición relativa, y el movimiento del sistema de visión y de los objetos físicos.

Para llevar a cabo estas funciones de visión, se proponen hipótesis sobre la naturaleza de los registros de imagen y sobre el movimiento que presentan los patrones visuales con el fin de establecer formalmente las características del problema.

Actualmente el estudio del movimiento en imágenes se ve motivado por variadas aplicaciones, entre las cuales destacan:

Televisión. Para codificación de imágenes de vídeo por compensación de movimiento, en varios servicios de telecomunicación tales como videoconferencias, televisión digital y, más recientemente, en transmisión de televisión digital de alta definición. El objetivo del análisis de movimiento aquí, es explotar la redundancia temporal y reducir la tasa de transmisión conservando la calidad de las imágenes reconstruidas en el receptor.

Autómatas Móviles (terrestres, submarinos, espaciales, nucleares). El objetivo es conferir al robot la capacidad de navegar autónomamente en ambientes dinámicos conocidos o desconocidos.

Imágenes de Satélite. Particularmente en meteorología para medir el movimiento de las nubes y establecer mapas de nubosidad para detectar tormentas, huracanes, etc.

Aplicaciones Militares. Para el rastreo de objetivos móviles y para la navegación autónoma de varios tipos de vehículos.

Imágenes Biomédicas. Ejemplos importantes son el análisis automático del movimiento del corazón y el estudio del movimiento del cuerpo humano.

Vigilancia de Sitios. Para detección de intrusión, monitoreo urbano y control de tráfico.

Interfaces y Realidad Virtual. En las nuevas interfaces máquina-hombre, especialmente dentro del dominio de la realidad virtual emergente, podría requerirse pronto un procesamiento e interpretación de movimiento facial (por ejemplo, el movimiento de los labios), o gestos humanos (por ejemplo, el movimiento de las manos y los dedos).

1.4 Objetivo

El objetivo de esta tesis es el desarrollo de un nuevo algoritmo de estimación de movimiento en secuencias de imágenes, basado en la técnica de descomposición local conocida como la transformada polinomial.

El trabajo está organizado de la siguiente manera. En el Capítulo 2 se presenta la descripción del problema y una sinopsis de los enfoques clásicos de estimación. El Capítulo 3 es un resumen de los resultados teórico relacionados con la transformada polinomial y su aplicación al procesamiento de imágenes. En los Capítulos 4 y 5 se derivan los resultados de dos modelos de imagen dinámica útiles para la estimación del flujo. También se discuten algunos detalles de la implantación de estas técnicas en computadora y su relación con otros enfoques. En el capítulo 6 se presenta la integración de los resultados previos y se discute la representación del campo de velocidades por medio de transformadas polinomiales para el suavizado del campo. Al final se presentan las conclusiones de esta tesis.

Capítulo 2

Antecedentes

2.1 Introducción

La estimación del flujo óptico debe basarse únicamente en la parte observable de la escena, esto es, en el registro de intensidades de patrones espaciotemporales captados por el sistema de visión. La mayoría de las técnicas dedicadas a resolver este problema, plantean la resolución de un conjunto de restricciones derivadas de hipótesis hechas sobre la naturaleza de los patrones observados y sobre el movimiento aparente. Por ejemplo la ecuación general de movimiento introducida por Horn y Schunck (sección 2.3.1), es una relación diferencial entre el flujo óptico y las variaciones de intensidad, que se deriva de la hipótesis de que los valores de intensidad se mantienen constante a lo largo de la trayectoria de movimiento. Adicionalmente se establece una restricción de *regularización* derivada de la hipótesis de que el flujo óptico es un campo vectorial globalmente *suave*.

En este capítulo se presenta una sinopsis estructurada de los problemas involucrados en el análisis y la estimación del movimiento en imágenes, así como una descripción cualitativa de los esquemas clásicos. La sección 2.2 trata con los problemas asociados a la detección y estimación de movimiento. En las secciones 2.3 a 2.5 se resumen las ideas básicas de los enfoques típicos de estimación de movimiento clasificados en: los que se basan en el cálculo del gradiente, los que se basan en la solución del problema de correspondencia y los que se basan en la energía espaciotemporal. Este último enfoque es menos popular que los dos anteriores, pero es de particular interés en nuestro estudio porque incorpora algunos aspectos relevantes del sistema visual. Como veremos en capítulos subsecuentes, el método propuesto está basado en la energía espaciotemporal de los registros de luminancia. También mostraremos como nuestro análisis de movimiento a lo largo de contornos basado en la transformada de Hermite conduce a la ecuación del gradiente discutida en este capítulo.

2.2 Descripción del problema

Como se mencionó en el capítulo anterior, el movimiento en la imagen, es decir, el

movimiento sobre la superficie de proyección del sistema de visión, se debe al movimiento relativo entre el sistema de visión y la escena. A la distribución de los vectores de velocidad aparente asociados a cada punto de la imagen se le denomina *flujo óptico*. Otro concepto que a menudo es usado equivalentemente con el de flujo óptico, pero que en general es diferente, es el de *flujo de imagen*, el cual es la proyección bidimensional de las velocidades físicas de los puntos visibles en la escena sobre el plano imagen. Las componentes normales del *flujo óptico* y del *flujo imagen* son iguales siempre y cuando las condiciones de iluminación sean uniformes y constantes, y los cambios espaciotemporales de las intensidades se deban exclusivamente al movimiento traslacional de objetos con superficies no especulares [Singh, 1991].

A primera vista parece trivial *sensar* el movimiento local en una imagen. Como ha sido ampliamente mostrado, una simple substracción, punto a punto, de la distribución de intensidades de la imagen en dos instantes de tiempo es suficiente para *sensar* el movimiento (o, estrictamente, la componente de velocidad ortogonal a la orientación de la estructura, la componente paralela se pierde generalmente debido al *problema de apertura*). Por este medio solamente en las regiones cercanas a las discontinuidades de la distribución de intensidades que se han movido se encontrarán resultantes diferentes de cero. Para desplazamientos pequeños y suponiendo un campo de velocidades globalmente suave esta diferencia temporal es proporcional al gradiente espacial de la distribución de intensidades siendo la constante de proporcionalidad igual a la magnitud de la velocidad. Sin embargo, no siempre se puede esperar que el campo sea localmente constante o globalmente suave, por lo que las discontinuidades en el movimiento deben tomarse en cuenta en el proceso de estimación de movimiento.

El *problema de apertura* [Hildreth, 1983] surge directamente del hecho de que el movimiento aparente que puede medirse para cualquier porción local de una estructura fuertemente orientada (un borde, por ejemplo), es solo la componente de movimiento ortogonal a la orientación local. Si una abertura (equivalente a un campo receptivo o a una ventana 2-D), es colocada sobre la porción central de un borde largo y recto, entonces, todo lo que es observado por el operador local es una porción no acotada de un borde recto. Si ahora ese borde se mueve en cualquier dirección entonces, dado que los extremos no llegan a ser visibles dentro de la abertura, no hay suficiente información para decir qué tan rápido se mueve en la dirección paralela a éste. De hecho, en el límite, si la dirección de movimiento es completamente paralela, ni el ojo humano (a través de la abertura), ni una computadora, serían capaces de determinar si el borde se ha movido. Así, para un borde recto y largo, sin otra estructura, es imposible determinar si existe componente tangencial de movimiento. Dicha componente puede inferirse solo si la estructura presenta cambios de orientación significativos formando así esquinas y si dicho cambio permanece dentro de la ventana durante un período de tiempo razonable (mínimo dos muestras consecutivas, en el caso discreto). El *problema de apertura* es, entonces, fundamental en cualquier sistema de análisis de imagen, incluido el sistema visual humano.

La forma más usual de tratar el movimiento en imágenes es buscando estructuras espaciales en muestras temporales sucesivas que tengan características similares y midiendo el desplazamiento entre ellas. Este enfoque trata con lo que comúnmente se refiere como el *problema de correspondencia* [Marr, 1982]; [Marr & Poggio, 1979]. Siendo un enfoque de apareamiento entre dos cuadros consecutivos de una secuencia, es difícil medir velocidades muy pequeñas de esta manera. En imágenes simples, con pocas estructuras o con varias estructuras con diferencias bien marcadas, dichas técnicas pueden operar muy bien. Sin embargo, para imágenes complejas o degradadas por ruido; sensar, seleccionar y hacer corresponder correctamente las estructuras se convierte en un serio problema, y en el límite, el desempeño de la técnica llega a depender directamente de la estadística del ruido y de la imagen. Además, en patrones regulares existe el riesgo de que el apareo sea erróneo o puede detectarse un movimiento inverso en el contexto global similar al que puede producirse como ilusión visual por un estroboscopio al iluminar ciertos patrones visuales en movimiento. En patrones con puntos aleatorios, en particular, el problema de correspondencia se torna muy complejo; en contraste, el ojo humano es capaz de predecir el movimiento con cierta facilidad [Julesz, 1971]; [Van Doorn & Koenderink, 1984]. Un problema al que se enfrenta el análisis de movimiento en secuencias discretas es el tener que trabajar con *diferencias enteras* entre píxeles, lo cual es típico de los enfoques de correspondencia de estructuras. Bajo condiciones de estructuras simples e ideales, los sistemas que operan sobre pares de imágenes pueden manejar muy bien desplazamientos grandes, pero son insensibles a movimientos pequeños. En particular, en un píxel base, el movimiento que puede detectarse está limitado a incrementos de píxeles enteros. Dicha insensibilidad limita severamente la utilidad de cualquier sistema práctico, requiriendo ya sea una mayor densidad de muestreo o aceptar campos de visión muy limitados con el fin de proporcionar la sensibilidad adecuada a pequeños movimientos absolutos.

Cuando un objeto se mueve dentro de la escena recubre o descubre ciertas áreas u objetos que pueden a su vez estar en movimiento. Este *recubrimiento* o *descubrimiento* introduce en la distribución de intensidades discontinuidades que dificultan tremendamente el modelado matemático de las imágenes dinámicas.

Por último, un sistema de interpretación no debe depender solamente de la habilidad de completar y precisar los datos, y estrictamente, de un conjunto fijo de condiciones o restricciones, de ser así el sistema fallaría al aparecer la incertidumbre. El sistema debe, tan pronto como sea posible, tomar en cuenta la incertidumbre inherente en las observaciones. En general, podemos distinguir tres fuentes de incertidumbre: datos incompletos (perdidos), imprecisión de los datos (debido a los sensores, a múltiples transformaciones digitales, etc.), y datos ambiguos (incluso con datos precisos y completos, pueden existir varias interpretaciones igualmente aceptables). El problema de la incertidumbre debe ser, por tanto, enfocado explícitamente.

2.3 Estimación basada en el gradiente

En un modelo de imagen dinámica se intenta relacionar los cambios de intensidades con el flujo o movimiento de los píxeles suponiendo, generalmente, que las imágenes son producidas bajo condiciones de iluminación constante. Como esta suposición en general no es cierta se tiene que las velocidades estimadas por algoritmos que se basan en los cambios espaciotemporales de las intensidades llegan a ser muy diferentes del flujo imagen (que es la proyección del campo de velocidades físicas de los puntos visibles sobre el plano imagen). En particular, se ha argumentado que las velocidades estimadas por métodos basados en el gradiente son raramente cercanos al flujo óptico. En otras palabras, el campo de velocidades que resulta de estos métodos es generalmente no coherente.

La coherencia del movimiento es fundamental en el análisis de movimiento y en aplicaciones de rastreo de objetos (*tracking*). Sin embargo, en aplicaciones de codificación, la coherencia no es un factor crítico. De ahí que los métodos basados en el gradiente sean considerados hasta la fecha como área de investigación.

2.3.1 Restricción fundamental

El punto de partida de estos métodos es la estimación de una restricción sobre los posibles movimientos de un punto de la imagen. La formalización clásica debida a Horn y Schunck es la siguiente [Horn & Schunck, 1981]. Sean las variaciones espaciotemporales de intensidades representadas por una función derivable $I(\mathbf{r}, t)$ y sea la proyección de la trayectoria de movimiento $\mathbf{s}(t)$ al instante t de un punto que se mueve en el espacio. Si asumimos que la intensidad registrada para el punto del objeto físico no cambia a medida que el objeto cambia de posición en el espacio, entonces la composición $I(\mathbf{r}, t)$ en $\mathbf{r} = \mathbf{s}(t)$ es constante. Así, derivando respecto a t se llega a

$$\mathbf{v} \cdot \nabla I = -I_t \quad (2.1)$$

donde el vector ∇I es el gradiente espacial de I , I_t es la derivada parcial de I respecto a t y $\mathbf{v} = (v_x, v_y)$ es el vector velocidad asociado a la proyección del punto en el espacio, es decir, la derivada de la trayectoria $\mathbf{s}(t)$.

Esta ecuación es referida como la ecuación del gradiente y constituye una restricción local del movimiento. Nótese que, aunque la única cantidad desconocida en esta ecuación es \mathbf{v} , claramente se ve que no puede ser determinada en forma única (por tratarse de una cantidad vectorial). Lo que la ecuación del gradiente hace, es definir una línea en el plano v_x, v_y sobre la cual descansa el vector \mathbf{v} . Esto nos da, en otras palabras, solo una componente del vector de movimiento como se ilustra en la Figura 2.1. La proyección (v_n) de \mathbf{v} sobre la dirección del gradiente se obtiene fácilmente de (2.1), esto es

$$\mathbf{v}_n = -l_i \frac{\nabla I}{|\nabla I|^2} \quad (2.2)$$

Así, conocidas las derivadas espaciales y temporal de la imagen, la ecuación del gradiente permite calcular la *componente* de velocidad en la dirección del gradiente, es decir, la *componente ortogonal (vernier motion)* a la orientación de una estructura localmente unidimensional tal como un borde o una línea. Este resultado refleja claramente el problema de apertura como la ambigüedad de interpretar cuantitativamente el movimiento traslacional de un borde visto a través de una apertura.

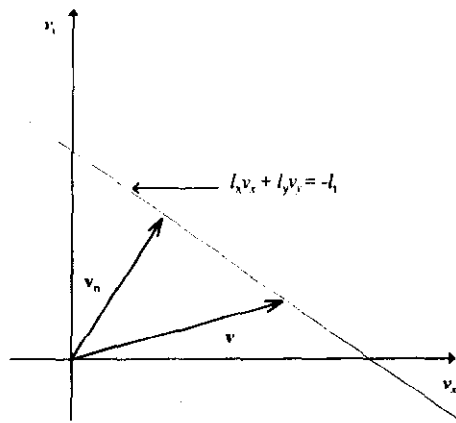


Figura 2.1 Relación entre el movimiento de una estructura orientada y un punto.

En la práctica, como las derivadas son aproximadas por diferencias finitas, se asume que las disparidades de movimiento entre tramas consecutivas son suficientemente pequeñas en relación a la escala espacial de la estructura. En general, se acepta que la invariabilidad de las intensidades respecto al movimiento se cumple aproximadamente cuando se tienen pequeños desplazamientos de objetos con superficies no especulares durante un corto período de observación. Por otro lado, en regiones poco texturizadas o con intensidad constante, las derivadas espaciales y temporal se hacen cero y como consecuencia ocurre una indeterminación en (2.2). Además el problema de recubrimiento introduce discontinuidades en la distribución de intensidades que invalidan el tratamiento matemático y práctico de los métodos basados en el gradiente.

2.3.2 Regularización del flujo

La forma usual en que los métodos basados en el gradiente intentan resolver el problema de apertura es mediante la *regularización* del flujo. En este proceso se lleva a cabo la optimización global de una restricción impuesta en el campo de velocidades. Se dice que los métodos basados en el gradiente requieren de la regularización del flujo porque el problema de estimación de movimiento en imágenes es un problema mal propuesto (*ill-posed problem*), es decir, no tiene solución única. El objetivo de la regularización es obtener un problema bien propuesto (*well-posed problem*) mediante introducción de contexto.

La optimización global resulta en general una técnica más robusta y mejor comportada que una técnica de optimización local, pero requiere un alto poder de cálculo. La idea general es establecer las características del campo de velocidades que satisface la ecuación del gradiente y alguna forma de suavidad. Por ejemplo, Horn y Schunck sugirieron que el campo debe ser tal que minimiza la siguiente función de costo [Horn & Schunck, 1981].

$$\iint \left[(\mathbf{v} \cdot \nabla I(\mathbf{x}, t) + I_t(\mathbf{x}, t))^2 + \alpha^2 \left(|\nabla v_x(\mathbf{x}, t)|^2 + |\nabla v_y(\mathbf{x}, t)|^2 \right) \right] dx \quad (2.3)$$

El primer término de la expresión es una medida de la desviación que se obtiene al suponer la conservación de las intensidades a lo largo de la trayectoria de movimiento y el segundo mide la suavidad del campo a partir del módulo del gradiente de cada componente de velocidad. La suavidad del campo es regulada por el parámetro α . En general la restricción de suavidad puede tomar diferentes formas, por ejemplo puede darse en términos del laplaciano de las componentes de velocidad o en términos de la divergencia o el rotacional del campo; y lo deseable es que el campo no se suavice sobre la frontera de oclusión, lo cual raramente se logra.

También se ha hecho una formulación estocástica del esquema de Horn y Schunck, la cual proporciona una vista más amplia del problema porque introduce un modelo estadístico del campo de velocidades y un modelo de ruido. En una formulación bayesiana, las *propiedades a priori* del modelo proporcionan las bases para el problema de regularización. Como esta regularización es generalmente expresada en términos de un contexto local, una forma natural de proceder es mediante el uso de campos aleatorios de Markov (MRF). En el marco de los MRF, el uso de un criterio de máximo *a posteriori* (MAP) para maximizar la probabilidad *a posteriori* de las características del movimiento (velocidades y discontinuidades) dadas las observaciones (variaciones de las intensidades), conduce al problema de minimización de una función global de energía. Esta minimización se logra mediante un algoritmo de relajación estocástica conocido como recocido simulado (*simulated annealing*) [Geman & Geman, 1984].

Método	Cantidad Conservada	Flujo real En patrones 2-D	Flujo normal en patrones 1-D	Comentarios
Horn-Schunck	Intensidad	No	No	
Nagel	Intensidad	Si	Si	El problema de derivadas
Enkelmann	Intensidad	Si	Si	Empieza un Control
Buxton-Quarta	Perfiles de intensidad	No	No	
Waxman-Wa- Bergholm	Perfiles de Actividad	No	No	Problemas de derivadas

Tabla 2.1.- Métodos basados en el gradiente

Otros métodos emplean la ecuación de restricción de movimiento sobre otras cantidades. La idea es que la conservación de intensidades no es una suposición realista y se postula que la conservación de las derivadas de la imagen es mas realista. En general, se asume que la información que se conserva es una versión filtrada (paso-bajas o paso-banda) de la distribución de intensidades o bien la conservación de ciertos perfiles o estructuras orientadas de la imagen. En la Tabla 2.1 se resumen los métodos más relevantes dentro del enfoque del gradiente. Todos ellos emplean la ecuación del gradiente sobre cierta cantidad que se asume se conserva sobre la trayectoria de movimiento. Las características de cada uno de ellos se resumen en términos de tres preguntas básicas:

- ¿Puede recuperar el flujo real en patrones no orientados?
- ¿Puede recuperar el flujo normal en patrones orientados?
- ¿Existe algún problema inherente, limitación o ventaja notable?

2.4 Estimación basada en correlación

A menos que se adopte un procesamiento multiresolución, el movimiento en imágenes donde ocurren desplazamientos grandes invalida la formulación que se refiere a las velocidades ópticas instantáneas como en los métodos basados en el gradiente. Una alternativa es adoptar un enfoque de correlación. Las técnicas basadas en correlación suponen típicamente la conservación de la distribución local de intensidades. Generalmente operan sobre dos imágenes sucesivas de una secuencia. En esencia, para cada bloque de píxeles en la

primera imagen. buscan el bloque de pixeles correspondiente en la segunda imagen. La salida es un vector de desplazamiento para cada bloque de pixeles en la imagen. Dicha búsqueda involucra encontrar la mejor correspondencia entre los pixeles considerados en la primera imagen y los candidatos en la segunda imagen. La mayoría de las investigaciones orientadas a resolver el problema del correspondencia usan las siguientes estrategias para determinar los pixeles con mejor correspondencia:

1. Los *pixeles candidatos* se determinan usando una estrategia de búsqueda de la más baja a la más alta resolución [Anandan, 1989] o imponiendo una restricción física sobre el desplazamiento máximo que puede tomar lugar entre dos imágenes consecutivas de la secuencia. La región cubierta por los pixeles candidatos es llamada *ventana de búsqueda*.
2. La *mejor correspondencia* entre bloques se encuentra maximizando una función de correspondencia (*matching*). La medida de correspondencia o medida de apareo, es usualmente una correlación entre una ventana pequeña alrededor del pixel actual y una ventana alrededor del pixel candidato. El uso de una vecindad para el cálculo de la medida de correspondencia es lógico por la suposición de la conservación de la distribución local de intensidades alrededor del pixel en consideración. Anandan y Scott usaron sumas ponderadas del error cuadrático correspondiente a los pixeles en las dos imágenes [Anandan, 1989]; [Scott, 1988].

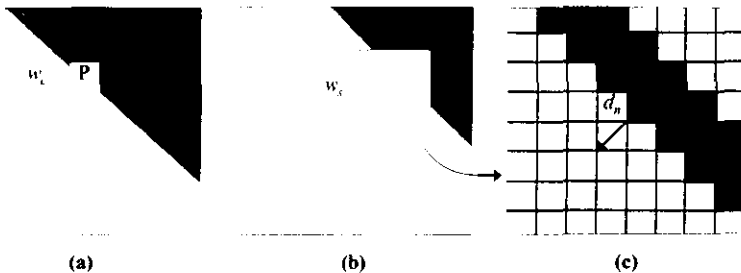


Figura 2.2.- El problema de apertura en los métodos basados en correlación: (a) imagen $t = 1$, (b) imagen $t = 2$ y (c) medida de correspondencia (los valores más oscuros indican una alta correlación)

Si la mejor correspondencia se puede determinar en forma única, entonces el vector de desplazamiento es conocido inmediatamente. Esto es cierto para puntos en la imagen que no sufren del problema de apertura, tales como esquinas y regiones texturizadas. En los puntos de regiones donde el gradiente de la distribución está fuertemente orientado, tales como bordes y

líneas, el comportamiento es muy diferente. En la Figura 2.2 se ilustra lo que pasa con la medida de correspondencia para los puntos cerca de un borde. La ventana de correlación w_c al rededor del pixel P en (a) se evalúa en la función de apareo para cada ventana similar al rededor de los pixeles candidatos que se encuentran dentro de la ventana de búsqueda w_b , mostrada en (b). La medida de correspondencia definida sobre todos los puntos de la ventana de correspondencia se muestra en (c), donde se puede ver que para varios puntos a lo largo del borde se tienen valores similares para la medida de correspondencia, por lo que no es posible resolver la ambigüedad; en otras palabras, el problema de apertura existe y no puede ser resuelto considerando únicamente la medida de correlación.

En su forma más aguda, el problema de apertura puede llevar a escenarios donde varios pixeles en w_c que tienen el mismo valor de correspondencia (el más alto) se vean como una línea recta. Lo mejor que se puede hacer en tales casos es estimar la componente del vector desplazamiento en la dirección del gradiente de intensidades - que es, normal al borde subyacente, señalado en la Figura 2.2-(c). Al igual que en los métodos basados en el gradiente, los métodos basados en correspondencia requieren de restricciones adicionales para poder recuperar el flujo real.

Algunas de las técnicas más importantes que emplean el enfoque de correlación se listan en la Tabla 2.2. Su caracterización obedece a los mismos criterios usados en los métodos basados en el gradiente.

Método	Característica	Flujo de datos	Características
Wong-Hall	Distribución global de la ventana paso-baja de las Intensidades	Si	Global
Nagel	Distribución local de la ventana paso-baja de las Intensidades	Si	Local
Enkelmann	Distribución global de la ventana paso-banda de las Intensidades	Si	Global Proporciona una medida de confianza

Tabla 2.2.- Métodos basados en Correlación

2.5 Estimación basada en la energía espaciotemporal

Algunos trabajos más recientes sobre del movimiento en imágenes se han enfocado en la característica de estímulos visuales en el dominio de la frecuencia [Adelson & Bergen, 1986]; [Fleet & Jepson, 1989]; [Heeger, 1987]. Una forma de ver el movimiento en el dominio de la frecuencia es considerar explícitamente los cambios de brillantez en los patrones visuales como una función de la posición y el tiempo, y examinar su representación frecuencial en las tres dimensiones. Sobre esta línea de investigación se han propuesto modelos de percepción visual humana en varios estudios recientes, en los que las propiedades de movimiento en el dominio de la frecuencias son consideradas en el diseño de filtros espaciotemporales sintonizados a diferentes velocidades y orientaciones. Dichos modelos son igualmente relevantes en visión automática y su principal atractivo recae en el hecho de que el proceso de estimación involucra operaciones básicas similares a las que realiza el sistema de visión humano, las cuales son fácilmente emuladas con un banco de filtros lineales operando en paralelo [Heeger, 1987]; [Heeger, 1988].

La forma más simple de modelar las variaciones espaciotemporales de los registros de intensidades se obtiene suponiendo que los cambios temporales de la imagen se deben exclusivamente a la traslación de patrones espaciales. Considérese por ejemplo el siguiente caso:

Sea $k(\mathbf{r})$ un patrón de imagen tal que al ser trasladado a una velocidad constante $\mathbf{v} = (v_x, v_y)$, genera la función espaciotemporal $l(\mathbf{r}, t)$ de los registros de imagen, esto es que $l(\mathbf{r}, t) = k(\mathbf{r} - \mathbf{v}t)$ para $t \in T$, donde T es el periodo de observación. Tomando la transformada de Fourier en ambos lados de la igualdad se puede ver que $L(\mathbf{k}, \omega) = K(\mathbf{k})\delta(\mathbf{v}\cdot\mathbf{k} + \omega)$, donde $\mathbf{k} = (k_x, k_y)$ es el vector de frecuencias espaciales, ω es la frecuencia temporal y el símbolo $\delta(\cdot)$ denota la función delta de Dirac. Como $\delta(\mathbf{v}\cdot\mathbf{k} + \omega)$ es cero fuera del *plano impulso*

$$\mathbf{v} \cdot \mathbf{k} + \omega = 0,$$

el espectro de potencia del patrón trasladado queda entonces confinado a dicho plano. El plano impulso pasa por el origen y su normal $\mathbf{N} = (v_x, v_y, 1)$ codifica directamente la velocidad de traslación del patrón k .

Si el patrón estático es suficientemente texturizado se puede usar un banco de filtros espaciotemporales sintonizados a diferentes velocidades para hacer un muestreo del espectro de potencia del patrón trasladado y detectar el plano impulso. Estos filtros deben permitir el computo local de las velocidades ópticas. Durante la última década se han propuesto varias transformadas para el diseño de estos filtros, incluyendo la transformadas de Fourier en tiempo corto [Weng, 1990], de Wigner-Ville, tensores, y la transformada de Gabor [Heeger, 1987, 1988]; [Watson & Ahumada, 1985]; [Adelson & Bergen, 1985, 1986]. En general, el desempeño de estas técnicas es bastante aceptable cuando se trata de patrones no orientados. El problema de apertura también ha sido considerado explícitamente en el diseño de filtros

que analizan orientaciones espaciotemporales y permiten el cálculo de la componente normal de velocidad [Fleet & Jepson, 1989].

La conformidad con experimento psicofísicos también ha sido de especial interés en el diseño de estos filtros [Watson & Ahumada, 1985], siendo la transformada de Gabor la propuesta para el modelo de los campos receptivos. El modelo alternativo que se propone en el presente trabajo difiere de la transformada de Gabor en que éste no elimina la ortogonalidad de las funciones base, resultando así en un modelo más eficiente.

2.6 Resumen

En este capítulo se presentó una sinopsis de los problemas y métodos para el análisis de movimiento en imágenes, exponiendo los modelos más ampliamente conocidos y las hipótesis que los sustentan. Vimos que el problema de apertura no depende del enfoque de análisis, sino de la inherente restricción de cómputo local impuesta por la complejidad de las deformaciones espaciotemporales de los registros de imagen. También se listaron las técnicas más populares en la literatura para los enfoques de gradiente y correlación. Hago notar aquí que dichos métodos se consideraron sin tomar en cuenta una etapa de suavizado del campo, el cual se lleva a cabo mediante un proceso de optimización global y se refiere a una restricción espacial sobre las velocidades ópticas. Como se señaló en la sección 2.3.2, el objetivo de la restricción de suavidad del campo es intentar resolver el problema de apertura por medio de introducción de contexto.

Capítulo 3

Transformadas Polinomiales

3.1 Introducción

Al igual que en el caso de la visión humana, para muchas aplicaciones de visión por computadora y procesamiento de imágenes se requiere que los datos de imagen, los cuales están dados como un arreglo de valores de intensidades, sean interpretados en patrones visualmente significativos, lo que generalmente requiere determinar las relaciones espaciotemporales entre las intensidades. Este tipo de procesamiento involucra análisis local, esto es, que la señal se multiplica por una ventana cuyo radio de influencia determina el conjunto de puntos que contribuyen en la operación de procesamiento básica y cuya forma establece el peso relativo de la contribución para cada punto. Con el fin de describir una imagen completamente, este procesamiento local tiene que repetirse para un número suficiente de posiciones de la ventana. La primera decisión que debe hacerse en este enfoque es entonces, elegir la forma y tamaño de la ventana. La segunda decisión es la elección del operador básico a emplear sobre la *señal ventaneada*, lo que equivale a buscar los patrones que son considerados más relevante a *priori* [Martens, 1990a].

Hacer una elección óptima de la *ventana* y del tipo de procesamiento es sin lugar a dudas un difícil compromiso si nos basamos solamente en argumentos teóricos. Es aquí donde los conocimientos sobre el sistema de visión humano aportan una buena referencia. En un principio las investigaciones sobre el sistema visual humano se concentraban en las funciones de Gabor para describir los procesos en la retina y la corteza cerebral teniendo como principal argumento que el perfil de los campos receptivos puede ser modelado con *funciones de Gabor*. Estudios posteriores demostraron, sin embargo, que las derivadas de Gaussianas pueden modelar más eficientemente a los campos receptivos, con la ventaja de requerir menos parámetros [Daugman, 1980], [Young, 1986]. Asimismo, se ha encontrado (mediante experimentos psicofísicos) que el procesamiento de estímulos temporales en el sistema visual es realizado por operadores similares a las derivadas de Gaussianas (derivadas de una distribución Gama) [den Brinker & Roufs, 1992]. Estos operadores son integrados formalmente en la técnica de descomposición local conocida como *transformada polinomial*

(TP), la cual presentamos en este capítulo. Para una elección específica de una ventana correspondiente a una distribución Gausiana, la descomposición resultante es la transformada de Hermite¹ (TH). La versión discreta de esta transformada resulta al considerar una ventana binomial, la cual se discute brevemente dentro de la sección 3.3. El otro caso particular se tiene cuando se elige una ventana correspondiente a una distribución Gama, resultando en una descomposición conocida como la transformada de Laguerre generalizada² (TLG). Para este caso no existe una teoría de la versión discreta, sin embargo en la sección 3.4 se discuten algunas ideas para aproximar el comportamiento continuo con una ventana discreta derivada por un proceso de límite. También se introduce en este capítulo una notación vectorial para la representación de imágenes estáticas y dinámicas en el dominio de la TP.

3.2 Definición

El análisis por descomposición local involucra dos pasos. En el primer paso, con el fin de limitar el análisis a una región de interés, la señal de entrada $l(x)$, para $x \in \mathfrak{I}$, es multiplicada por una función estrictamente positiva $w(x)$ llamada *ventana*. El conjunto \mathfrak{I} es un subconjunto (continuo o discreto) del espacio euclidiano D -dimensional, siendo $D = 2$ y 3 para imágenes estáticas y secuencias de imágenes respectivamente. Para describir a la señal completamente se coloca la ventana en un número suficientes de posiciones $p \in P$, dentro de la rejilla de muestreo $P \subseteq \mathfrak{I}$. En el segundo paso, cada una de las porciones de la señal dentro de la ventana se descompone en una suma ponderada de la base ortonormal $\{\varphi_n(x-p)w(x-p), n \in \aleph\}$. Esta descomposición se expresa en términos generales como

$$w(x-p) \left[l(x) - \sum_{n \in \aleph} l_n(p) \varphi_n(x-p) \right] = 0 \quad (3.1)$$

donde el conjunto discreto \aleph indiza las funciones de la base. Los coeficientes de la expansión se obtienen al multiplicar (3.1) por $\varphi_n(x-p)w(x-p)$ y aplicar la condición de ortonormalidad, lo que resulta en

$$l_n(p) = \int_{\mathfrak{I}} l(x) \varphi_n(x-p) dx \quad (3.2)$$

para $n \in \aleph$ y $p \in P$, donde

¹ El resumen de la teoría de la transformada de Hermite está basado en *The Hermite Transform -Theory* (Martens, 1990a) con los ajustes expuestos en *Local Orientation Analysis in Images by Means of the Hermite Transform* (Martens, 1997) en la función de peso.

² El resumen de la transformada de Laguerre generalizada está basado en *Evidence for a generalized Laguerre transform of temporal events by the visual system* (den Brinker & Roufs, 1992). Aquí se redefinen algunos parámetros, se introduce una constante de normalización en la ventana y se hacen los mismos ajustes que en el caso Hermitiano para la función de peso.

$$a_n(x) = \varphi_n(x)w^2(x) \quad (3.3)$$

es la *función de análisis* de orden n . También es común denominar la función $a_n(-x)$ como *filtro de análisis* de orden n , ya que los coeficientes de la expansión pueden ser obtenidos convolucionando la señal de entrada con estos filtros y seleccionando la salida en cada punto p de la rejilla P .

La reconstrucción de la señal a partir de los coeficientes de la expansión se obtiene mediante la suma de las descomposiciones locales sobre todas las posiciones $p \in P$, siempre que la función periódica, referida como *función de peso*,

$$w_p^2(x) = \sum_{p \in P} w^2(x-p) \quad (3.4)$$

sea diferente de cero para toda $x \in \mathfrak{I}$. Multiplicando (3.1) por $w(x-p)$, sumando sobre las posiciones $p \in P$, y dividiendo por la función de peso, resulta la siguiente expresión para la reconstrucción de la señal:

$$l(x) = \sum_{p \in P} \sum_{n \in \mathfrak{N}} l_n(p) s_n(x-p) \quad (3.5)$$

donde

$$s_n(x) = \frac{a_n(x)}{w_p^2(x)} \quad (3.6)$$

es la *función de síntesis* de orden n , también referida como *filtro de interpolación*.

Cuando las funciones $\varphi_n(x)$, $n \in \mathfrak{N}$, son polinomios $\rho_n(x)$, la descomposición de la señal de entrada $l(x)$ en los coeficientes (3.2) recibe el nombre de *transformada polinomial* (TP), mientras que a la reconstrucción de la señal con (3.5) a partir de los coeficientes se le denomina *transformada polinomial inversa* (TPI). En este caso, los polinomios ortonormales $\rho_n(x)$ están determinados en forma única por la ventana $w^2(x)$ [Szegő, 1959]. Para ventanas separables, esto es, $w(x) = w_1(x_1) \cdot w_2(x_2) \cdots w_D(x_D)$ con $x = (x_1 \ x_2 \ \cdots \ x_D)$, los polinomios ortonormales también son separables, es decir, $\rho_n(x) = \rho_{n_1}^{1}(x_1) \cdot \rho_{n_2}^{2}(x_2) \cdots \rho_{n_D}^D(x_D)$ con $n = n_1 + n_2 + \dots + n_D$. Cada polinomio unidimensional $\rho_{n_i}^i(x_i)$ se puede obtener en términos de la ventana unidimensional $w_i^2(x_i)$.

Para el caso unidimensional los polinomios $\rho_n(x)$ que resultan de ortonormalizar la base canónica $\{1, x, x^2, \dots\}$ están dados por el determinante

$$\rho_n(x) = \frac{1}{\sqrt{M_{n-1}M_n}} \begin{pmatrix} c_0 & c_1 & c_2 & \dots & c_n \\ c_1 & c_2 & c_3 & \dots & c_{n+1} \\ \vdots & \vdots & \vdots & & \vdots \\ c_{n-1} & c_n & c_{n+1} & \dots & c_{2n+1} \\ 1 & x & x^2 & \dots & x^n \end{pmatrix} \quad (3.7)$$

para $n = 0, 1, \dots$, donde $M_n = |c_{i+j}|_{j,i=0,\dots,n}$, $M_{-1} = 1$, y

$$c_n = \int_{-\infty}^{+\infty} x^n w^2(x) dx \quad (3.8)$$

es el momento de orden n de la función $w^2(x)$.

De la ortogonalidad de los polinomios $\rho_n(x)$ se tiene que

$$\int x^m a_n(x) dx = 0$$

para $m = 0, \dots, n-1$, lo que implica el filtro $a_n(-x)$ es insensible a variaciones polinomiales de grado menor a n .

Una condición suficiente para garantizar la convergencia de la expansión polinomial es que $l(x)$ sea analítica y acotada. Así, el error de aproximación entre la señal y un polinomio puede hacerse arbitrariamente pequeño tomando un grado suficientemente grande para la expansión.

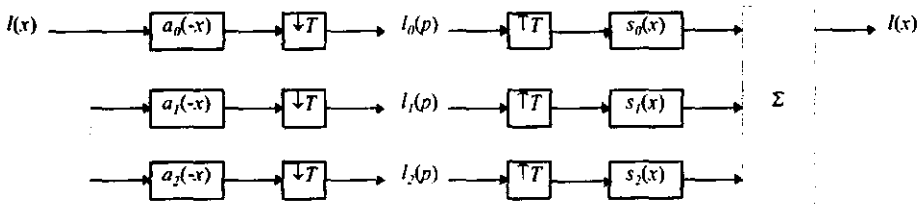


Figura 3.1 Análisis y síntesis de la transformada polinomial

En la Figura 3.1 se ilustran esquemáticamente los bloques de la TP y la TPI. Los bloques etiquetados con la letra T y una flecha representan un submuestreo si la flecha apunta hacia abajo o un sobremuestreo si la flecha apunta hacia arriba, considerando en ambos casos el mismo intervalo T . Los bloques etiquetados con $a_n(-x)$ y $s_n(x)$ representan los operadores lineales de análisis y síntesis respectivamente.

3.3 Transformada de Hermite

La transformada de Hermite (TH) es un caso particular de la TP que ha cobrado especial interés en el campo de procesamiento de imágenes durante los últimos años. Esta transformada resulta al elegir una ventana específica que corresponde a una distribución Gaussiana. Las razones para centrar la atención en ventanas Gaussianas son varias. Primero, la teoría es ampliamente conocida; su tratamiento posee ciertas bondades de tal suerte que las propiedades de la transformada pueden ser fácilmente derivadas y evaluadas; su aproximación discreta es matemáticamente tratable y sus propiedades están estrechamente relacionadas con el caso continuo. En segundo lugar, las ventanas Gaussianas en posiciones separadas por el doble de la desviación σ son un buen modelo de los campos receptivos traslapados encontrados en experimentos fisiológicos [Sakitt & Barlow, 1982]. Tercero, la descomposición involucra los operadores derivadas de Gaussianas, encontrados ampliamente en el campo de visión por computadora [Bevington & Mersereau, 1987] y en el modelado psicofísico del sistema visual humano [Marr & Hildreth, 1980]. Y por último, y no menos importante, las ventanas Gaussianas minimizan el producto de incertidumbre en el dominio espacial y frecuencial.

Como la transformada hermitiana D -dimensional se puede definir en función de transformadas hermitianas 1-D sobre cada coordenada, sólo se presenta aquí el caso unidimensional. La ventana gaussiana 1-D se define como

$$w^2(x; \sigma) = \frac{1}{\sqrt{\pi}\sigma} e^{-(x/\sigma)^2} \quad (3.9)$$

donde σ es un parámetro de escala y el factor $1/\pi^{1/2}\sigma$ se introduce para normalizar la energía. Los polinomios ortonormales respecto a $w^2(x)$ corresponden, salvo por una constante, a los polinomios de Hermite, es decir,

$$\rho_n(x; \sigma) = \frac{1}{\sqrt{2^n n!}} H_n(x/\sigma) \quad (3.10)$$

para $n = 0, \dots, \infty$, donde $H_n(x)$ denota el polinomio de Hermite de grado n

$$H_n(x) = n! \sum_{k=0}^{\lfloor n/2 \rfloor} (-1)^k \frac{(2x)^{n-2k}}{k!(n-2k)!}$$

Por simplicidad escribiremos $w(x)$ y $\rho_n(x)$ en lugar de $w(x; \sigma)$ y $\rho_n(x; \sigma)$ ya que para esta definición de la transformada de Hermite la escala σ es constante. La definición de la transformada hermitiana a múltiples escalas se puede ver en [Martens, 1990b].

Las propiedades más importantes de la transformada de Hermite se determinan sobre los filtros de análisis, cuya respuesta al impulso puede obtenerse como derivadas sucesivas de la ventana, esto es

$$a_n(-x) = \frac{1}{\sqrt{2^n n!}} \frac{d^n}{d(x/\sigma)^n} w^2(x) \quad (3.11)$$

para $n = 0, \dots, \infty$. A partir de esta relación se obtiene directamente su representación en frecuencia

$$A_n^*(k) = \frac{(j\sigma k)^n}{\sqrt{2^n n!}} e^{-(\sigma k/2)^2} \quad (3.12)$$

donde k denota frecuencia espacial.

La función de peso requerida para la construcción de los filtros de interpolación $s_n(x) = a_n(x)/w_\lambda^2(x)$ de la transformada de Hermite inversa se expresa en términos de su serie de Fourier por

$$w_\lambda^2(x) = \frac{1}{\lambda} \left[1 + 2 \sum_{i=0}^{\infty} e^{-\left(\frac{i\pi\sigma}{\lambda}\right)^2} \cos\left(i \frac{2\pi x}{\lambda}\right) \right] \quad (3.13)$$

para un muestreo uniforme igual a λ . En general es deseable limitar el número de descomposiciones locales de la señal (especialmente en las aplicaciones de codificación), por lo que conviene escoger λ tan grande como sea posible. Desafortunadamente y sobre todo en aplicaciones digitales, para valores grandes de λ la división por $w_\lambda^2(x)$ puede introducir cierta sensibilidad a la variabilidad espacial en los filtros de interpolación.

3.3.1 Aproximación Discreta

Como se ha mencionado, uno de los principales problemas que se tienen en la práctica al trabajar con imágenes y secuencias de imágenes discretas, es precisamente el hecho de tener que trabajar con las muestras de una señal continua; siendo estas muestras la única información disponible para extraer las características relevantes tales como posición, orientación, velocidad, forma, textura, etc. Para llevar a cabo dicha extracción de características en forma eficiente a partir de los datos, se requieren herramientas de análisis en cuyo desarrollo no se tenga que ligar explícitamente una señal continua a la señal discreta en cuestión, de ahí que en el presente trabajo se incluyan algunos de los resultados del enfoque discreto de la transformada de Hermite (THD).

Para poder realizar físicamente la transformada polinomial de una señal discreta $f[x]$, se requiere que la ventana sea de longitud finita, esto es que $w[x] = 0$ para $x < N_1$ y $x > N_2$. De esta manera la expansión (3.1) tiene un orden finito $N = N_2 - N_1$ y puede ser realizada en computadora. Los resultados de la sección 3.2 son también válidos para el caso discreto, solo se tienen que remplazar las integrales por sumas finitas sobre el rango apropiado.

El equivalente de la ventana Gaussiana en el dominio discreto es la bien conocida distribución binomial

$$w^2[x; N] = \frac{1}{2^N} C_N^x \quad (3.14)$$

para $x = 0, \dots, N$. Los polinomios ortonormales asociados a esta ventana corresponden a los polinomios de Krawtchouk [Szegő, 1959]

$$\rho_n[x] = \frac{1}{\sqrt{C_N^n}} \sum_{k=0}^n (-1)^{n-k} C_{N-x}^{n-k} C_x^k$$

Por otro lado, para valores grandes de N , la ventana binomial se aproxima a una Gaussiana. Específicamente

$$\lim_{N \rightarrow \infty} w^2[x - N/2] = \frac{1}{\sqrt{\pi N/2}} e^{-\left(x/\sqrt{N/2}\right)^2} \quad (3.15)$$

El mismo proceso de límite convierte a los polinomios de Krawtchouk en los polinomios continuos (3.10),

$$\lim_{N \rightarrow \infty} \rho_n[x - N/2] = \frac{1}{\sqrt{2^n n!}} H_n\left(x/\sqrt{N/2}\right)$$

lo que implica que las propiedades de la THD de orden N , pueden predecirse con cierta precisión a partir las propiedades de la TH con parámetro de escala $\sigma = \sqrt{N/2}$.

Una restricción que conviene imponer a la ventana binomial es que N sea par, de esta manera las funciones de análisis se definen desplazando la ventana binomial $N/2$ unidades para tener origen como centro de simetría, es decir,

$$a_n[x] = \rho_n[x - N/2] w^2[x - N/2] \quad (3.16)$$

para $x = -N/2, \dots, N/2$. Estas funciones pueden expresarse también como

$$a_n[x - N/2] = \frac{(-1)^n}{\sqrt{C_N^n}} \Delta^n \{C_N^n w^n[x]\} \tag{3.17}$$

donde $\Delta^n(\cdot)$ es el operador diferencia de orden n .

N	n	$d_n[x], x = -N/2, \dots, N/2$						
2	0	1		2		1		
	1	-1		0		1		
	2	1		-2		1		
4	0	1	4	6	4	1		
	1	-1	-2	0	2	1		
	2	1	0	-2	0	1		
	3	-1	2	0	-2	1		
	4	1	-4	6	-4	1		
6	0	1	6	15	20	15	6	1
	1	-1	-4	-5	0	5	4	1
	2	1	2	-1	-4	-1	2	1
	3	-1	0	3	0	-3	0	1
	4	1	-2	-1	4	-1	-2	1
	5	-1	4	-5	0	5	-4	1
	6	1	-6	15	-20	15	-6	1

Tabla 3.1.-Filtros de análisis de la THD de orden 2, 4 y 6.

En la Tabla 3.1 se listan los coeficientes de los filtros denormalizados

$$d_n[x] = \frac{2^n}{\sqrt{C_N^n}} a_n[N/2 - x]$$

para $N = 2, 4$ y 6 .

Estos operadores han probado ser buenos detectores de las características locales de imágenes estáticas con la importante ventaja de que pueden ser eficientemente realizados para valores de N muy grandes aprovechando el hecho que todos ellos resultan de convoluciones sucesivas entre las secuencias $[1,1]$ y $[1,-1]$ [Hashimoto & Skalansky, 1987].

3.4 Transformada de Laguerre generalizada

Martens ha sugerido que la transformada de Hermite en tres dimensiones puede ser usada para codificar los eventos espaciotemporales de estímulos visuales [Martens, 1990a]. Por otro lado, existe la hipótesis de que la codificación de eventos temporales en el sistema visual toma la forma de la transformada de Laguerre generalizada (TLG) como lo sugieren experimentos psicofísicos recientes [den Brinker & Roufs, 1992]. Como mostraremos en esta sección, ambas transformadas tienen sorprendentes similitudes en la operación de análisis, con la diferencia en la causalidad de los filtros. En esta sección derivamos las condiciones bajo las cuales el análisis de la transformada de Laguerre generalizada puede ser aproximada por la transformada de Hermite.

La teoría de la TLG se introduce siguiendo esencialmente la misma idea que en el caso gaussiano, pero teniendo ahora una señal temporal $l(t)$ conocida hasta el tiempo actual t_0 . Evidentemente la ventana de análisis debe considerar solamente valores sobre la porción conocida de la señal, esto es, $w(t-t_0) \equiv 0$ para $t > t_0$. El análisis de la porción de la señal dentro de la ventana debe hacerse para varios instantes discretos t_0 . Por simplicidad se consideran instantes uniformemente espaciados por un periodo de muestreo T . La ventana en este caso es la distribución Gama con argumento $-t$ la cual se define como

$$w^2(t; \alpha, \beta) = \frac{1}{\beta \Gamma(\alpha)} (-t/\beta)^{\alpha-1} e^{-t/\beta} u_{-1}(-t) \quad (3.18)$$

donde $\alpha, \beta > 0$ son los parámetros de sesgo y escala respectivamente y $u_{-1}(t)$ es la función escalón unitario ($u_{-1}(t) = 1$ para $t \geq 0$ y $u_{-1}(t) = 0$ para $t < 0$). En este desarrollo restringimos el parámetro α a valores enteros mayores que uno. Se ha encontrado que para valores de α entre 3 y 10 inclusive y $\beta \approx 45/(\alpha-1)^{1/2}$ los filtros ajustan perfectamente los resultados de experimentos psicofísicos [den Brinker & Roufs, 1992].

Los polinomios asociados a la ventana (3.18) corresponden, salvo por una constante, a los polinomios de Laguerre generalizados (o polinomios de Laguerre cuando $\alpha = 1$), es decir,

$$\rho_n(t) = \left(\frac{\Gamma(\alpha) n!}{\Gamma(n+\alpha)} \right)^{1/2} L_n^{\alpha-1}(-t/\beta) \quad (3.19)$$

donde $L_n^{(\alpha)}$ denota el polinomio de Laguerre de grado n con parámetro α definido como

$$L_n^{(\alpha)}(t) = \sum_{k=0}^n \frac{\Gamma(n+\alpha)}{\Gamma(k+\alpha)(n-k)!k!} (-t)^k$$

Los filtros de la transformada de Laguerre satisfacen una propiedad similar a (3.11) para el caso hermitiano, la cual se expresa como

$$a_n(-t) = \left(\frac{\Gamma(\alpha)}{n!\Gamma(n+\alpha)} \right)^{1/2} \frac{d^n}{d(t/\beta)^n} w^2(-t; \alpha+n) \quad (3.20)$$

o en el dominio de la frecuencia como

$$A_n^*(\omega) = \left(\frac{\Gamma(n+\alpha)}{n!\Gamma(\alpha)} \right)^{1/2} \frac{(j\beta\omega)^n}{(1+j\beta\omega)^{n+\alpha}} \quad (3.21)$$

Como puede verse estos filtros también llevan a cabo la operación de derivada y una operación de suavizado de la señal, pero a diferencia de los operadores derivadas de Gaussiana que intervienen en la TH el *difuminado* que estos operadores hacen sobre la señal varía con el orden de derivación dado que el denominador de (3.21) depende de n .

Para valores grandes de α y β pequeño, y siempre que $\alpha\beta^2$ se mantenga constante, se demuestra que los filtros (3.21) se aproximan a los filtros hermitianos retardados $\alpha\beta$ unidades. La demostración es inmediata si escribimos (3.21) como

$$A_n^*(\omega) = \left(\frac{\Gamma(n+\alpha)}{n!\Gamma(\alpha)} \right)^{1/2} \left(\frac{j\beta\omega}{j\beta\omega+1} \right)^n (1+\beta^2\omega^2)^{-\alpha/2} e^{-j\alpha \tan^{-1}\beta\omega}$$

y empleamos las relaciones asintóticas

$$\frac{\Gamma(n+\alpha)}{\Gamma(\alpha)} \sim \alpha^n, \quad \frac{j\beta\omega}{j\beta\omega+1} \sim j\beta\omega, \quad (1+\beta^2\omega^2)^{-\alpha/2} \sim \exp\left(-\frac{\alpha\beta^2\omega^2}{2}\right), \quad \alpha \tan^{-1}\beta\omega \sim \alpha\beta\omega$$

para $\alpha \rightarrow \infty$, $\beta \rightarrow 0$ y $\alpha\beta^2 = cte$, es decir que,

$$A_n^*(\omega) \sim \frac{(j\beta\sqrt{2\alpha\omega})^n}{\sqrt{2^n n!}} e^{-(\beta\sqrt{\alpha/2\omega})^2} e^{-j\alpha\beta\omega} \quad (3.22)$$

para $\alpha \rightarrow \infty$, $\beta \rightarrow 0$ y $\alpha\beta^2 = cte$. Es claro que (3.22) corresponde al filtro hermitiano de orden n retardado $\alpha\beta$ unidades y con parámetro de escala $\sigma^2 = 2\alpha\beta^2$.

El resultado anterior implica que para α suficientemente grande y β suficientemente pequeño, la TH es una buena aproximación de la TLG, sin embargo como hemos hecho notar, los valores de α que tienen una interpretación válida en el sistema visual son en general pequeños. Además, elegir valores grandes de α implica que la señal es principalmente analizada en término de eventos pasados mientras que para valores pequeños de α la señal se interpreta más en término de eventos presentes.

Las funciones de síntesis requeridas para la reconstrucción de la señal se derivan de la definición general resultando $s_n(t) = a_n(t) / w_T^2(t)$, donde la función de peso $w_T^2(t)$ se expresa por su serie de Fourier

$$w_T^2(t) = \frac{2\pi}{T} \left\{ 1 + 2 \sum_{i=0}^{\infty} \left[1 + \left(i \frac{2\pi\beta}{T} \right)^2 \right]^{-\alpha/2} \cos \left(i \frac{2\pi\beta t}{T} + \varphi_i \right) \right\} \quad (3.23)$$

con

$$\varphi_i = \alpha \tan^{-1} \left(i \frac{2\pi\beta}{T} \right) \quad (3.24)$$

En este caso se observa que las funciones de síntesis tienen una respuesta no causal ya que los filtros de análisis son causales. Así, en un sistema práctico, la reconstrucción de la señal original a partir de los coeficientes de la TLG solo es posible en forma aproximada permitiendo un retardo suficientemente grande y un truncamiento para $t < 0$. El retardo necesario para producir un error máximo aceptable dependerá directamente del parámetro α y del número de términos considerados en la expansión.

3.4.1 Aproximación Discreta

Una de las desventajas de usar la transformada de Laguerre, y quizá la mayor de ellas, es que no se cuenta con una aproximación discreta estudiada. En esta sección presentamos una aproximación de la ventana Gama, sin embargo dada la complejidad matemática que presenta, no proporcionamos una expresión matemática para los polinomios ortonormales, pero el cálculo numérico de éstos puede hacerse mediante la expresión (3.7).

La aproximación discreta de una distribución Gama corresponde a la distribución binomial

inversa [Pessing, 1963]

$$P(x; m, p) = C_{x+m-1}^x p^m (1-p)^x \tag{3.25}$$

para $x = 0, 1, \dots$ con parámetros $0 < p \leq 1$ y $m > 0$. La aproximación se establece mediante el límite

$$\lim_{p \rightarrow 0} P(t/\beta p - m + 1) = \frac{p}{\beta \Gamma(m)} (t/\beta)^{m-1} e^{-t/\beta}$$

con β y $m > 0$ constantes. La ventana que nos interesa es una versión truncada y reflejada de la distribución binomial inversa.

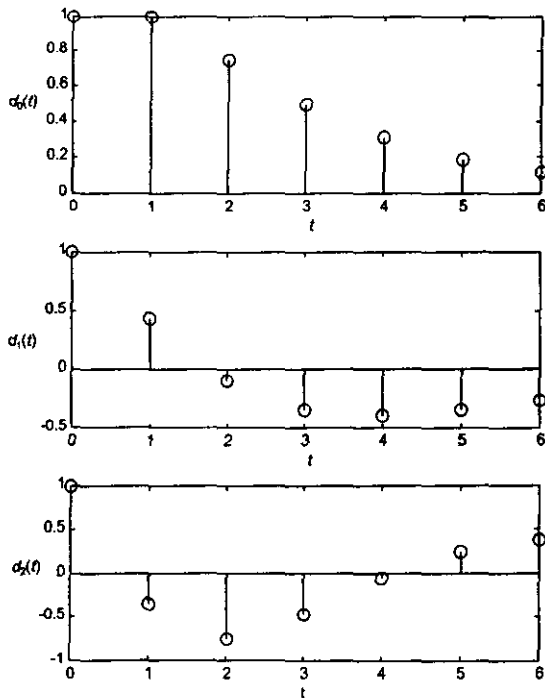


Figura 3.2 Filtros de la aproximación discreta de la TLG (N = 6).

En particular para $p = 0.5$,

$$w^2[t; N] = \frac{K}{2^{m+N-t}} C_{N-t+m-1}^{N-t} \quad (3.26)$$

con $t = 0, 1, \dots, N$, donde K es una constante de normalización y $N \geq 2m$ para que el sesgo sea positivo.

La Figura 3.1 muestra los filtros hasta orden 2 para $m = 2$ y $N = 6$. Los filtros están escalados de tal forma que el máximo valor es 1.

3.5 Representación de imágenes

Una imagen dinámica monocromática $l(\mathbf{r}, t)$ es una señal de luminancia que depende de una posición $\mathbf{r} = (x, y)$ en el plano imagen y del instante t . Usamos entonces $l(\mathbf{r}, t)$ y $l(x, y, t)$ equivalentemente. Cuando se fija un instante de tiempo específico o no se presentan cambios temporales en los registros de luminancia durante el período de observación, se dice que se tiene una imagen estática. En tal caso la dependencia del tiempo se elimina de la notación, es decir, se usa $l(\mathbf{r})$.

En la sección 3.2 se introdujo el conjunto \mathbb{N} para indizar las funciones de la expansión. Sin embargo, para el caso específico de imágenes se define un orden de indizado en relación con el grado de los polinomios ortonormales. Denotamos por $\rho_{m,n}(\mathbf{r})$ al conjunto de polinomios bidimensionales de orden n , para $m = 0, \dots, n$ asociados a la expansión polinomial de una imagen estática y por $\rho_{m,m,p}(\mathbf{r}, t)$ al conjunto de polinomios tridimensionales de orden n , para $m = 0, \dots, n$ y $n = 0, \dots, p$ asociado a la expansión polinomial de una imagen dinámica. Un indizado similar se usa para las funciones definidas en términos de estos polinomios y para los coeficientes de la expansión.

Como se tienen exactamente $n+1$ polinomios bidimensionales y $[(n+1)^2+n+1]/2$ polinomios tridimensionales de grado n , conviene introducir una notación para referirnos a los coeficientes polinomiales de orden n . En una expansión bidimensional usaremos el vector columna

$$\mathbf{l}_n = \begin{bmatrix} l_{0,n} \\ l_{1,n-1} \\ \vdots \\ l_{n-1,1} \\ l_{n,0} \end{bmatrix} \quad (3.27)$$

para referirnos a los coeficientes de orden n correspondientes a una posición genérica de la ventana. Y para los de la expansión tridimensional del mismo orden usaremos la matriz

triangular superior

$$\mathbf{L}_n = \begin{bmatrix} l_{0,0,n} & l_{0,1,n-1} & \dots & l_{0,n-1,1} & l_{0,n,0} \\ 0 & l_{1,0,n-1} & \dots & l_{1,n-2,1} & l_{1,n-1,0} \\ 0 & 0 & \vdots & \vdots & \vdots \\ \vdots & \vdots & \dots & l_{n-1,0,1} & l_{n-1,1,0} \\ 0 & 0 & 0 & 0 & l_{n,0,0} \end{bmatrix} \quad (3.28)$$

Esta notación se extiende para las funciones indizadas, así, $\rho_n(\mathbf{r})$ y $P_n(\mathbf{r},t)$ denotan el vector y la matriz triangular de polinomios 2-D y 3-D de grado n respectivamente.

La expansión local usando esta notación vectorial se expresa como

$$\left[l(\mathbf{r}) - \sum_{n=0}^{\infty} l_n^T \rho_n(\mathbf{r}) \right] w(\mathbf{r}) = 0 \quad (3.29)$$

para la imagen estática $l(\mathbf{r})$, y como

$$\left[l(\mathbf{r},t) - \sum_{n=0}^{\infty} tr \left\{ l_n^T P_n(\mathbf{r},t) \right\} \right] w(\mathbf{r},t) = 0 \quad (3.30)$$

para la imagen dinámica $l(\mathbf{r},t)$, donde $tr(\cdot)$ denota la traza de la matriz (\cdot) .

De las expresiones anteriores observamos que en ambos casos los términos de la expansión pueden ser escritos como un producto escalar entre los vectores (o matrices) de coeficientes y de polinomios, si se definen los productos escalares entre dos vectores columna \mathbf{a} , \mathbf{b} y dos matrices cuadradas \mathbf{A} , \mathbf{B} con su respectiva norma euclidiana como

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{a}^T \mathbf{b}, \quad \mathbf{a} \cdot \mathbf{a} = \|\mathbf{a}\|^2$$

y

$$\mathbf{A} \cdot \mathbf{B} = tr \left\{ \mathbf{A}^T \mathbf{B} \right\} \quad \mathbf{A} \cdot \mathbf{A} = \|\mathbf{A}\|^2$$

respectivamente.

Usando estas definiciones podemos escribir la generalización del teorema de Parseval para la descomposición local de imágenes como

$$\int l^2(\mathbf{r})w^2(\mathbf{r})d\mathbf{r} = \sum_{n=0}^{\infty} \|l_n\|^2$$

$$\iint l^2(\mathbf{r},t)w^2(\mathbf{r},t)d\mathbf{r}dt = \sum_{n=0}^{\infty} \|L_n\|^2$$
(3.31)

En ocasiones es común referirse a cada término de los miembros de la derecha como energía de orden n . En los capítulos subsecuentes se hace la distinción entre la energía de orden cero y la energía de orden superior. La primera es la energía contenida en el coeficiente de orden cero y se le denomina energía de DC; la segunda es la suma de las energías de orden 1,2... y se le denomina energía de AC. La razón de esta nomenclatura es porque el único coeficiente que contiene la componente de DC de la señal es justamente el coeficiente de orden cero.

3.6 Resumen

En general, los parámetros libres de la transformada polinomial, son la ventana y la rejilla de muestreo. Para una ventana Gaussiana y una rejilla cuadrada, el parámetro libre es la desviación estándar σ (en el caso discreto, el orden de la ventana binomial N) y intervalo de muestreo λ . Cuando la ventana corresponde a la distribución Gama y se elige una rejilla uniformemente espaciada, los parámetros libres son la escala β y el parámetro de sesgo α (y en el caso discreto la longitud de la ventana N y el parámetro de sesgo m). En ambos casos los parámetros pueden ser elegidos con base en los resultados psicofísicos, o en la experiencia sobre aplicaciones conocidas.

El uso de los filtros de Laguerre en lugar de los filtros de Hermite para la codificación de eventos temporales es intuitivamente razonable si pensamos que el sistema visual tiende a interpretar una escena más por los eventos inmediatos que por los eventos pasados. Lo que sucede al considerar valores de α muy grandes es que la interpretación instantánea de un estímulo se basa justamente en eventos pasados más que en eventos actuales.

Alternativamente a la ventana binomial inversa (3.26) para la TLD se puede emplear una ventana binomial sesgada,

$$w^2[t; N] = C_N^t p^t (1-p)^{N-t}$$
(3.32)

con $t = 0, \dots, N$ y con $p > 0.5$ para producir un sesgo positivo. Si se escoge un valor de p apropiado, para valores no muy grandes de N se observa que los filtros generados se aproximan a aquellos generados con la ventana binomial inversa. Los polinomios ortonormales respecto a la ventana binomial sesgada siguen siendo los polinomios de Krawtchouk (ver [Szegő, 1959] para la expresión general de estos polinomios).

Capítulo 4

Representación 2-D de imágenes dinámicas

4.1 Introducción

En este capítulo se presenta el desarrollo teórico de una nueva técnica para estimar el movimiento de patrones no orientados usando la información contenida en los coeficientes de la transformada polinomial. También se proporciona una sinopsis de los pasos más importantes del método y las consideraciones prácticas para su realización en un programa de computadora. Al final se discuten resultados experimentales obtenidos para secuencias específicas.

La hipótesis fundamental de nuestro análisis es que la imagen dinámica es generada por patrones visuales que se desplazan bajo condiciones de iluminación invariables siguiendo una trayectoria de movimiento continua en todo el período de observación. Partiendo de esta suposición se minimiza una medida de distorsión entre el patrón bidimensional y la imagen dinámica de entrada con el fin de estimar ciertos parámetros de orientación espaciotemporal y, a partir de estos, se calculan las velocidades locales¹.

Aunque no es el objetivo de esta tesis, este enfoque permite la reconstrucción de la secuencia original a partir del patrón estático y de las características de movimiento (u orientación) determinadas, lo que lo hace atractivo para aplicaciones de codificación de vídeo.

4.2 Coordenadas intrínsecas

Para extraer el patrón estático a partir de la imagen dinámica es necesario el uso de un sistema de *coordenadas intrínsecas* a la trayectoria de movimiento. Un sistema de coordenadas intrínsecas permite analizar las características estáticas del patrón que genera la imagen dinámica porque se ajusta a la forma de la trayectoria. Por ejemplo, considere una

¹ Una versión que no considera orientaciones espaciotemporales fue reportada en "Polynomial transform application to motion estimation", *Proc. of SPIE*, vol. 3457, 1998 (Silván-Cárdenas & Escalante-Ramírez).

trayectoria como la de la Figura 4.1-a. El sistema de coordenadas que conviene emplear para referir las variaciones de intensidad en el instante t_0 debe ser una versión rotada del sistema original de tal forma que la velocidad en el instante de interés es nula en ese nuevo sistema y no hay variaciones de intensidad a lo largo de la nueva coordenada temporal. Es evidente que el ángulo de rotación debe ser el que define el vector \mathbf{T} tangente a la trayectoria.

Para efectuar la rotación eliminando la incompatibilidad de unidades espacial y temporal, conviene hacer el cambio de variable $z = ut$, donde u es una constante con unidades de velocidad (Figura 4.1-b).

La relación entre las variables originales y las variables rotadas queda descrita por

$$\begin{aligned} x' &= x \cos \phi - z \operatorname{sen} \phi & x &= x' \cos \phi + z' \operatorname{sen} \phi \\ z' &= x \operatorname{sen} \phi + z \cos \phi & z &= -x' \operatorname{sen} \phi + z' \cos \phi \end{aligned} \quad (4.1)$$

y la relación entre el ángulo de rotación y la velocidad en el punto de interés está dada por $v_0 = u \tan \phi$.

Para demostrar que esta transformación de coordenadas efectivamente elimina el movimiento en el punto de interés, evaluamos dx'/dz' tomando una trayectoria descrita en forma implícita por $f(x,z) = f(z' \operatorname{sen} \phi + x' \cos \phi, z' \cos \phi - x' \operatorname{sen} \phi) = 0$, y aplicamos la fórmula de la derivada de una función implícita $dx/dz = -(\partial f / \partial z) / (\partial f / \partial x) = \tan \phi$, es decir,

$$\begin{aligned} \frac{dx'}{dz'} &= -\frac{\partial f / \partial z'}{\partial f / \partial x'} \\ &= -\frac{(\partial f / \partial z)(\partial z / \partial z') + (\partial f / \partial x)(\partial x / \partial z')}{(\partial f / \partial z)(\partial z / \partial x') + (\partial f / \partial x)(\partial x / \partial x')} \\ &= -\frac{\operatorname{sen} \phi - \tan \phi \cos \phi}{\cos \phi - \tan \phi \operatorname{sen} \phi} = 0 \end{aligned}$$

Para el caso de imágenes, la trayectoria que describe un punto a medida que se desplaza sobre el plano imagen se puede representar en tres dimensiones (con el correspondiente cambio de variable de la coordenada temporal). Para proceder como en el caso anterior, primero se hace una rotación al rededor del eje z para alinear el eje x con la dirección de movimiento y luego se aplica la transformación (4.1) sobre las coordenadas que resultan de esa rotación.

La Figura 4.2 ilustra la rotación de los ejes. Observe que la posición final del eje z (eje z') es colineal con el vector tangente \mathbf{T} .

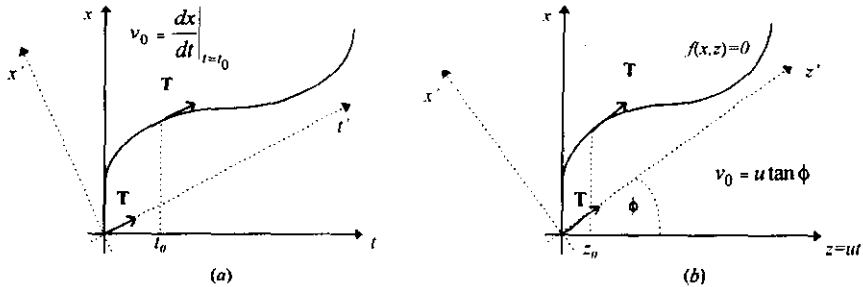


Figura 4.1.- Transformación de coordenadas. (a) trayectoria original, (b) trayectoria en coordenadas espaciales ($z = ut$)

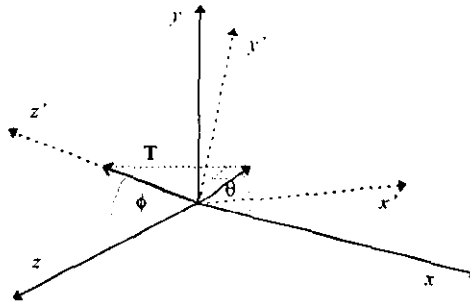


Figura 4.2.- Rotación del sistema tridimensional (el eje z' queda alineado en la dirección del vector T)

Si θ es el ángulo que forma la proyección del vector tangente sobre el plano xy y el eje x , entonces $T = (\cos\theta\text{sen}\phi, \text{sen}\theta\text{sen}\phi, \cos\phi)$ y la transformación completa resulta

$$\begin{aligned} x' &= (x \cos\theta + y \text{sen}\theta)\cos\phi - z \text{sen}\phi \\ y' &= -x \text{sen}\theta + y \cos\theta \\ z' &= (x \cos\theta + y \text{sen}\theta)\text{sen}\phi + z \cos\phi \end{aligned} \tag{4.2}$$

El vector velocidad en el punto de interés se determina entonces a partir de la dirección del vector tangente, es decir, $v = u \tan\phi (\cos\theta, \text{sen}\theta)$.

4.3 Aproximación local a 2-D

Como en el nuevo sistema de referencia el movimiento en el instante de interés es nulo,

podemos plantear el problema de estimación como:

Determinar el patrón bidimensional k y los ángulos (θ, ϕ) para los cuales la imagen $l(x, y, z)$ puede ser mejor ajustada por $k(x', y')$, con x' y y' dados por la transformación (4.2).

Para resolver el problema definimos el mejor ajuste en términos del error cuadrático ponderado²,

$$WSE_k^2(\theta, \phi) = \iiint [k(x', y') - l(x, y, z)]^2 w^2(x, y, z) dx dy dz \quad (4.3)$$

el cual se minimiza sobre todos los patrones k y ángulos (θ, ϕ) . Si el ángulo óptimo $\phi = \pi/2$, entonces el patrón k es puramente temporal, como por ejemplo un cambio brusco de iluminación sobre una superficie uniforme. Si $\phi \neq \pi/2$, entonces k es un patrón que se mueve a una velocidad constante $u \tan \phi$ en la dirección dada por el ángulo θ .

Se define la ventana bidimensional

$$w_{\theta, \phi}^2(x', y') = \int w^2(x, y, z) dz' \quad (4.4)$$

con

$$\begin{aligned} x &= (x' \cos \phi + z' \sin \phi) \cos \theta - y' \sin \theta \\ y &= (x' \cos \phi + z' \sin \phi) \sin \theta + y' \cos \theta \\ z &= -x' \sin \phi + z' \cos \phi \end{aligned}$$

proyectando $w^2(x, y, z)$ sobre el plano definido por su normal \mathbf{T} . Esta ventana bidimensional es independiente de los ángulos (θ, ϕ) si $w(x, y, z)$ tiene la propiedad de simetría rotacional.

Sustituyendo la expansión

$$w_{\theta, \phi}(x', y') \left[k(x', y') - \sum_{j=0}^{\infty} \sum_{i=0}^i k_{(\theta, \phi)_{i, j}} g_{(\theta, \phi)_{i, j-i}}(x', y') \right] = 0 \quad (4.5)$$

de $k(x', y')$ sobre la base ortonormal $\{g_{(\theta, \phi)_{i, j}}(x', y') : i = 0, \dots, j; j = 0, \dots, \infty\}$ de polinomios asociados a esta ventana y la expansión de la imagen $l(x, y, z)$ sobre la base $\{p_{(\theta, \phi)_{m, n, p}}(x', y') : m = 0, \dots, n; n = 0, \dots, p; p = 0, \dots, \infty\}$ de polinomios ortonormales sobre $w^2(x, y, z)$ en (4.3); y

² Para simplificar la notación, sólo se discute el caso con la posición de la ventana en el origen; la generalización a cualquier posición es inmediata.

derivando respecto a $k_{(\theta,\phi),j-i}$ resultan los coeficientes óptimos

$$k_{(\theta,\phi),j-i} = \sum_{p=0}^j \sum_{n=0}^p \sum_{m=0}^n l_{m,n-m,p-n} h_{m,n-m,p-n}^{i,j-i}(\theta, \phi) \quad (4.6)$$

donde

$$h_{m,n-m,p-n}^{i,j-i}(\theta, \phi) = \iiint \mathcal{G}_{(\theta,\phi),j-i}(x', y') \rho_{m,n-m,p-n}(x, y, z) w^2(x, y, z) dx dy dz \quad (4.7)$$

es una *función de ángulo* completamente determinada por $w^2(x, y, z)$ (ver sección 4.3.1).

Tomando en cuenta que los coeficientes polinomiales $l_{m,n-m,p-n}$ pueden ser derivados convolucionando la señal de entrada $l(x, y, z)$ con los filtros de análisis $d_{m,n-m,p-n}(x, y, z) = \rho_{m,n-m,p-n}(-x, -y, -z) w^2(-x, -y, -z)$ y seleccionando la salida en $x = y = z = 0$, podemos escribir (4.6) como

$$k_{(\theta,\phi),j-i} = l(x, y, z) * d_{(\theta,\phi),j-i}(x, y, z) \Big|_{(x,y,z)=(0,0,0)} \quad (4.8)$$

donde

$$d_{(\theta,\phi),j-i}(x, y, z) = \sum_{p=0}^j \sum_{n=0}^p \sum_{m=0}^n d_{m,n-m,p-n}(x, y, z) h_{m,n-m,p-n}^{i,j-i}(\theta, \phi) \quad (4.9)$$

son *filtros direccionables* que analizan la imagen en diferentes orientaciones espaciotemporales. Para los ángulos óptimos, estos filtros extraen las características estáticas de los patrones que generan la imagen dinámica.

Desarrollando el cuadrado en (4.3) y considerando los coeficientes óptimos (4.6), se observa que el error de aproximación

$$WSE^2(\theta, \phi) = \sum_{p=0}^{\infty} \sum_{n=0}^p \sum_{m=0}^n l_{m,n-m,p-n}^2 - \sum_{j=0}^{\infty} \sum_{i=0}^j k_{(\theta,\phi),j-i}^2 \quad (4.10)$$

se minimiza al maximizar la *energía direccional*

$$\sum_{j=0}^{\infty} \sum_{i=0}^j k_{(\theta,\phi),j-i}^2 \quad (4.11)$$

respecto a θ y ϕ , donde los coeficientes polinomiales pueden ser calculados mediante (4.6) o (4.8).

Si la porción de la imagen $l(x,y,z)$ dentro de la ventana es efectivamente generada por el desplazamiento de un patrón 2-D que contiene suficiente información local entonces es posible determinar en forma única los ángulos que maximizan la energía direccional. Bajo estas condiciones el error de aproximación es nulo para la dirección óptima y los coeficientes de la imagen original pueden ser recuperados a partir de los coeficientes óptimos del patrón 2-D mediante

$$l_{m,n-m,p-n} = \sum_{j=0}^{\infty} \sum_{i=0}^j k_{(\theta,\phi),j-i} h_{m,n-m,p-n}^{i,j-i}(\theta, \phi) \quad (4.12)$$

Si la imagen dinámica no satisface las condiciones de movimiento uniforme y constante dentro de la ventana, entonces esta última expresión es todavía válida como el mejor ajuste bidimensional de la secuencia.

4.3.1 Función de ángulo

En la práctica, los polinomios $g_{(\theta,\phi),j-i}(x',y')$ y $\rho_{m,n-m,p-n}(x,y,z)$ se obtienen ortonormalizando $\{((x \cos \theta + y \sin \theta) \cos \phi - z \sin \phi)^i (-x \sin \theta + y \cos \theta)^{j-i} : i = 0, \dots, j; j = 0, 1, \dots\}$ y la base canónica $\{x^m y^{n-m} z^{p-n} : m = 0, \dots, n; n = 0, \dots, p; p = 0, 1, \dots\}$ respectivamente con respecto a la ventana $w^2(x,y,z)$. Así, al definir $w^2(x,y,z)$, las funciones de ángulo quedan fijas. Estas funciones se pueden ver como los coeficientes de la expansión de los polinomios $g_{(\theta,\phi),j-i}(x',y')$ sobre la base de polinomios $\rho_{m,n-m,p-n}(x,y,z)$.

De la propiedad de ortonormalidad de los polinomios $g_{(\theta,\phi),j-i}(x',y')$ se obtienen las siguientes propiedades para las funciones de ángulo

$$\begin{aligned} h_{0,0,0}^{0,0}(\theta, \phi) &= 1, \\ h_{0,0,0}^{i,j-i}(\theta, \phi) &= 0, \text{ para } j > 0, \\ h_{m,n-m,p-n}^{i,j-i}(\theta, \phi) &= 0, \text{ para } p > j, \end{aligned} \quad (4.13)$$

$$\sum_{p=0}^{\infty} \sum_{n=0}^p \sum_{m=0}^n h_{m,n-m,p-n}^{i,j-i}(\theta, \phi) h_{m,n-m,p-n}^{k,l-k}(\theta, \phi) = \delta_{i,k} \delta_{j,l}$$

donde la primera supone una ventana con energía normalizada. De ésta se puede concluir que el coeficiente de DC no interviene en la determinación de la velocidad y que la información de movimiento contenida en este coeficiente sólo se puede analizar a escalas mayores.

En lo sucesivo, a menos que se indique lo contrario, cada vez que se mencione la energía de la señal nos referiremos a la energía de AC, es decir, a la energía contenida en los coeficientes de orden 1, 2, etc.

De la segunda y tercera propiedad se determinan las funciones de ángulo no nulas numeradas en la Tabla 4.1. Las celdas marcadas con × corresponden a funciones de ángulo nulas. En general el número total de funciones de ángulo hasta orden 2 es 33 (de las cuales 32 son las requeridas).

$i,j-i$			0 0	0 1	1 0	0 2	1 1	2 0
$m,n-m,p-n$								
0	0	0	1√	×	×	×	×	×
0	0	1	×	2	4√	7	16	25
0	1	0	×	3√	5√	8	17	26
1	0	0	×	4√	6√	9	18	27
0	0	2	×	×	×	10	19	28√
0	1	1	×	×	×	11	20√	29√
1	0	1	×	×	×	12	21√	30√
0	2	0	×	×	×	13√	22√	31√
1	1	0	×	×	×	14√	23√	32√
2	0	0	×	×	×	15√	24√	33√

Tabla 4.1.- Orden de las funciones de ángulo no nulas para $p = 0,1$ y 2

La elección de una ventana con simetría rotacional reduce el número de funciones de ángulo requeridas. En particular, para una ventana Gaussiana se tienen funciones de ángulo no nulas dadas por

$$h_{m,n-m,p-n}^{i,j-i}(\theta, \phi) = \sum_{k=m}^n \sqrt{C_{n-p+i}^{n-k} C_{p-i}^{k-m}} \alpha_{k-p+i,p-i-k+m}(-\theta) \alpha_{k-m,n-k}(\theta) \alpha_{n-p+i,p-n}(-\phi) \cdot \delta_{j,p}$$

para $i \geq p-n$, con

$$\alpha_{m,n-m}(\theta) = \sqrt{C_n^m} \cos^m(\theta) \text{sen}^{n-m}(\theta)$$

lo que es una importante simplificación sobre el caso general ya que, como puede verse, para calcular los coeficiente 2-D de un orden dado, sólo se requieren los coeficientes 3-D del mismo orden. Estas funciones de ángulo están marcadas con √ en la Tabla 4.1. El número de funciones requeridas en este caso es de 19.

4.3.2 Criterio de detección

Si la imagen no contiene suficiente información local, entonces no es necesario hacer la aproximación ya que se obtendría una distribución de energía direccional uniforme que no proporciona suficiente información para determinar una dirección óptima. Por lo tanto, es necesario definir una medida de variación local que nos permita tomar la decisión de realizar o no la aproximación a 2-D. Esta medida de variación local de la imagen corresponde al *coeficiente de contraste*

$$C = \log \left[1 + \sum_{p=1}^N \sum_{n=0}^p \sum_{m=0}^n t_{m,n-p,p-n}^2 \right]$$

el cual es cero si la imagen puede ser aproximada por su valor medio dentro de la ventana. Si los valores de imagen se desvían de su valor medio, entonces el coeficiente de contraste es una medida logarítmica de la energía de la desviación hasta orden N . Martens (1990) ha hecho notar que este coeficiente de contraste proporciona una buena indicación de la posición y contraste de estructuras visuales. Así, el análisis direccional se lleva a cabo si C es mayor que el mínimo contraste perceptible δ . En la mayoría de los experimentos presentados aquí se consideró un umbral $\delta = \gamma C_{\mu}$, con C_{μ} el promedio del coeficiente de contraste sobre todas las posiciones de análisis y la constante $\gamma \approx 0.01$.

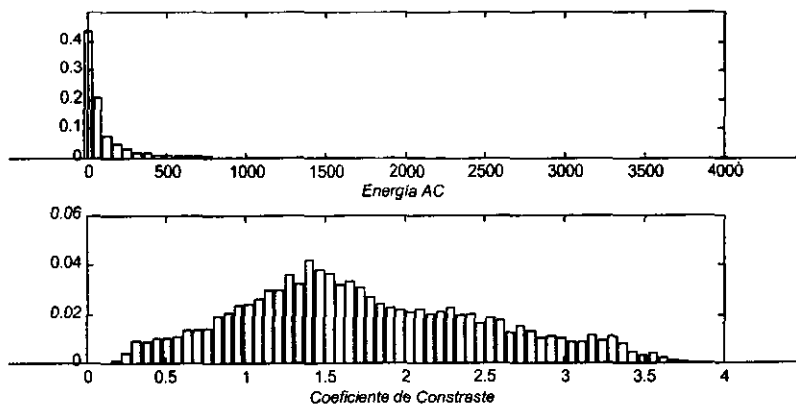


Figura 4.3.-Histogramas de frecuencias de la energía y del coeficiente de contraste típicos

El hecho de usar una medida logarítmica en lugar de la medida original reduce el rango dinámico de la cantidad original facilitando la elección de un umbral empírico para la detección de los patrones 2-D. La Figura 4.3 muestra la forma típica de los histogramas de la energía AC y del coeficiente de contraste de una imagen texturizada. Como puede observarse, el histograma del coeficiente de contraste está *mejor* distribuido sobre todo su rango.

De la misma forma se puede usar el error cuadrático ponderado para definir una medida de la calidad del ajuste local que además puede ser usada para rechazar o aceptar la estimación de la orientación. Esta medida es el coeficientes de contraste del error de aproximación:

$$\Delta C = \log[1 + WSE]$$

el cual se compara contra un umbral ΔC_{thr} para decidir si la aproximación es aceptable o no. Este umbral debe eliminar el error que no es perceptible al ojo. Para un observador, los errores son menos perceptibles cuando son desviaciones de valores grandes que cuando son desviaciones de valores pequeños. Para el coeficiente de contraste introducido arriba se puede usar $\Delta C_{thr} = \delta + \lambda (C - \delta)$ para $\lambda \approx 0.4$ [Martens, 1990b].

4.4 Formulación discreta

La formulación discreta de la técnica de aproximación local a dos dimensiones es totalmente similar al caso continuo. Las expresiones integrales del caso continuo corresponden a sumas finitas sobre un dominio apropiado y el orden máximo de la expansión de una señal discreta se determina por el tamaño de la ventana. El parámetro de escala para el caso discreto está en función de la longitud de la ventana. Aquí debe notarse que lo que establece la diferencias de escala espacial y temporal es el parámetro de velocidad u . En el caso continuo es claro que, una vez definidas las escalas unitarias, este parámetro está dado por el cociente entre la escala espacial y la escala temporal. En el caso discreto, éste corresponde al cociente entre el orden espacial y el orden temporal. Por ejemplo, para una ventana Gaussiana en las tres direcciones la escala espacial está dada por el parámetro σ y la escala temporal por $\beta = \sigma / u$, es decir que $u = \sigma / \beta$. Y para el caso discreto, suponiendo que se toma un orden N en las dimensiones espaciales y M en la dimensión temporal para la expansión de la imagen, resulta el parámetro $u = N/M$. Este parámetro de velocidad establece el rango de velocidades en la cual la transformada polinomial es más sensible.

Para el cálculo de las funciones de ángulo, al considerar el cambio de variable $z = ut$, resulta la ventana del mismo orden N en las tres dimensiones, es decir que se cambia únicamente la longitud temporal. Además, no se calcula la ventana 2-D para cada par de ángulos (θ, ϕ) , sino que, en lugar de ello, se obtienen directamente los polinomios 2-D ortonormalizando la base $\{1, x', \dots, x'^N, y', x' y', \dots, x'^N y', \dots, y'^N, x' y'^N, \dots, x'^N y'^N\}$ sobre la ventana $w^2[x, y, z]$ con x' y y' dados por (4.2) para los ángulos

$$\theta = \arctan(y/x) \quad \text{para } x, y = -N/2, \dots, N/2$$

$$\phi = \arctan(r/z) \quad \text{para } r^2 = x^2 + y^2, z = -N/2, \dots, N/2$$

Para ello puede usarse el proceso de ortonormalización Gram-Smith. Por razones que se explican en el Capítulo VI conviene ortonormalizar los polinomios de esta base siguiendo el orden dado arriba, especialmente cuando la ventana no es isotrópica (con simétrica rotacional). A partir de éstos y los polinomios discretos 3-D se calculan las funciones de ángulo definidas por (4.7).

Nótese que el costo computacional (en especial, el tiempo de máquina) que requiere el cálculo de las funciones de ángulo para una aplicación práctica no debe ser de tanta consideración como lo es el cálculo de los coeficientes de la expansión, ya que éstas funciones necesitan ser calculadas sólo una vez para una ventana específica, de hecho, en un enfoque jerárquico no es necesario calcular las funciones de velocidad en cada nivel de la pirámide si se elige la misma ventana de análisis en todos de los niveles. Por otro lado, dado que las funciones de ángulo son usadas en el cálculo de los coeficientes del patrón estático, conviene reducir el número de éstas con el fin de agilizar el cómputo de la energía direccional. Una forma de agilizar este cálculo es reducir el número coeficientes de la expansión. Simulaciones en computadora muestran que basta tomar los coeficientes hasta orden 2 para describir ciertas primitivas de imagen; sin embargo, las regiones muy texturizadas requieren un orden más grande para ser descritas con una precisión aceptable. La otra alternativa, que es más viable, es considerar una ventana isotrópica como ya lo hemos mostrado.

4.5 Estimación del flujo óptico local

El primer paso para estimar el campo de velocidades corresponde a la obtención los coeficientes de la expansión $l_{m,n,p} : 0 \leq m, n \leq N, 0 \leq p \leq M$. En este paso conviene usar ventanas separables para que los filtros de análisis resultantes también sean separables y el filtrado pueda hacerse sobre las tres dimensiones en forma independiente usando una configuración en cascada. Aquí se debe elegir también el tamaño y forma de la ventana, así como el intervalo entre posiciones consecutivas de la ventana es decir la rejilla de muestreo espacial.

Como no se tiene experiencia de aplicaciones de la transformada polinomial al procesamiento de secuencias de imágenes, la elección de los parámetros de la transformada se pueden hacer tomando en cuenta lo que se conoce de la transformada de imágenes estáticas. Sabemos, por ejemplo, que la elección de la escala espacial mínima se establece en términos del tamaño de las estructuras más pequeñas que tienen un significado perceptible en la imagen. Para el caso de imágenes dinámicas, las simulaciones hechas en computadora muestran que la escala espacial también determina la magnitud de la velocidad máxima que es posible estimar dentro de cierto margen de precisión. Por ejemplo en una ventana de 3×3 píxeles no es factible registrar velocidades mayores a 2 píxeles por trama. La escala temporal, por su parte, establece la sensibilidad mínima de movimiento. Por ejemplo, un desplazamiento

de 1 pixel cada 4 tramas no puede ser detectado usando una ventana de longitud menor a 5 tramas.

La estructura general del procedimiento para estimar flujo óptico dadas la expansión espaciotemporal y las funciones de ángulo asociadas a la ventana de análisis, es la siguiente:

Para cada posición sobre la rejilla de muestreo espacial donde el coeficiente de contraste es mayor que un umbral prefijado,

- *Calcular los coeficientes de la aproximación 2-D. Para los ángulos prefijado se determinan los coeficientes de los patrones 2-D a partir de los coeficientes 3-D y las funciones de ángulo mediante (4.7), o sea*

$$k_{(\theta,\phi),j-i} = l_{0,0,1}h_{0,0,1}^{i,j-i}(\theta,\phi) + \dots + l_{j,0,0}h_{j,0,0}^{i,j-i}(\theta,\phi)$$

para $i = 1, \dots, j$ y $j = 0, \dots, N$.

- *Maximizar la energía direccional 2-D. En este paso se determinan los valores de (θ, ϕ) tales que:*

$$k_{(\theta,\phi),0,1}^2 + k_{(\theta,\phi),1,0}^2 + \dots + k_{(\theta,\phi),N,N}^2$$

es máximo así como el coeficiente de contraste del error de ajuste ΔC .

- *Calcular el vector flujo. Si $\phi \neq \pi/2$ y si $\Delta C < \Delta C_{thr}$ se calculan las componentes del vector de flujo:*

$$v_x = u \tan \phi \cos \theta$$

$$v_y = u \tan \phi \sen \theta$$

El proceso se ilustra en el siguiente diagrama de bloques. La entrada del proceso corresponde a la expansión polinomial. El bloque $H^{(2)}(\theta, \phi)$ mapea los coeficientes 3-D, representados en notación matricial por L_n , a coeficientes 2-D orientados, representados los vectores $k_{(\theta,\phi),n}$.

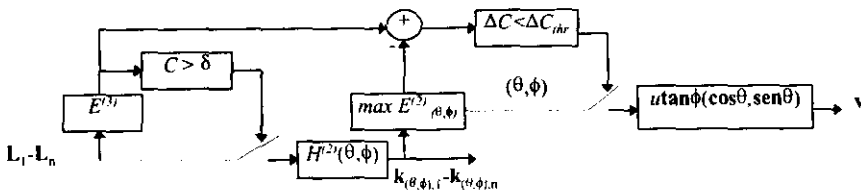


Figura 4.4.- Estimación del flujo local basada en la energía direccional

El bloque etiquetado con $C > \delta$ efectúa la comparación entre el coeficiente de contraste y un umbral prefijado δ para determinar si debe o no hacerse la aproximación, mientras que el bloque etiquetado con $\Delta C < \Delta C_{thr}$ efectúa comparación entre el error del coeficiente de contraste y un umbral prefijado para aceptar o rechazar la orientación óptima. La salida es un vector velocidad para cada posición de la ventana.

4.5.1 Resultados experimentales

La Figura 4.5-a muestra la imagen sintética TEXTURE la cual se animó usando el flujo mostrado en la Figura 4.5-b. Esta secuencia sintética se empleó en el primer experimento para mostrar el impacto que tiene el número de coeficientes en la estimación del flujo. En la Figura 4.5-c y d se muestran dos estimaciones del flujo empleando una ventana binomial de longitud 5 en las tres direcciones. La primera (c) se obtuvo con los coeficientes hasta de orden 2 y la segunda (d) considerando coeficientes hasta de orden 4. A simple vista se observa que la precisión de las estimaciones con un número mayor de coeficientes es mejor que aquella obtenida con pocos coeficiente. Sin embargo, el orden máximo que debe considerarse no sólo depende de la precisión requerida si no también, de las capacidades de procesamiento disponible, ya que el número de coeficientes crece en forma cúbica con el orden máximo y el tiempo de cálculo puede llegar a ser prohibitivo.

El comportamiento del método en una secuencia real se ejemplifica en el siguiente experimento hecho sobre la secuencia INTERVIEW. En la Figura 4.6 se muestran 3 tramas representativas de la porción de la secuencia empleada para este experimento. El campo de velocidades mostrado es una versión muestreada de la estimación usando la misma ventana del caso anterior. Es posible que algunas de las observaciones hechas a continuación no se aprecien muy bien por submuestreo del campo. En general los resultados son coherentes con el movimiento aparente pero podemos observar lo siguiente. Primero, el método es impreciso sobre la parte superior del pelo debido al problema de recubrimiento y al problema de apertura (en algunos puntos sobre el contorno el flujo la estimación es aproximadamente la componente normal, pero en la mayoría de ellos esto no es cierto). Segundo, aunque el movimiento global es hacia arriba, el flujo local no resulta uniforme y tiende a segmentarse sobre la parte de los brazos debido a las deformaciones de la ropa y a la cuantización de los ángulos. Tercero, sobre la parte superior derecha el método falla debido al problema de apertura. En el siguiente capítulo se presentan resultados más detallado sobre este último caso.

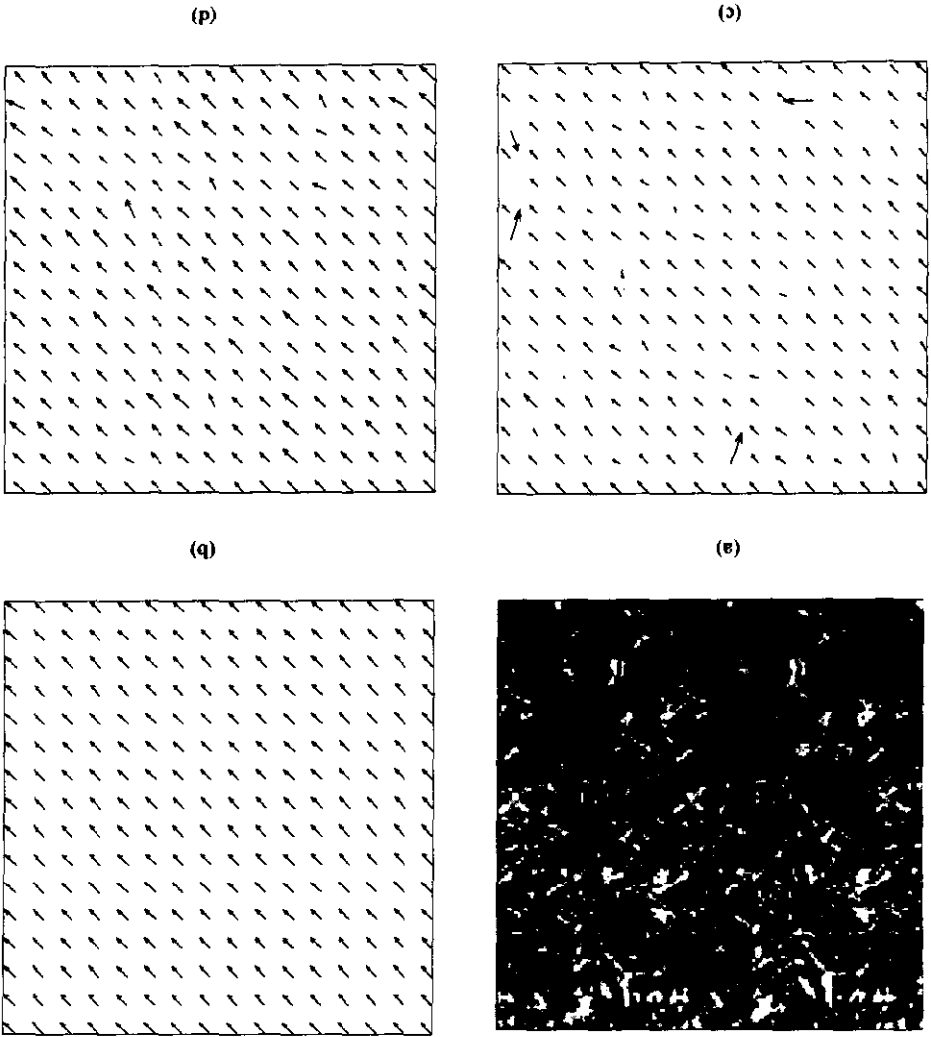


Figura 4.5.- (a) Imagen sintética TEXTURE. (b) Flujo óptico real. (c) Flujo estimado usando coeficientes hasta de orden 2. (d) Flujo estimado usando coeficientes hasta de orden 4.

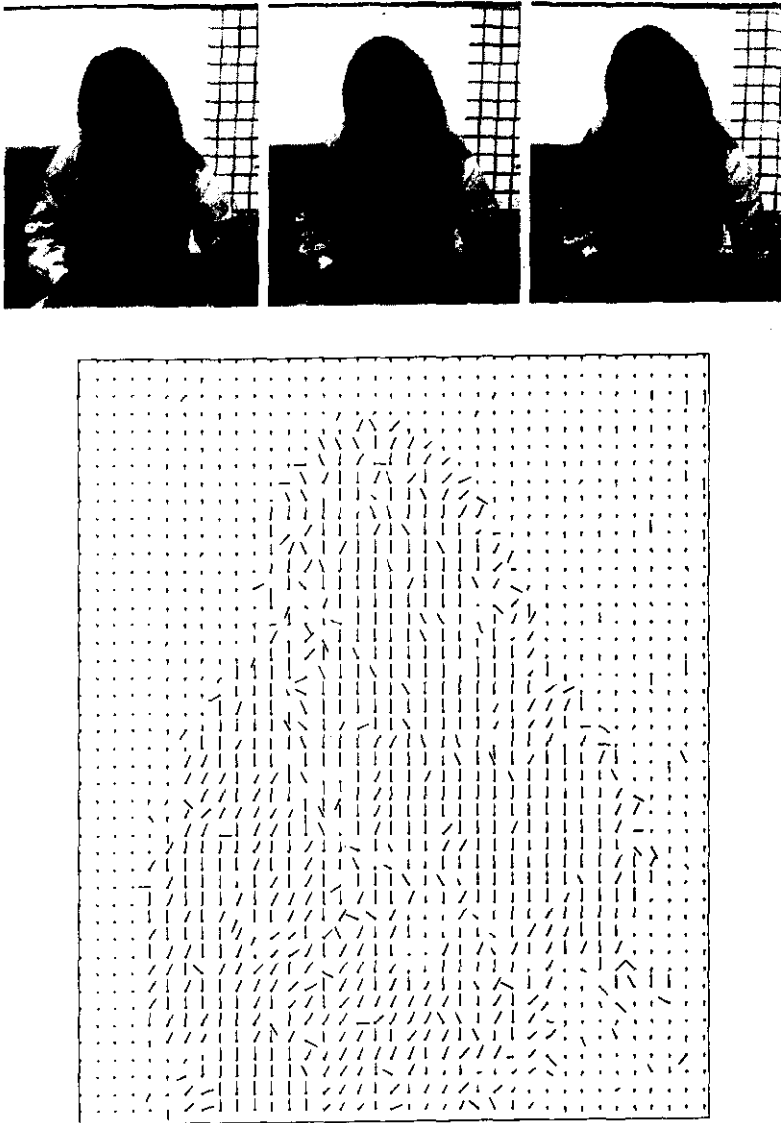


Figura 4.6.- Tramas 1, 4 y 7 de la secuencia INTERVIEW y flujo óptico estimado (muestreado con un factor de 8)

Capítulo 5

Representación 1-D de imágenes dinámicas

5.1 Introducción

Parte fundamental de la interpretación de una escena descansa en la tarea de detección de contornos. Es por ello que en muchas investigaciones se ha puesto especial interés en la determinación del movimiento a lo largo de contornos [Hildreth, 1984]. Los esquemas clásicos para medir el movimiento sobre el contorno atacan principalmente tres problemas: (a) detectar el contorno, (b) medir localmente el movimiento en cada punto sobre el contorno, y (c) reconstruir globalmente el campo de velocidades sobre todo el contorno. Una forma alternativa de proceder es tomar explícitamente en cuenta las dimensiones espaciotemporales del problema y echar mano de filtros espaciotemporales para detectar los contornos en movimiento directamente.

En este capítulo se describe un procedimiento para la extracción del flujo normal de estructuras orientadas (contornos) basado en un modelo unidimensional de la imagen. La formulación es completamente similar a la técnica introducida en el capítulo anterior¹. En la sección 5.4 se discute la interpretación de este enfoque en términos de la ecuación del gradiente. Al final del capítulo se incluyen resultados experimentales del método.

5.2 Problema de apertura

En el capítulo anterior mostramos cómo el movimiento en imágenes puede ser caracterizado por la definición de una orientación espaciotemporal. También concluimos que existen casos en los que no es posible determinar en forma única esa orientación óptima. En el peor de los casos, si la porción de la imagen dentro de la ventana es plana, entonces no existe una dirección óptima. Otro caso de particular interés es ocasionado por el problema de

¹ Una versión del método que no emplea orientaciones explícitamente se reportó en "Motion estimation by means of polynomial transforms", *Memorias de Computación Visual 98. Méx.*, D. F. abril 1998 (Silván-Cárdenas & Escalante-Ramírez).

apertura. La Figura 5.1 muestra la distribución típica de energía direccional en el plano $\theta\phi$ para un patrón orientado. Los puntos más oscuros indican valores altos de energía, mientras que los puntos blancos indican valores pequeños. Como puede observarse, existen muchos valores de θ y ϕ con valores de energía máxima. Lo deseable es que, en vista del problema de apertura, el método sea al menos capaz de recuperar la componente normal de movimiento.

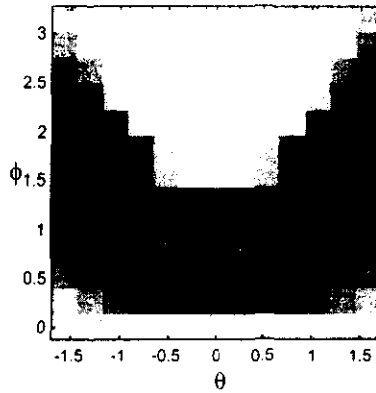


Figura 5.1.- Distribución de la energía direccional de un borde. Los puntos más oscuros corresponden a valores de energía altos y los puntos más claros a valores de energía pequeños.

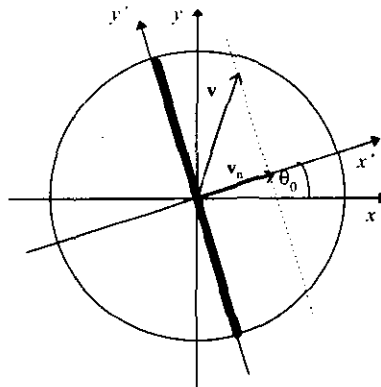


Figura 5.2.- Problema de apertura. Sólo la componente normal del borde puede ser determinada a partir de la información local de la imagen.

Los puntos marcados con + corresponden a posiciones igualmente aceptables para estimar el vector velocidad. Estos puntos describen la trayectoria definida por

$$\phi = \arctan\left(\frac{v_x \cos\theta_0 + v_y \operatorname{sen}\theta_0}{\cos(\theta - \theta_0)}\right) \quad (5.1)$$

para $-\pi/2 < \theta, \phi \leq \pi/2$, de donde se ve que es imposible determinar las componentes de movimiento v_x y v_y . Lo que se puede hacer en lugar de ello es estimar la magnitud de la componente normal que corresponde al numerador del argumento de **arctan**. Esto se logra fijando θ en la dirección normal a la orientación local de la imagen, θ_0 , como se ilustra en la Figura 5.2.

Al alinear el ángulo θ de esta forma, la dirección que sigue la orientación local (eje y') no contiene información relevante de la estructura de la imagen y puede ser eliminada de su representación, de esto se desprende que el patrón que aproxima a la imagen dinámica es ahora unidimensional. De (5.1) la componente normal queda determinada por $v_n = \operatorname{utan}\phi$ ($\cos\theta_0, \operatorname{sen}\theta_0$).

5.3 Aproximación local a 1-D

Usando el criterio del error cuadrático ponderado, ahora minimizamos

$$WSE_k^2(\theta, \phi) = \iiint [k((x \cos\theta + y \operatorname{sen}\theta)\cos\phi - z \operatorname{sen}\phi) - l(x, y, z)]^2 w^2(x, y, z) dx dy dz$$

sobre todos los patrones unidimensionales k y los ángulos (θ, ϕ) . El ángulo θ define la dirección normal a la orientación del patrón 1-D respecto al eje x y el escalar $\phi \neq \pi/2$, la magnitud de la velocidad normal. Para el análisis del patrón 1-D usamos ahora la ventana

$$w_{\theta, \phi}^2(x') = \iint w^2(x', y', z') dy' dz'$$

en términos de $w(x, y, z)$. Esta ventana 1-D es la proyección de la ventana (4.4) sobre el eje que forma un ángulo θ con el eje x . Procediendo como en el capítulo anterior se encuentra la solución óptima

$$k_{(\theta, \phi), i} = \sum_{p=0}^i \sum_{n=0}^p \sum_{m=0}^n l_{m, n-m, p-n} h_{m, n-m, p-n}^{(i)}(\theta, \phi) \quad (5.2)$$

para los coeficientes de la expansión del patrón 1-D sobre las funciones de análisis asociadas a la ventana $w_{\theta, \phi}^2(x')$, donde

$$h_{m,n-m,p-n}^{(i)}(\theta, \phi) = \iiint g_{(\theta, \phi), i}(x') \rho_{m,n-m,p-n}(x, y, z) w^2(x, y, z) dx dy dz \quad (5.3)$$

son *funciones de ángulo* que proyectan los coeficientes 3-D en coeficientes 1-D a lo largo de la dirección especificada y están determinadas en forma única por la ventana $w(x,y,z)$ ya que los momentos de $w^2_{\theta, \phi}(x')$ requeridos para construir los polinomios $g_{\theta, \phi, i}(x')$, para $i = 0, 1, \dots$, pueden basarse en la función $w^2(x,y,z)$, es decir,

$$\begin{aligned} c_{(\theta, \phi), p} &= \iiint [(x \cos \theta + y \sin \theta) \cos \phi - z \sin \phi]^p w^2(x, y, z) dx dy dz \\ &= \sum_{n=0}^p \sum_{m=0}^n \sqrt{C_n^m C_p^n} \alpha_{m,n-m}(\theta) \alpha_{n,p-n}(-\phi) c_{m,n-m,p-n} \end{aligned} \quad (5.4)$$

donde $c_{m,n-m,p-n}$ son los momentos de la ventana 3-D.

La velocidad normal $v_n = u \tan \phi (\cos \theta, \sin \theta)$ se estima para los ángulo θ y ϕ que maximizan la energía direccional de los patrones 1-D

$$\sum_{j=0}^{\infty} k_{(\theta, \phi), j}^2$$

o el coeficiente de contraste direccional

$$C_{(\theta, \phi)} = \log \left(1 + \sum_{j=1}^{\infty} k_{(\theta, \phi), j}^2 \right)$$

Aquí se puede usar el mismo criterio usado en el capítulo anterior para decidir si debe o no proyectarse a 1-D y si la aproximación es aceptable o no.

5.4 Relación con la ecuación de movimiento de Horn-Schunck

Para el caso específico de una ventana Gaussiana en las tres dimensiones encontramos que las funciones de ángulo tienen una expresión bastante sencilla:

$$h_{m,n-m,p-n}^j(\theta, \phi) = \frac{1}{\sqrt{C_p^n C_n^m}} \alpha_{m,n-m}(\theta) \alpha_{n,p-n}(-\phi) \delta_{j,p}$$

Por otro lado, si la estructura local de la imagen está fuertemente orientada, entonces basta considerar los coeficientes de primer orden para determinar el máximo de la energía

direccional. En particular se ha mostrado que bordes y líneas pueden ser descritas completamente usando coeficientes hasta de orden dos. Al maximizar la energía direccional contenida en el coeficiente de primer orden $k_{\theta_0, \phi_0, t}$ sobre los ángulos (θ, ϕ) , encontramos la siguiente expresión para calcular la dirección y magnitud del vector del flujo normal:

$$\theta_0 = \arctan\left(\frac{l_{0,1,0}}{l_{1,0,0}}\right) \tag{5.5}$$

$$v_n = \frac{-ul_{0,0,1}}{l_{1,0,0} \cos\theta_0 + l_{0,1,0} \sin\theta_0}$$

para la ventana con posición en el origen y el parámetro de velocidad $u = \sigma/\beta$. Esta última expresión es válida para cualquier punto (x,y) e instante t . Además, usando la propiedad derivativa de los filtros hermitianos:

$$l_{m,n-m,p-n}(x,y,t) = \frac{1}{\sqrt{2^p (p-n)!(n-m)!m!}} \frac{\partial^p}{\partial(x/\sigma)^m \partial(y/\sigma)^{n-m} \partial(t/\beta)^{p-n}} l_{0,0,0}(x,y,t)$$

y considerando una dirección arbitraria θ en lugar de θ_0 , se puede escribir como

$$\mathbf{v} \cdot \nabla l_{0,0,0} = -\partial_t l_{0,0,0}$$

que es la ecuación de movimiento de Horn-Schunck (2.1) para $l_{0,0,0}$. Este resultado indica que nuestra aproximación usando la energía de primer orden equivale a la ecuación del gradiente aplicada a la versión paso-bajas de la secuencia. También es importante observar que:

1. al aplicar el filtro de orden cero a $l(x,y,t)$ (para obtener $l_{0,0,0}(x,y,t)$) se eliminan las discontinuidades validando así la formulación del gradiente,
2. la diferencia con el método del gradiente es que la ambigüedad debida al problema de apertura se eliminó desde el inicio del planteamiento alineando el ángulo θ en la dirección normal a la orientación local. En la aproximación (5.5) esto equivale a alinear el ángulo θ en la dirección del gradiente espacial de $l_{0,0,0}(x,y,t)$.
3. para escalas muy grandes, los bordes tienden a aplanarse demasiado y se pierde localización de la estructura, lo que implica que los coeficientes de orden superior no deben ser eliminados en ese caso.

5.5 Estimación del flujo normal

En primer lugar, para calcular las funciones de ángulo llevamos a cabo el siguiente procedimiento:

- Calcular los momentos 3-D de la ventana. Se calculan las constantes $c_{m,n,p}$ para $m = 0, \dots, n$; $n = 0, \dots, p$; $p = 0, \dots, 2M$. Si la ventana es separable entonces estos momentos son evaluados en términos de los momentos de funciones 1-D.
- Para cada par de ángulos (θ, ϕ) , tal que $\theta = \arctan(y/x)$ y $\phi = \arctan(r/z)$ con $r^2 = x^2 + y^2$, $-N/2 \leq x, y \leq N/2$ y $0 \leq z \leq N/2$ hacer lo siguiente:
 - calcular los momentos orientados con la expresión (5.4),
 - determinar los coeficientes de los polinomios 1-D evaluando el determinante (3.7)
 - evaluar la expresión (5.3) que corresponde a la definición de las funciones de ángulo.

Una vez calculadas las funciones de ángulo, se lleva a cabo el siguiente procedimiento para determinar los vectores del flujo normal (note la similitud con el método de aproximación bidimensional presentado en el capítulo anterior):

Dados los coeficientes de la expansión y la funciones de ángulo asociadas a la ventana de análisis, para cada posición de la rejilla de muestreo donde el coeficiente de contraste es mayor que un umbral prefijado,

- Calcular los coeficientes del patrón 1-D. Para los ángulos prefijado se determinan los coeficientes de la expansión 1-D a partir de la expansión 3-D y las funciones de ángulo mediante (5.2), o sea, $k_{(\theta, \phi)j} = l_{0,0,1}h_{0,0,1}^j(\theta, \phi) + \dots + l_{j,0,0}h_{j,0,0}^j(\theta, \phi)$ para $j = 0, \dots, N$.
- Maximizar la energía direccional. En este paso se determina el valor de (θ, ϕ) tal que, $k_{(\theta, \phi)1}^2 + k_{(\theta, \phi)2}^2 + \dots + k_{(\theta, \phi)N}^2$ es máximo, así como el error de aproximación ΔC .
- Calcular el vector normal. Si $\phi \neq \pi/2$ y si $\Delta C < \Delta C_{thr}$ se calculan las componentes del vector normal:

$$v_{n,x} = u \tan \phi \cos \theta$$

$$v_{n,y} = u \tan \phi \sen \theta$$

Este método se esquematiza en el siguiente diagrama, el cual es similar al de la Figura 4.4 del modelo bidimensional. El bloque etiquetado con $H^{(1)}(\theta, \phi)$ proyecta los coeficientes L_n a los coeficientes $k_{(\theta, \phi),n}$. Como en el caso de la Figura 4.4, los bloques $C > \delta$ y $\Delta C > \Delta C_{thr}$ determinan si es necesario hacer la proyección y si la dirección óptima es aceptable o no, respectivamente.

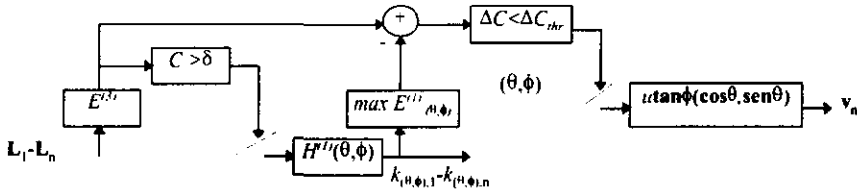


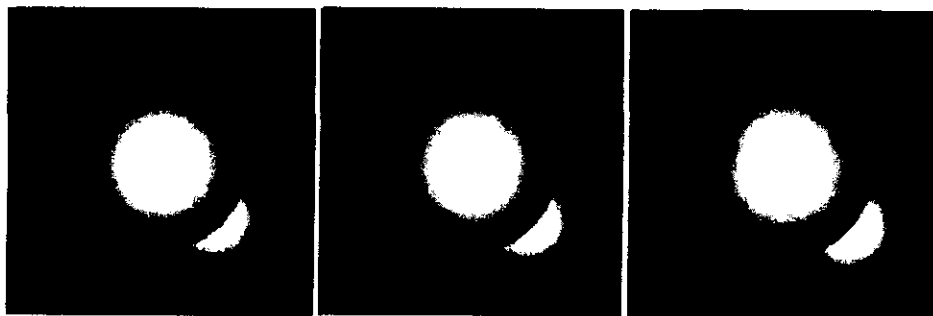
Figura 5.3.- Estimación del flujo normal basada en la energía direccional

5.5.1 Resultados experimentales

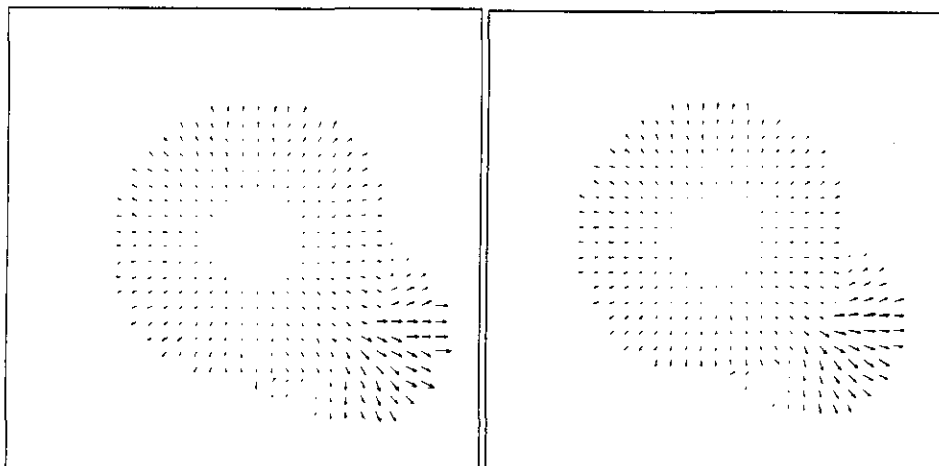
La Figura 5.4-(a) muestra 3 tramas de la secuencia SPHERE usada en el primer experimento. La escena consta de dos esferas sobre un fondo al infinito. La esfera de enfrente se expande con velocidad constante, y la esfera del fondo experimenta un desplazamiento hacia abajo y a la derecha. En las Figuras 5.4 (a) y (b) se muestran las estimaciones del flujo normal que se obtienen usando los coeficientes hasta de orden 4 y usando la aproximación de primer orden para el caso hermitiano. Los resultados en este caso son bastante similares como era de esperarse, pero no así la velocidad de cómputo la cual es mucho menor usando la aproximación de primer orden.

El segundo experimento mostrado en la Figura 5.5 compara los modelos 1-D y 2-D para un caso específico. La imagen mostrada en (a) es la porción de 64×64 píxeles de la parte superior derecha de la secuencia INTERVIEW. El movimiento de esta región se debe exclusivamente al movimiento de la cámara la cual se desplaza con velocidad casi constante con dirección hacia arriba, es decir, que la rejilla permanece aproximadamente estática respecto al fondo. En la Figura 5.5 (b) y (c) se muestran las estimaciones correspondientes a los modelos 2-D y 1-D. En ambos casos se utilizó el método de detección introducido en el capítulo anterior con $\gamma = 0.01$ y $\lambda = 0.8$ (en las posiciones no estimadas se pusieron ceros), y coeficientes hasta de orden 4 para la misma ventana del experimento anterior. Se observa que las estimaciones son aproximadamente complementarias, es decir, el método 2-D no proporciona estimación sobre la región que en algún instante (dentro de los 5 cuadros considerados para la estimación) queda cubierta por las líneas horizontales y sobre las partes planas del fondo; mientras que la aproximación 1-D recupera el flujo normal sobre la región de las barras horizontales. En este caso el flujo normal corresponde justamente al flujo real.

Por lo tanto, un modelo más completo deberá incluir los modelos de aproximación 1-D y 2-D en una forma conveniente. En el siguiente capítulo se discute la integración de estos dos modelos para estimar el flujo real en regiones texturizadas y el flujo normal a lo largo de contornos.



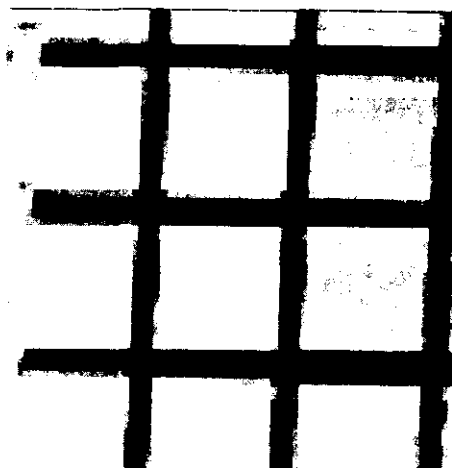
(a)



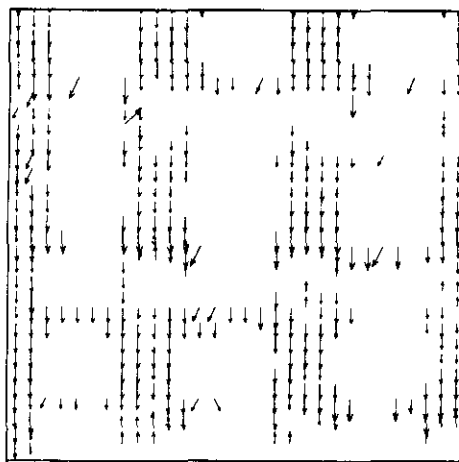
(b)

(c)

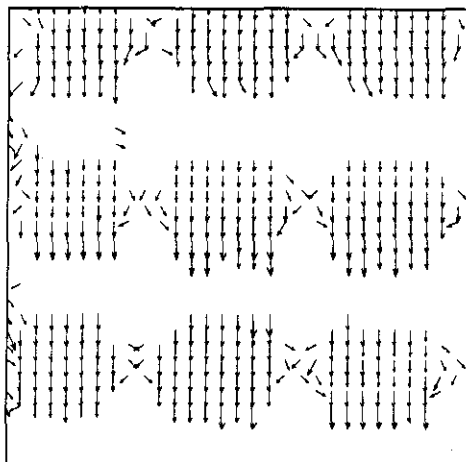
Figura 5.4.- (a)Tramas 1, 4 y 7 de la secuencia SPHERE. (b) Flujo normal calculado con coeficientes hasta de orden 4. (c) Flujo normal calculado con la aproximación de primer orden de la ec. (5.5).



(a)



(b)



(c)

Figura 5.5.- (a) Fragmento de la secuencia INTERVIEW. (b) Estimación usando el modelo 2-D. (c) Estimación usando el modelo 1-D. En ambos casos $\lambda = 0.8, \gamma = 0.01$

Capítulo 6

Integración

6.1 Introducción

Como se mostró en el capítulo anterior las estimaciones de los modelos 2-D y 1-D son, en cierta forma, complementarias. La primera es pobre sobre los contornos, pero bastante coherente sobre las regiones suficientemente texturizadas. La segunda, en cambio, determina la componente normal del flujo a lo largo de los contornos, pero es una estimación pobre sobre las regiones con patrones no orientados.

El objetivo de este capítulo es presentar un algoritmo que integra convenientemente ambas representaciones, así como un procedimiento de propagación de velocidades que, bajo ciertas condiciones, puede preservar las discontinuidades del campo. En primer lugar, se deducen algunas relaciones entre las representaciones estudiadas en los dos capítulos anteriores. Después, en la sección 6.3, se introduce el método para suavizar el campo preservando las fronteras de oclusión que son fáciles de detectar. La propuesta de propagación de velocidades discutida aquí está motivada por resultados expuestos en [Singh, 1991], los cuales fueron ajustados al esquema de la transformada polinomial. Aunque no se presentan los detalles técnicos, se proporciona una deducción cualitativa del esquema, cuya eficacia se puede comprobar en los resultados incluidos al final del capítulo. En la sección 6.4 se discuten aspectos de la inserción del enfoque de estimación en un esquema jerárquico. Al final se proporciona la descripción del algoritmo que integra los modelos 1-D y 2-D, así como la técnica de propagación de velocidades basado en la transformada polinomial de las componentes del flujo. También se reportan experimentos hechos sobre secuencias sintéticas y reales.

6.2 Relación entre los modelos 2-D y 1-D

En el Capítulo IV se estableció el orden en que la base canónica 2-D debe ser ortonormalizada. La razón de usar ese orden de ortonormalización es que la base de

polinomios ortonormales 1-D queda contenida dentro de la base de polinomios 2-D, es decir que

$$\iint g_{(\theta,\phi),k}(x')g_{(\theta,\phi),i,j-i}(x',y')^2g_{\theta,\phi}^2(x',y')dx'dy' = \delta_{i,j}\delta_{j,k}$$

Por lo que es válido escribir $g_{(\theta,\phi),k}(x') = g_{(\theta,\phi),i,0}(x',y')$ para los polinomios 1-D y 2-D y

$$h_{m,n-m,p-n}^i(\theta,\phi) = h_{m,n-m,p-n}^{i,0}(\theta,\phi) \tag{6.1}$$

para las funciones de ángulo.

Podemos decir entonces que el modelo 1-D queda incluido dentro del modelo 2-D, ya que el conjunto de coeficientes del patrón 1-D es subconjunto del conjunto de coeficientes del patrón 2-D. La energía direccional del patrón 2-D se puede escribir entonces como

$$\begin{aligned} E^{(2)}(\theta,\phi) &= \sum_{j=1}^{\infty} k_{(\theta,\phi),j,0}^2 + \sum_{j=1}^{\infty} \sum_{i=0}^{j-1} k_{(\theta,\phi),j,j-i}^2 \\ &= E^{(1)}(\theta,\phi) + \sum_{j=1}^{\infty} \sum_{i=0}^{j-1} k_{(\theta,\phi),j,j-i}^2 \end{aligned}$$

donde $E^{(1)}(\theta,\phi)$ es la energía direccional del patrón 1-D. La energía complementaria correspondiente al último término la denotaremos por $E^{(2-1)}(\theta,\phi) = E^{(2)}(\theta,\phi) - E^{(1)}(\theta,\phi)$. De ésta forma, se observa que los resultados de la aproximación 1-D se pueden aprovechar para hacer la aproximación 2-D requiriendo calcular el conjunto de funciones de ángulo para la proyección 2-D una sola vez. En este punto se usa el error de aproximación 1-D mínimo $WSE^{(1)}$ (o su coeficiente de contraste) para decidir si debe hacerse la aproximación 2-D. En un sistema práctico debemos decidir también si el error mínimo $WSE^{(2)}$ es aceptable o debe tomarse alguna acción adicional. Que $WSE^{(2)}$ sea muy grande significa que hay mucha información local prediciendo un movimiento complejo, la existencia de una frontera de oclusión, o se experimentan cambios en la iluminación (o una combinación). En el mejor de los casos, se puede estimar la tendencia de dicho movimiento complejo de la siguiente manera.

Defina una distribución $R(\theta,\phi)$ del error de aproximación WSE para cada dirección de proyección definida por (θ,ϕ) , entonces si $R(\theta,\phi)$ es insesgada, la mejor estimación es

$$\begin{aligned}
 u_x &= \int_{-\pi/2}^{\pi/2} \int_{-\pi/2}^{\pi/2} R(\theta, \phi) \mu \tan \phi \cos \theta d\theta d\phi \\
 u_y &= \int_{-\pi/2}^{\pi/2} \int_{-\pi/2}^{\pi/2} R(\theta, \phi) \mu \tan \phi \operatorname{sen} \theta d\theta d\phi
 \end{aligned}
 \tag{6.2}$$

En los experimentos incluidos en esta tesis se eligió

$$R(\theta, \phi) = R \exp \left\{ -\frac{WSE(\theta, \phi)}{\varepsilon E_L} \right\}$$

donde R es la constante de normalización, E_L es la energía de $l(x,y,z)$ dentro de la ventana de análisis y $0 < \varepsilon < 1$ es constante. Es difícil justificar la elección de esta distribución, pero una de las razones es que la exponencial con argumento negativo es muy bien comportada al rededor de cero.

6.3 Suavizado del campo

El objetivo del segundo paso en la recuperación del flujo óptico consiste en propagar las velocidades usando la información de los vecinos imponiendo una restricción de *suavidad* en el campo. Esta restricción establece que el vector velocidad de un punto dado no debe ser muy diferente de los vectores de las posiciones vecinas. Sin embargo, esto no siempre es cierto, especialmente sobre las fronteras de oclusión en vista del problema de recubrimiento.

En el marco la transformada polinomial el suavizado del campo puede ser formulado como la minimización de la energía de orden superior del flujo, sujeta a la energía direccional máxima. La idea es que la expansión local de las componentes del vector de velocidad en regiones donde el movimiento es homogéneo puede ser perfectamente representadas por el coeficiente de orden cero, ya que la energía contenida en los coeficientes de orden superior es nula o casi nula. Sin embargo, en los puntos cercanos a la oclusión, el campo es discontinuo y la energía de AC es proporcional a la magnitud del cambio de las componentes del flujo en el punto de discontinuidad. En este caso la descripción del flujo dentro de la ventana debe tomar en cuenta también los coeficientes de orden superior. Afortunadamente, para una descripción local del campo basta considerar coeficientes no mayor a los de segundo orden ya que el movimiento en imágenes que puede tener una interpretación en el mundo físico puede ser modelado por un polinomio de segundo grado en x y y [Verri & Poggio, 1989]. Además si se piensa en la frontera de oclusión como un *borde vectorial* (esto es, un cambio abrupto en las componentes del flujo), entonces este puede ser modelado eficientemente por los coeficientes de primero y segundo orden de la expansión polinomial [Martens, 1990b]. Simulaciones hechas en computadora muestran que la eliminación de los coeficientes de segundo orden no produce un impacto significativo en la preservación de las fronteras de oclusión y en cambio,

(con su eliminación) se elimina una mayor cantidad de *artefactos* introducidos por la estimación inicial, con la conveniencia de requerir menos tiempo de cálculo.

No es difícil mostrar que la minimización de la energía de AC menos la energía de primer orden se logra minimizando

$$\iint (\mathbf{v} - \mathbf{v}_{0,0})^T \mathbf{E}_1^{-1} (\mathbf{v} - \mathbf{v}_{0,0}) w^2(x, y) dx dy \quad (6.3)$$

donde

$$\mathbf{E}_1 = \begin{bmatrix} v_{x1,0}^2 + v_{x0,1}^2 & v_{x1,0} v_{y1,0} + v_{x0,1} v_{y0,1} \\ v_{x1,0} v_{y1,0} + v_{x0,1} v_{y0,1} & v_{y1,0}^2 + v_{y0,1}^2 \end{bmatrix}$$

es la *matriz de energía de primer orden* y $v_{x_{m,n-n}}$ y $v_{y_{m,n-n}}$ denotan los coeficiente de orden n de la expansión (espacial) de las componentes del vector velocidad $\mathbf{v} = [v_x, v_y]^T$. La función $w(x, y)$ es la ventana elegida para el análisis del campo. Si no se tiene una mejor opción, se puede elegir la misma ventana (espacial) usada en el análisis de la distribución de intensidades.

Por otra parte, a la estimación obtenida a partir del patrón local de la imagen, digamos $\mathbf{u} = [u_x, u_y]^T$, se le puede asociar una incertidumbre para determinar *en cuánto* debe ser corregida por la información de sus vecinos. Por ejemplo, se puede usar la matriz de covarianza

$$\mathbf{U} = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}$$

con

$$u_{11} = \int_{-\pi/2}^{\pi/2} \int_{-\pi/2}^{\pi/2} (u \tan \phi \cos \theta - u_x)^2 R(\theta, \phi) d\theta d\phi$$

$$u_{22} = \int_{-\pi/2}^{\pi/2} \int_{-\pi/2}^{\pi/2} (u \tan \phi \sen \theta - u_y)^2 R(\theta, \phi) d\theta d\phi$$

$$u_{21} = u_{12} = \int_{-\pi/2}^{\pi/2} \int_{-\pi/2}^{\pi/2} (u \tan \phi \cos \theta - u_x) (u \tan \phi \sen \theta - u_y) R(\theta, \phi) d\theta d\phi$$

como medida de incertidumbre. La estimación \mathbf{u} es más precisa cuando \mathbf{U} toma valores pequeños y es óptima si

$$\iint (\mathbf{v} - \mathbf{u})^T \mathbf{U}^{-1} (\mathbf{v} - \mathbf{u}) w^2(x, y) dx dy \quad (6.4)$$

es mínima. La minimización simultánea de (6.3) y (6.4) conduce a [Singh, 1991]

$$\mathbf{E}_1^{-1} (\mathbf{v} - \mathbf{v}_{0,0}) + \mathbf{U}^{-1} (\mathbf{v} - \mathbf{u}) = \mathbf{0} \quad (6.5)$$

de donde \mathbf{u} y \mathbf{U} son conocidos a partir de la información local de la secuencia y \mathbf{v} , \mathbf{E}_1 y $\mathbf{v}_{0,0}$ deben ser determinados. Esta expresión representa un sistema no lineal en los valores de \mathbf{v} en la vecindad definida por la ventana, el cual puede ser resuelto en forma aproximada usando relajación Gauss-Seidel. La solución se expresa como

$$\begin{aligned} \mathbf{v}^0 &= \mathbf{u} \\ \mathbf{v}^{k+1} &= (\mathbf{E}_1^{-1} + \mathbf{U}^{-1})^{-1} (\mathbf{E}_1^{-1} \mathbf{v}_{0,0}^k + \mathbf{U}^{-1} \mathbf{u}) \end{aligned}$$

Una interpretación de esta relación recurrente se da en términos de la minimización del error cuadrático medio suponiendo que $\mathbf{v}_{0,0}$ y \mathbf{u} son estimadores insesgados de \mathbf{v} [Singh, 1991]. La diferencia de mi propuesta con la propuesta de Singh es que aquí utilizo la matriz de energía de primer orden en lugar de una matriz de correlación (NOTA: la matriz de correlación usada por Singh equivale a la matriz de energía de AC), motivado principalmente por la observación de que la frontera de oclusión corresponde a estructuras orientadas las cuales son mejor detectadas con la energía de primer orden porque es menos sensible al ruido que las energías de orden mayor.

6.4 Esquema jerárquico

En el Capítulo V señalamos que el coeficiente de DC de la señal $I(x,y,z)$ no interviene en el proceso de estimación del flujo óptico. Sin embargo, en el esquema jerárquico este coeficiente es mapeado a coeficientes de AC en los niveles más altos (escalas mayores). Por lo tanto, para poder analizar la información de movimiento contenida en este coeficiente, es necesario obtener la aproximación local en todos los niveles de la jerarquía aprovechando la característica de que, dentro del esquema jerárquico, los coeficientes en los niveles más bajo (escala menor o amplitud de ventana pequeña) son calculados a partir de los coeficientes de los niveles más altos (escala mayor o amplitud de ventana grande). De esta manera, las aproximaciones 2-D y 1-D pueden predecirse y refinarse en los niveles bajos a partir de la aproximación en los niveles más altos con una reducción significativa de cómputo. Además, las escalas asociadas a los diferentes niveles de la jerarquía se eligen de tal forma que la operación de mapeo de los coeficientes entre capas consecutivas se mantiene para toda la pirámide [Martens, 1990b].

Note que, bajo este esquema, no sólo se puede variar la escala de análisis σ (la cual debe estar ligada a la amplitud de la ventana), sino también el parámetro de velocidad u . Al variar este parámetro la escala temporal cambia respecto a la escala espacial. La variación de estos dos parámetros (σ, u) permite variar no solo el rango velocidades a la que la transformada polinomial es más sensible si no también el rango de las posibles estructuras visuales que pueden ser detectadas dentro de la ventana.

La elección de valores apropiados para estos parámetros se puede hacer tomando como referencia lo que se conoce sobre el sistema visual. Existen indicios de que el sistema visual consta de dos subsistema, uno sensible a frecuencias espaciales altas y frecuencias temporales bajas, y el otro sensible a frecuencias espaciales bajas y frecuencia temporales altas [Shulman & Mulvanny, 1983]. Esto puede ser simulado usando dos transformadas Hermite-Laguerre, una con u pequeña y $\sigma (= \beta u)$ grande y la otra con u grande y σ pequeña. Desafortunadamente, aún no se tiene claro cómo se lleva a cabo la integración de las mediciones tomadas en los diferentes subsistemas (o capas de la pirámide en el esquema jerárquico). La forma más simple de refinamiento es proyectar (mediante interpolación de orden cero o de primer orden) la estimación obtenida a una escala grande sobre una capa de menor escala y refinar siguiendo esencialmente un mismo proceso de estimación. Un mayor esfuerzo deberá ser dedicado a esta área de investigación.

Un estudio detallado de las implicaciones del esquema jerárquico de la transformada polinomial en el método de estimación propuesto aquí queda fuera del alcance de este trabajo.

6.5 Descripción del método

Para calcular el flujo real sobre regiones texturizadas y el flujo normal a lo largo de contornos, se compara el coeficiente de contraste del error de la aproximación 1-D contra un umbral prefijado para decidir si debe usarse la aproximación 2-D o si se ha detectado un patrón orientado. El mismo juicio se aplica a la aproximación 2-D y si el error es muy grande se calcula el valor esperado (6.2) como aproximación a la velocidad promedio dentro de la ventana.

La Figura 6.1 muestra la integración de los modelos 1-D y 2-D. Aunque en el diagrama se usa ΔC_{thr} en las dos aproximaciones, en la práctica el umbral ΔC_{thr} para la aproximación 1-D puede ser diferente para la aproximación 2-D.

El procedimiento puede resumirse en los siguientes pasos:

Dados los coeficientes de la expansión y las funciones de ángulo asociadas a la ventana de análisis, para cada posición de la rejilla de muestreo donde el coeficiente de contraste es mayor que un umbral prefijado,

⇒ Calcular los coeficientes del patrón 1-D. Se determinan los coeficientes de la expansión 1-D $k_{(\theta,\phi),j,0} = l_{0,0,1}h_{0,0,1}^{j,0}(\theta,\phi) + \dots + l_{j,0,0}h_{j,0,0}^{j,0}(\theta,\phi)$ para $j = 0, \dots, N$.

⇒ Maximizar la energía direccional $E^{(1)}$. En este paso se determina el valor de (θ,ϕ) tal que, $E^{(1)}(\theta,\phi) = k_{(\theta,\phi),1,0}^2 + k_{(\theta,\phi),2,0}^2 + \dots + k_{(\theta,\phi),N,0}^2$ es máximo, así como el error de aproximación $WSE^{(1)}$.

⇒ Si $\Delta C^{(1)} \leq \Delta C_{thr}$: (aceptar la aproximación 1-D)

◊ Calcular el vector normal. Si $\phi \neq \pi/2$, se calculan las componentes del vector normal:

$$v_{n_x} = u \tan \phi \cos \theta$$

$$v_{n_y} = u \tan \phi \sin \theta$$

◊ Fin del proceso

⇒ Si $\Delta C^{(1)} > \Delta C_{thr}$, (hacer la aproximación 2-D):

◊ Calcular los coeficientes complementarios para la expansión 2-D. Se determinan los coeficientes $k_{(\theta,\phi),j-i} = l_{0,0,1}h_{0,0,1}^{i,j-i}(\theta,\phi) + \dots + l_{j,0,0}h_{j,0,0}^{i,j-i}(\theta,\phi)$ para $i = 0, \dots, j-1$ y $j = 1, \dots, N$.

◊ Maximizar la energía direccional $E^{(2)}$. En este paso se determina el valor de (θ,ϕ) tal que,

$$E^{(2)}(\theta,\phi) = E^{(1)}(\theta,\phi) + k_{(\theta,\phi),0,1}^2 + k_{(\theta,\phi),1,1}^2 + \dots + k_{(\theta,\phi),N,N}^2$$

es máxima, así como el error de aproximación $WSE^{(2)}$

◊ Si $\Delta C^{(2)} \leq \Delta C_{thr}$ (se ha detectado un patrón 2-D):

⇒ Calcular el vector de flujo. Si $\phi \neq \pi/2$, se calculan las componentes del vector de flujo:

$$v_x = u \tan \phi \cos \theta$$

$$v_y = u \tan \phi \sin \theta$$

⇒ Fin del proceso

◊ Si $\Delta C^{(2)} > \Delta C_{thr}$ (El patrón es 3-D):

⇒ Calcular la velocidad promedio u_x y u_y , mediante (6.2)

◊ Fin del proceso

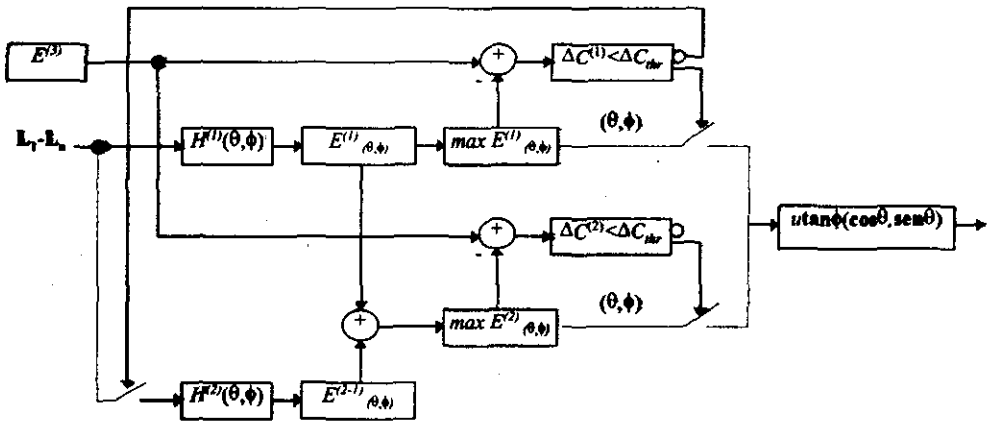


Figura 6.1.-Algoritmo de estimación de movimiento basado en las aproximaciones 1-D y 2-D

Para insertar la etapa de propagación de velocidades en el método anterior sólo hay que incluir el cálculo de la matriz de covarianza U la cual debe usar el error de la proyección 2-D (si se quiere intentar corregir el problema de apertura). Y después llevar a cabo el siguiente método (donde u denota la estimación proporcionada por el método anterior):

Dada la estimación u y la incertidumbre U asociada, para cada punto

- ⇒ Fijar una tolerancia e inicializar $v_0 = u$ y $k = 0$
- ⇒ Calcular los coeficiente $v_{0,0}^k, v_{1,0}^k, v_{0,1}^k$, de la expansión de v^k
- ⇒ Calcular la matriz de energía de primer orden E_1^k
- ⇒ Actualizar la estimación

$$v^{k+1} = (E_1^{-1} + U^{-1})^{-1} (E_1^{-1} v_{0,0}^k + U^{-1} u)$$

- ⇒ Si los cambios entre las componentes de v_{k+1} y v_k son menor que la tolerancia fijada terminar el proceso, de otra forma regresar al segundo paso incrementando k .

El método anterior asume que las matrices E_1 y U son no singulares, sin embargo eso no siempre es cierto. Una forma de resolver el problema de la precisión numérica cuando se tienen matrices singulares (o casi singulares) es mediante la diagonalización ortogonal. La diagonalización de una matriz simétrica A consiste a grandes rasgos en determinar la matriz

ortogonal \mathbf{Q} (formada por los vectores característicos ortonormales de \mathbf{A}) y la matriz de diagonal \mathbf{D} (formada por los valores característicos) y expresar la matriz como $\mathbf{A} = \mathbf{Q}\mathbf{D}\mathbf{Q}^T$, de donde $\mathbf{A}^{-1} = \mathbf{Q}^T\mathbf{D}^{-1}\mathbf{Q}$. La inversa de \mathbf{A} requiere entonces que sus valores característicos sean diferentes de cero (o no menor que un valor mínimo). Si son cero se sustituyen por un valor mínimo (generalmente 0.1, 0.01, etc.).

6.5.1 Resultados experimentales

En todos los experimentos descritos a continuación (salvo en aquellos donde se indica) se usó una ventana binomial de longitud cinco en las tres dimensiones tanto para la secuencia como para el flujo (en todos los casos se empleó una sola escala). Para la propagación de velocidades se iteró hasta que el segundo dígito decimal de las componentes de todo el campo se mantuvo igual entre dos iteraciones consecutivas. Por claridad se muestran versiones muestreadas del campo obtenido y en algunas gráficas se ponen puntos en la parte inicial de los vectores en lugar de flechas en la parte terminal.

El primer experimento ilustra el comportamiento de la técnica sobre la frontera de oclusión con diferentes velocidades. En este experimento usamos la imagen TEXTURE de la Figura 6.2-(a) con el flujo mostrado en (b). Los resultados mostrados en (c) y (d) corresponden a la estimación inicial y la estimación después de propagar las velocidades. La estimación inicial del flujo es bastante precisa sobre las regiones del fondo y el cuadrado salvo por unos puntos sobre la parte del fondo donde se detectaron patrones orientados. La estimación sobre la región del rectángulo muestra *artefactos* (puntos aislados cuyas velocidades quedan fuera de contexto) y en general es menos precisa ya que la diferencia ($2-1 = 1$) entre magnitud de la velocidad y la constante de velocidad u es mayor que en el caso del cuadro (que es $1.4142-1 = 0.4142$) y el fondo (que es $1-1 = 0$), es decir que la magnitud se aleja del rango de velocidades al que la transformada es más sensible ($u = 1$). La estimación sobre la frontera de oclusión es deficiente con un error de aproximación por encima del error promedio. La estimación en algunos puntos se aproxima al promedio de las velocidades de oclusión. La propagación de las velocidades elimina perfectamente los artefactos y preserva la discontinuidad del campo (Figura 6.2-(d)). La discontinuidad sobre las porciones rectas es mejor reconstruida que sobre la esquina del rectángulo (principalmente por la eliminación de los coeficientes de segundo orden). La frontera de arriba del cuadro y la frontera de abajo del rectángulo no se reconstruyen muy bien porque se encuentran muy cerca del borde de la imagen.

El segundo experimento con la secuencia SPHERES muestra como el método de propagación de velocidades puede corregir hasta cierto punto el problema de apertura. La estimación inicial corresponde a la estimación de la Figura 6.3-(a). El flujo obtenido después de 34 iteraciones se muestra en Figura 6.3. Se observa que la estimación inicial sobre la esfera de enfrente se mantiene, mientras que la dirección del flujo sobre la esfera de atrás se aproxima mejor a la real.

Los siguientes experimentos ilustran el comportamiento de la técnica en secuencias reales. Las secuencias INTERVIEW, MISSAME y CALTRAIN son ampliamente conocidas en la

comunidad científica dedicada al procesamiento de secuencias de imágenes. Aquí se discuten brevemente los resultados de la estimación del flujo de la secuencia INTERVIEW (algo similar se puede decir para los otros dos casos). Al aplicar aproximación local se obtiene el flujo de la Figura 6.4. Sobre la región de la cabeza el método distingue tres regiones: las dos más visibles corresponden al contorno en movimiento (el cual se ajusta perfectamente a 1-D) y a la parte del fondo y la cabeza (los cuales se ajustan perfectamente a 2-D); la tercera es una franja delgada que corresponde a la frontera de oclusión localizada entre el contorno en movimiento y el fondo. Es claro que el método recupera inicialmente el flujo normal sobre el contorno, proporciona el flujo local sobre regiones texturizadas (el fondo y el cuerpo) e intenta encontrar la velocidad promedio sobre la frontera de oclusión. En esta última estimación se introduce, sin embargo, una mayor cantidad de artefactos debido principalmente a la suposición de que la distribución $R(\theta, \phi)$ es insesgada lo cual no es cierto en general. La estimación del flujo global se obtiene después de propagar las velocidades (Figura 6.5) y al mismo tiempo se preservan las fronteras de oclusión y se resuelve (aproximadamente) el problema de apertura.

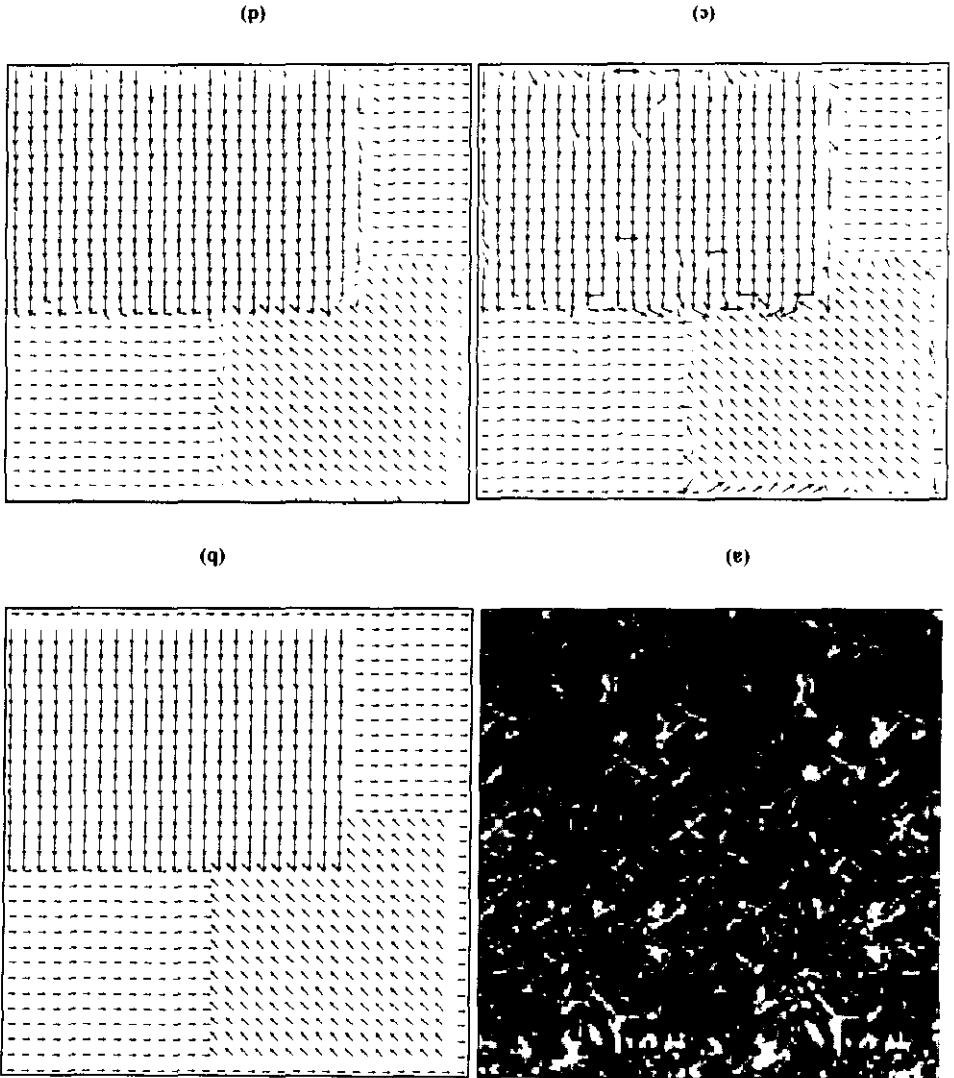
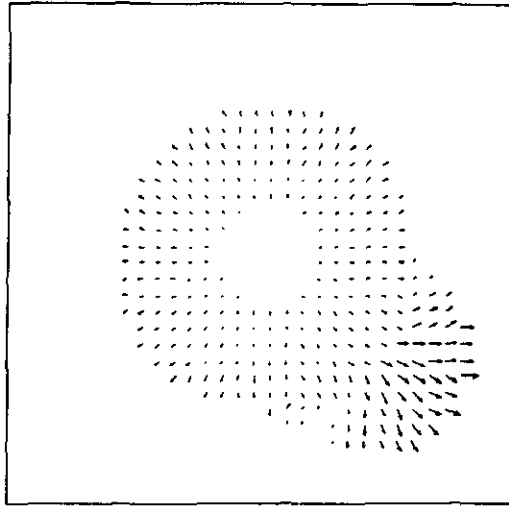
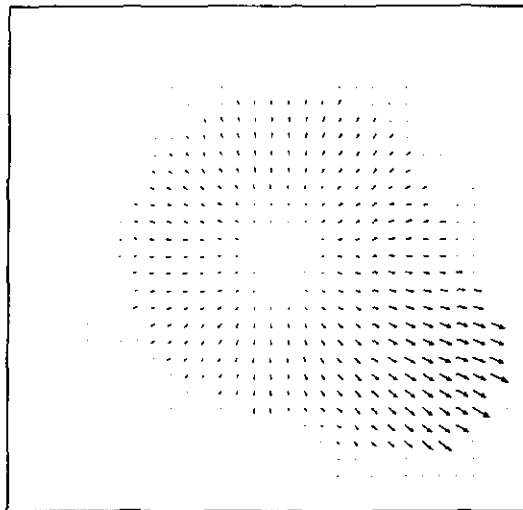


Figura 6.2.- (a) Imagen de textura.(b) Flujo real. Se tomaron dos secciones de la imagen, una cuadrada y otra rectangular, y se desplazaron independientemente de tal forma que el rectángulo recubre al cuadrado. Las velocidades del fondo, el rectángulo y el cuadrado son $(-1,0)$, $(0,2)$ y $(1,-1)$ respectivamente. (c) Estimación inicial. (d) Estimación después de propagar las velocidades (el algoritmo converge después de 11 iteraciones).



(a)



(b)

Figura 6.3.- Estimación para la secuencia SPHERES. (a) Estimación inicial. (b) Estimación después de propagar velocidades (34 iteraciones)

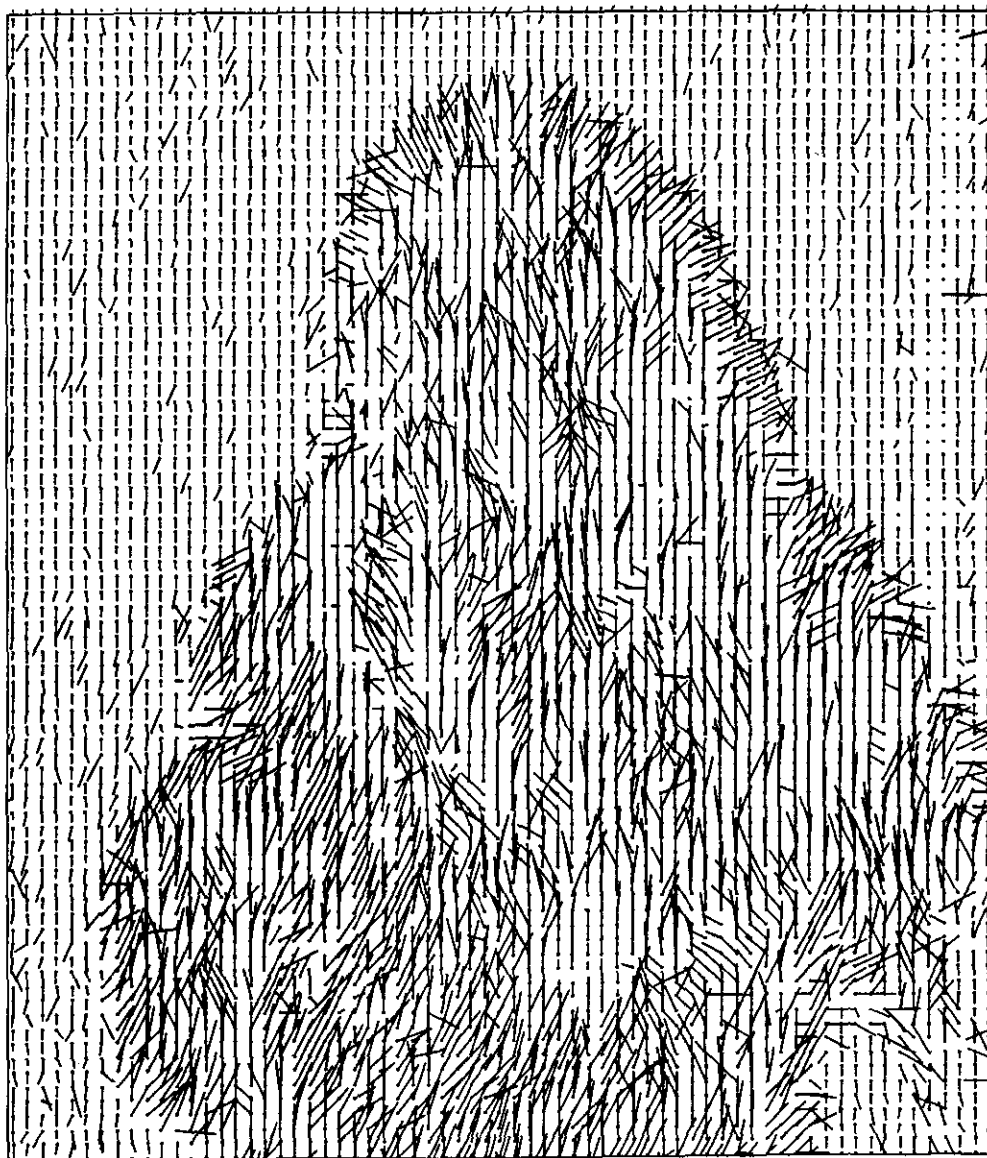


Figura 6.4.- Estimación inicial para la secuencia INTERVIEW.

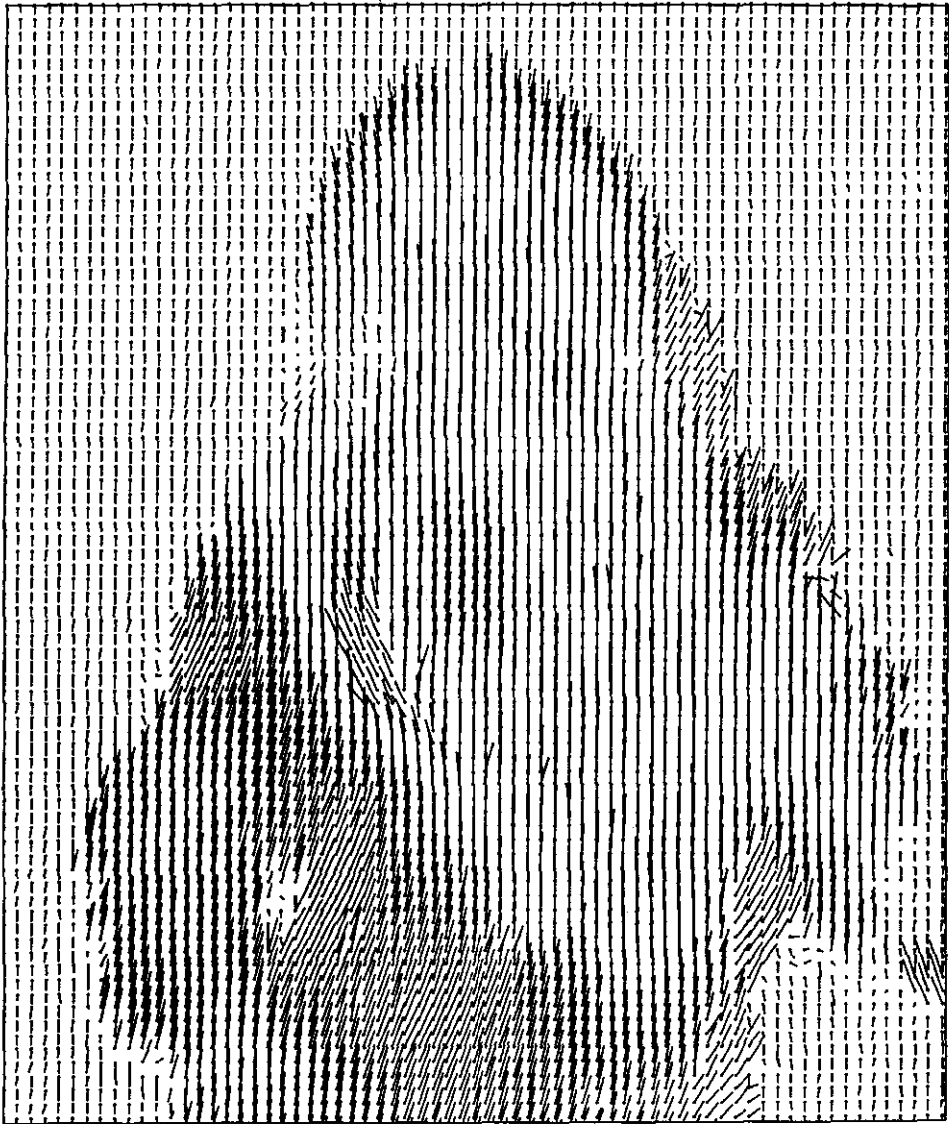


Figura 6.5.- Estimación después de propagar velocidades (20 iteraciones).

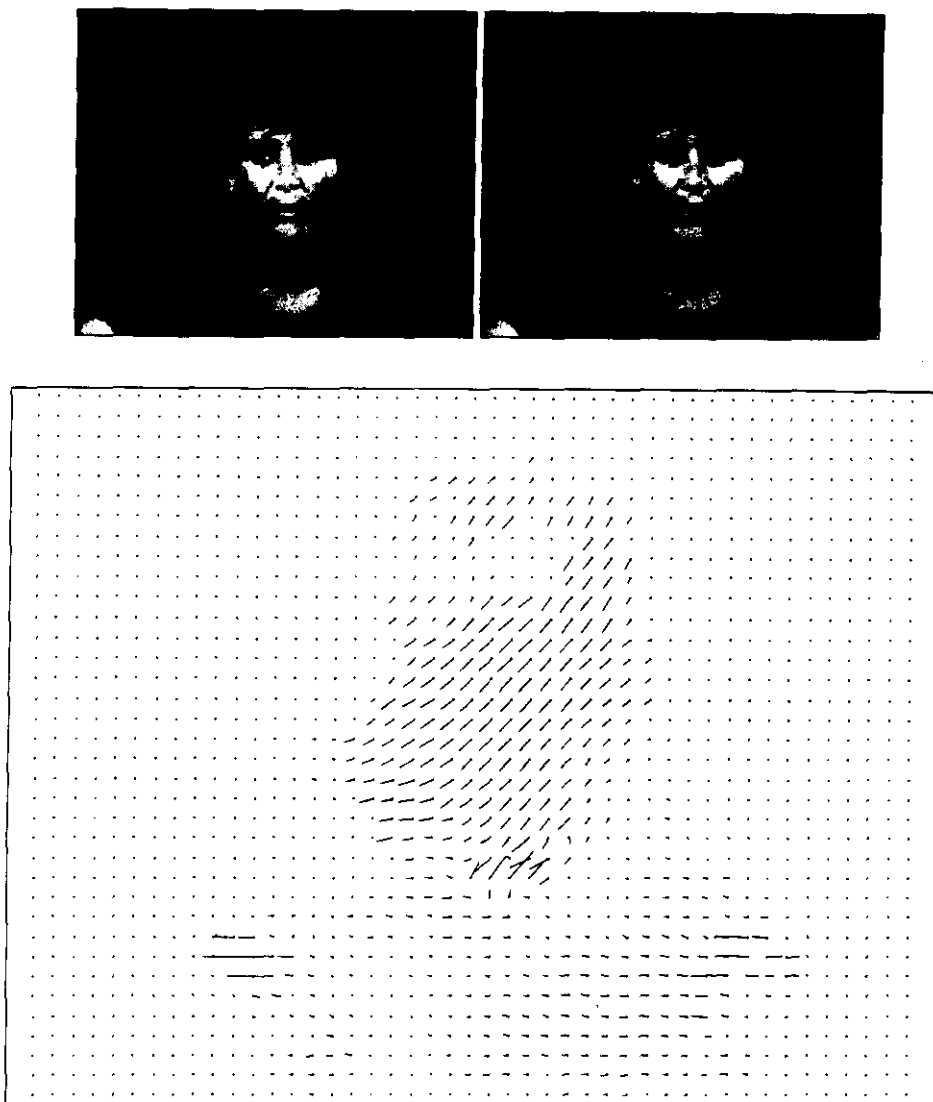


Figura 6.6.- Tramas 1 y 3 de la secuencia MISSAME y estimación después de propagar las velocidades (el algoritmo converge en 15 iteraciones). En este experimento se usó una ventana binomial de 5×5 píxeles \times 3 tramas sobre las tres primeras tramas de MISSAME.

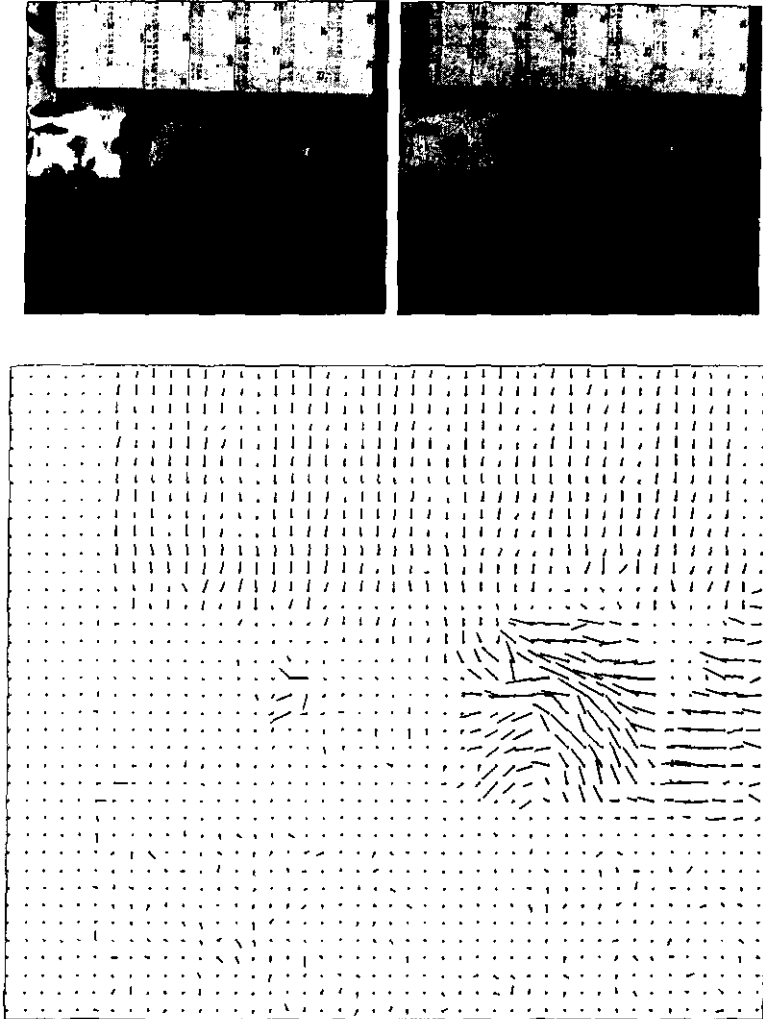


Figura 6.7.- Tramas 3 y 5 de la secuencia CALTRAIN y estimación después de propagar las velocidades (15 iteraciones). En este experimento se uso una ventana binomial de 5×5 pixeles \times 3 tramas sobre las tres primeras tramas de CALTRAIN.

**ESTA TERCERA NO DEBE
QUEDAR DE LA BIBLIOTECA**

Capítulo 7

Conclusiones

En la búsqueda de modelos eficientes para el tratamiento de problemas concernientes a la percepción visual se ha observado que aquellos modelos que incorporan aspectos del sistema visual humano sobresalen ventajosamente sobre las técnicas clásicas dedicadas al procesamiento ciego de la señal, incluso en las tareas básicas tales como segmentación preliminar y detección de *bordes significativos* o *regiones coherentes*. En general, los modelos de percepción visual coinciden en que el procesamiento de un estímulo visual se lleva a cabo en dos etapas: una etapa de medición inicial y una etapa de interpretación de alto nivel.

Fleet y Jepson (1989) apuntaron que el objetivo de la medición del movimiento, en la etapa inicial, es rica en codificación de estructuras de imagen en término de propiedades generales a partir de las cuales es más fácil detectar y analizar estructuras más complejas. Argumentaron que dicho proceso de medición debe ser independiente y no debe requerir de la interpretación previa de la imagen. En contraste, las técnicas basadas en correspondencia de estructuras llevan a cabo dos pasos, (Capítulo 2): en el primero llevan a cabo la extracción de estructuras en tramas aisladas y en el segundo la correspondencia temporal de símbolos similares. Tan solo la primera etapa involucra una gran cantidad de interpretación de la escena al remover ruido y otras características irrelevantes. Por su parte los enfoques basados en el gradiente están más allá del objetivo de la medición inicial. Las estimaciones de velocidad obtenidas por estos métodos son a menudo restringidas a contornos espaciales específicos, usualmente de gran amplitud, alto grado de curvatura o consistentes sobre varias escalas espaciales. Además, la forma particular de la restricción depende de varias suposiciones sobre el flujo. En cambio, el uso de filtros espaciotemporales para la extracción del movimiento no depende del reconocimiento de tipos de estructuras espaciales en tramas aisladas. Esto es de gran importancia especialmente en dominios ruidosos donde es difícil obtener las características espaciales confiablemente. Asimismo, el conjunto de filtros no se diseña necesariamente para una clase restringida de movimientos, y sí proporcionan explícitamente una distribución de movimiento local en una forma que facilita las etapas iniciales de interpretación tales como la segmentación preliminar y la detección de una amplia variedad de tipos movimiento en 3-D.

Adelson y Bergen (1986) mostraron que la energía extraída con los filtros de Gabor puede usarse para estimar el movimiento traslacional en una forma que es equivalente a los métodos basados en gradientes. Heeger (1987) estimó el movimiento traslacional usando mínimos cuadrados del ajuste local de la energía espaciotemporal con el plano impulso en el dominio de la frecuencia. El trabajo presentado en esta tesis difiere de estas dos propuestas en tres aspectos importantes. Primero, en lugar de usar los filtros de Gabor usamos los filtros de la transformada polinomial, los cuales, al ajustarse a las características del sistema visual humano, resultan en los filtros de las transformadas de Hermite (espacial) y de Laguerre (temporal). Estos filtros pueden verse como unidades de procesamiento lineales que en conjunto describen completamente a la señal con la ventaja de ser ortonormales. Segundo, no se requieren hipótesis sobre la forma del espectro de frecuencias del estímulo visual, de hecho, no se requiere la respuesta en frecuencia de la señal, porque todo el análisis es hecho en el dominio espaciotemporal usando el error cuadrático ponderado del ajuste local de la señal a 1-D y 2-D. Y tercero, en nuestro análisis las velocidades están ligadas a direcciones espaciotemporales que son fácilmente estimadas. Esta última característica permite codificar no solo patrones espaciales en movimiento sino también patrones puramente temporales (a los cuales no es posible asociarles un movimiento), resultando ideal para un esquema de codificación de secuencias de imágenes. En general, el modelo permite identificar cinco clases de patrones visuales para algunos de los cuales no es posible estimar el movimiento en forma directa. Estos son:

1. *Patrones planos* ($C^{(3)} < \delta$): corresponden a las regiones uniformes y deben ser analizados a escalas mayores para poder determinar su movimiento. También con la propagación las velocidades se intenta determinar el flujo en estas regiones.
2. *Patrones puramente temporales* ($C^{(3)} \geq \delta$, $\Delta C < \Delta C_{thr}$ y $\phi = \pi/2$): por ejemplo un cambio brusco en la iluminación (*flickering*). No tienen movimiento asociado.
3. *Patrones orientados que experimentan un desplazamiento detectable* ($C^{(3)} \geq \delta$, $\Delta C^{(3-1)} < \Delta C_{thr}$ y $\phi \neq \pi/2$): por ejemplo, bordes o líneas. Es posible estimar su velocidad normal promedio dentro de la ventana temporal si el desplazamiento máximo es comparable con el tamaño de la ventana espacial.
4. *Patrones no orientados que experimentan un desplazamiento detectable* ($C^{(3)} \geq \delta$, $\Delta C^{(3-2)} < \Delta C_{thr}$ y $\phi \neq \pi/2$): por ejemplo, regiones texturizadas o esquinas. Es posible estimar su velocidad promedio dentro de la ventana temporal si el desplazamiento máximo es comparable con el tamaño de la ventana espacial.
5. *Patrones espaciales con movimiento complejo* ($C^{(3)} \geq \delta$, $\Delta C^{(3-2)} \geq \Delta C_{thr}$): por ejemplo, una frontera de oclusión. Bajo ciertas condiciones es posible estimar la tendencia promedio dentro de la ventana espaciotemporal, pero requiere de un análisis de la distribución de energía direccional más detallado.

En suma, con base en el residuo del error de aproximación se puede determinar si la información local es bastante próxima a la de un patrón orientado, si es bidimensional o si existe mucha información que sugiere que el movimiento local es muy complejo para ser modelado como una traslación simple. Es posible estudiar otros tipos de movimiento usando este enfoque. Por ejemplo, se ha usado la energía espaciotemporal para estudiar propiedades del movimiento turbulento [Heeger & Pentland, 1988]. También podría ser de interés considerar superficies semitransparentes donde se asocia más de un movimiento a una misma posición.

En el Capítulo 6 se introdujo un método de propagación de velocidades que suaviza el campo preservando discontinuidades. El análisis del campo de velocidades con transformadas polinomiales se ve motivado por las características polinomiales del movimiento encontradas en el estudio del flujo de imagen [Verri & Poggio, 1989] y por la eficacia de los filtros de primero y segundo orden en la detección y descripción de bordes [Martens, 1990b]. El método utiliza la propuesta de Singh (1991) ajustado al esquema de la transformada polinomial. Es un método iterativo que propaga la información de movimiento local hasta las fronteras de disrupción en forma robusta, elimina artefactos, preserva las discontinuidades fuertemente orientadas y puede resolver el problema de apertura.

Trabajos futuros sobre esta línea de investigación deberán dedicar mayor esfuerzo en los siguientes puntos. Primero, una revisión más rigurosa de los resultados teóricos. Segundo, se debe llevar a cabo una evaluación experimental y sistemática de los algoritmos presentados aquí en relación con otras técnicas sobre un conjunto de imágenes de referencia y comprobar la confiabilidad de las estimaciones. Tercero, optimar los parámetros introducidos para la detección en términos de percepción visual. Cuarto, investigar los mecanismos en que la información extraída en cada nivel de la jerarquía (en el enfoque jerárquico) puede ser *compartida* y *refinada* más eficientemente entre las diferentes capas. Quinto, la generalización al estudio de movimientos complejos para aplicaciones específicas.

Referencias

- Adelson, E. H. & Bergen, J. R., 1985, "Spatiotemporal energy models for the perception of motion", *Journal of the Optical Society of America A*, 2:284-299
- Adelson, E. H. & Bergen, J. R., 1986, "The extraction of spatiotemporal energy in human and machine vision", *IEEE Workshop on Motion, Representation and Analysis*, 151-155
- Anandan, P., 1989. "A computational framework and an algorithm for the measurement of visual motion". *International Journal on Computer Vision*, 2:283-310.
- Bevington, J. & Mersereau, R., 1987, "Differential operator based edge and line detection" in *Proceeding ICASSP*, 249-252.
- Blommaert, F. J. J. & Martens, J. B., 1990. "An object oriented model for brightness perception". *Spatial Vision* 5, 15-41.
- Daugman, J., 1980. "Two-dimensional spectral analysis of cortical receptive fields profiles", *Vision Res.*, vol. 20, 847-856
- den Brinker, A. C. & Roufs, J. A. J., 1992, "Evidence for a generalized laguerre transform of temporal events by visual system". *Biological Cybernetics*. 395-402.
- Escalante-Ramírez, B. & Martens, J. B., 1992. "Noise reduction in Computerized Tomography images by means of polynomial transforms", *Journal of Visual Communications and Image Representation*.
- Escalante-Ramírez, B. & Venegas-Martínez, S., 1995. "Pyramidal predictive image coding with polynomial transforms", *Still-Image Compression*, Proceedings SPIE, 2418:107-117.
- Fleet, D. J. & Jepson, A. D., 1985. "Velocity extraction without form interpretation". *IEEE Workshop on computer vision: Representation and Control*, pp 179-185
- Fleet, D. J. & Jepson, A. D., 1989. "Hierarchical construction of orientation and velocity selective filters". *IEEE Trans. on Pattern Analysis and Machine Intelligence* 11(3): 315-325
- Geman, S. & Geman, D., 1984. "Stochastic relaxation, Gibbs distribution, and the bayesian restoration of images". *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6(6):721-741.

- Gibson, J. J., 1950. *The perception of the visual world*. Houghton Mifflin, Boston, MA.
- Hashimoto, M. & Skalansky, J., 1987. "Multi-order derivatives for detecting local image characteristics". *Computer Vision, Graphics, and Image Processing* 29, 28-55.
- Heeger, D. J., & Pentland, A. P., 1988. "Seeing structure through chaos", In *Proceeding of IEEE Workshop Motion: Representation and analysis*, Charleston May 1986. pp 131-136
- Heeger, D. J., 1987. "A model for the extraction of image flow", *Journal of the Optical Society of America A*, 4:1455-1471
- Heeger, D. J., 1988. "Optical flow using spatiotemporal filters", *International Journal of the Optical Society of America*, pp 279-302
- Hildreth, E. C., 1983, "The computation of velocity field", MIT Artificial Intelligence Laboratory Memo
- Hildreth, E. C., 1984, "Computation underlying the measurement of visual motion", *Artificial Intelligence*, 23:309-354
- Horn, B. K. P. & Schunck B.G., 1981, "Determining the optical flow", *Artificial Intelligence* 17. pp 185-203.
- Huck, F. O., 1983. "Information theory analysis of sensor array imaging systems for computer vision". *Proceeding of the SPIE*, Vol 397, pp 82.
- Julesz, B., 1971, "Foundations of Cyclopean Perception", *University of Chicago Press*.
- Koenderink, J & Van Doorn, A. J., 1975. "Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer". *Optical acta*. 22:773-791.
- Koenderink, J & Van Doorn, A. J., 1978. "Visual detection of spatial contrast: influence of location in the visual field, target, extent and illuminance level". *Biological Cybernetics* 30, 157-167.
- Koenderink, J & Van Doorn, A. J., 1982. "Invariant Features of contrast detection: an explanation in terms of self-similar detector arrays". *Journal of the Optical Society of America* 72, 83-87.
- Marr, D. & Hildreth, E., 1980, "A theory of edge detection", *Proceeding Roy. Soc. Londo B*, vol. 207, pp. 187-217
- Marr, D. & Poggio, T., 1979, "A computational theory of vision", *Proceeding Roy. Soc. Londo B*, 204, 301
- Marr, D., 1982. *Vision*, W.H. Freeman and Company, San Francisco.
- Martens, J. B., 1990a. "The Hermit Transform - theory". *IEEE Trans. Acoust. Speech Signal Processing*, 38: 1595-1606
- Martens, J. B., 1990b. "The Hermit Transform - applications *IEEE Trans. Acoust. Speech Signal Processing*, 38: 1607-1618

- Martens, J. B., 1997, "Local orientation analysis in images by means of the Hermite transform". *IEEE Trans on Image Processing*, vol. 6, pp 2-15
- Mitiche, A & Bouthemy, P., 1996. "Computation and analysis of Image Motion: A Synopsis of Current Problems and Methods". *International Journal of Computer Vision* 19(1), 29-55
- Moctezuma, M. & Escalante Ramírez, B., 1997. "Coastline detection with polynomial transform and markovian segmentations". *IEEE Geoscience and Remote Sensing Society, IGARSS'97*.
- Pessing, V., 1963. "Some asymptotic properties of the negative binomial distributions", *Annals of Mathematical Statistic*, 32, 922-923 (abstract).
- Sakitt, B. & Barlow, H., 1982. "A model for the economical encoding of visual image in cerebral cortex". *Biological Cybernetics.*, vol 43. pp 97-108.
- Scott, G. L., 1988, *Local and Global interpretation of moving Images*, Morgan Kaufman Publishers, Inc., Los Altos, Calif.
- Schreiber, W. F., 1967. "Picture coding". *Proceeding of the IEEE* 55, 320-330.
- Silván-Cárdenas, J. L. & Escalante-Ramírez, B. "Motion estimation by means of polynomial transforms", en *Memorias de Computación Visual 98*. Méx., D.F. abril 1998.
- Silván-Cárdenas, J. L. & Escalante-Ramírez, B. "Polynomial transform application to motion estimation", Mathematical modeling and estimation techniques in Computer Vision (E. Dougherty, J.L. Davison, T. Petreux, eds.), *Proceeding of SPIE*. vol. 3457, 1998.
- Singh, A., 1991, "Optic Flow Computation: A Unified Perspective". *IEEE Computer Society Press*. Los Alamitos, Cal.
- Spillman L., Ransom-Hogg A. & Oehler R., 1987, "A comparison of perceptive and receptive fields in man and monkey", *Human Neurobiology*, 6, 55.
- Stockham, T. G. Jr., 1972. "Image Processing in the context of visual model". *Proceeding of the IEEE* 60, 828-842.
- Szegő, G., 1959. "Orthogonal Polynomials" (American Mathematical Society). Colloquium Publications
- Tompson, W. B. & Kearney, J. K., 1986. "Inexact Vision", *Proceeding of the IEEE Workshop on motion: Representation and Analysis*, Charleston, pp 15-21
- Torre, V. & Poggio, T., 1986, "On edge detection". *IEEE Trans Pattern-Analysis Mach, Intell.*, vol. 8, pp. 147-163
- Van Doorn, A. J. & Koenderink, J. J., 1984, "Spatiotemporal integration in the detection of coherent motion". *Vision Research*. 24, 47
- Verri, A. & Poggio, T., 1989. "Motion field and optical flow: Qualitative properties". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:490-498

- Watson, A. B., & Ahamuda, A. J., 1985, "Model of human vision motion sensing", *Journal of the Optical Society of America A*, 2:322-342
- Weng, J., 1990. "A theory of image matching". In *Proceeding 3rd International Conference on Computer Vision*, 200-209
- Young, R., 1986, "The Gaussian derivative theory of spatial vision: Analysis of cortical cell receptive field line-weighting profiles". General Motors Res. Labs., Rep. 4920
- Yuille, A. L. & Grzywacz, N. M., 1992. "A mathematical analysis of the motion coherence theory". *Int. Journal of Computer Vision*, 3:155-175