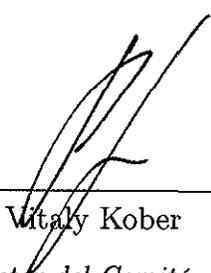


TESIS DEFENDIDA POR

**Pablo Mario Aguilar González**

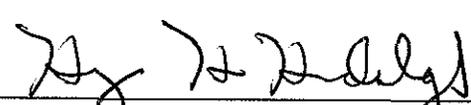
Y APROBADA POR EL SIGUIENTE COMITÉ



---

Dr. Vitaly Kober

*Director del Comité*



---

Dr. Hugo Homero Hidalgo Silva

*Miembro del Comité*



---

Dr. Pedro Gilberto López Mariscal

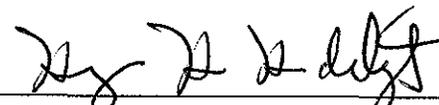
*Miembro del Comité*



---

Dr. Juan Humberto Sossa Azuela

*Miembro del Comité*



---

Dr. Hugo Homero Hidalgo Silva

*Coordinador del programa de  
posgrado en Ciencias de la Computación*



---

Dr. David Hilario Covarrubias Rosales

*Director de Estudios de Posgrado*

24 noviembre de 2011

**CENTRO DE INVESTIGACIÓN CIENTÍFICA Y DE  
EDUCACIÓN SUPERIOR DE ENSENADA**



---

**PROGRAMA DE POSGRADO EN CIENCIAS  
EN CIENCIAS DE LA COMPUTACIÓN**

---

**DISEÑO DE FILTROS ADAPTATIVOS PARA EL RECONOCIMIENTO  
DE OBJETOS MEDIANTE IMÁGENES RUIDOSAS DE REFERENCIA**

**TESIS**

que para cubrir parcialmente los requisitos necesarios para obtener el grado de  
**DOCTOR EN CIENCIAS**

Presenta:

**PABLO MARIO AGUILAR GONZÁLEZ**

Ensenada, Baja California, México, noviembre de 2011.

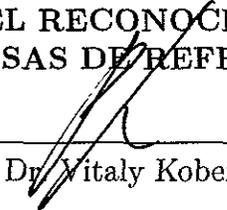
---

**RESUMEN** de la tesis de **PABLO MARIO AGUILAR GONZÁLEZ**, presentada como requisito parcial para la obtención del grado de **DOCTOR EN CIENCIAS** en **CIENCIAS DE LA COMPUTACIÓN**. Ensenada, Baja California, noviembre de 2011.

**DISEÑO DE FILTROS ADAPTATIVOS PARA EL RECONOCIMIENTO DE OBJETOS MEDIANTE IMÁGENES RUIDOSAS DE REFERENCIA**

Resumen aprobado por:

---



Dr. Vitaly Kober

Director de Tesis

El reconocimiento de patrones basado en correlación ha sido ampliamente estudiado. Los filtros de correlación comúnmente se diseñan optimizando analíticamente un criterio de desempeño para un modelo de señales. Esto permite la derivación de filtro óptimos para una aplicación dada. Sin embargo, la derivación comúnmente se realiza bajo el supuesto de que se conocen explícitamente la apariencia y forma del objeto de interés. En situaciones prácticas, el objeto se muestra en una imagen de la vida real y su apariencia se extrae mediante segmentación manual. Aún cuando la referencia se captura en un ambiente controlado, las imágenes inevitablemente contienen ruido debido al proceso de formación de imagen. Además, el desempeño de los filtros de correlación se ve afectado por cambios en la apariencia del objeto en la escena. En esta tesis se proponen nuevos modelos de señales y se derivan filtros óptimos para detectar un objeto en escenas aditivas y no traslapadas. Los modelos propuestos consideran la presencia de ruido en la imagen de referencia usada en el diseño de filtros. Al considerar el ruido como parte del modelo, no se requiere segmentación manual. De esta manera, se puede reconocer objetos en forma automática. Los parámetros necesarios para el diseño del filtro se estiman a partir de una imagen ruidosa de referencia. Se proponen diferentes estimaciones para cada modelo. Simulaciones por computadora muestran que los filtros y estimaciones propuestas permiten la detección automática confiable de objetos aún en la presencia de ruido moderado. El desempeño de los filtros se evalúa en cuanto a su capacidad de discriminación y errores de localización. Se propone un algoritmo adaptativo para el diseño de filtros compuestos para tomar en cuenta cambios de apariencia del objeto debidos a distorsiones geométricas. El algoritmo propuesto se basa en funciones discriminantes sintéticas y usa información sobre el objeto, el ruido en la escena, y un ejemplo de fondo de la escena. Con un proceso iterativo, la capacidad de discriminación del filtro se mejora hasta alcanzar un nivel deseado. Simulaciones por computadora muestran que los filtros adaptativos propuestos pueden detectar objetos con distorsiones aún cuando las imágenes de referencia contengan ruido fuerte. Se muestra una aplicación del algoritmo propuesto para rastrear objetos en secuencias de videos en tiempo real. El sistema utiliza un banco de filtros diseñado para garantizar una precisión de localización deseada con el mínimo número de filtros.

**Palabras Clave:** Filtros de correlación, reconocimiento de patrones, filtros adaptativos compuestos.

**ABSTRACT** of the thesis presented by **PABLO MARIO AGUILAR GONZÁLEZ**, as a partial fulfillment of the requirements of the **DOCTOR OF SCIENCE** degree in **COMPUTER SCIENCE**. Ensenada, Baja California, November 2011.

## **DESIGN OF ADAPTIVE FILTERS FOR OBJECT RECOGNITION USING NOISY REFERENCE IMAGES**

Correlation pattern recognition has been extensively studied over the years. Correlation filters are commonly designed by means of analytical optimization of a performance criterion for a given signal model. This allows the derivation of optimum filters for a given application. However, the filter derivation is commonly carried out assuming that the appearance and shape of an object of interest are explicitly available. In practical situations, a target is given in a real-life image, and its appearance is extracted by manual segmentation. Even in situations, in which a reference is captured in a controlled environment, images are inevitably corrupted by noise due to the image formation process. Furthermore, changes in the appearance of the target in a scene affect the detection performance of correlation filters. In this thesis, new signal models are proposed, and optimum filters are derived for detecting a target in additive and nonoverlapping scenes. The proposed signal models take into account the presence of noise in the reference image used for filter design. By considering the noise as part of the model, no manual segmentation step is required. So, it allows automatic target recognition. The parameters required for optimum filter design are estimated from a single noisy reference image. Different estimation methods are proposed for each signal model. Computer simulations show that the proposed filters and suggested estimations allowing reliable automatic target recognition even in the presence of moderate levels of noise. Filter performance is evaluated in terms of discrimination capability and location errors. An adaptive algorithm for composite filter design is proposed to take into account the appearance changes of the target owing to geometric distortions. The proposed algorithm is based on synthetic discriminant functions and uses information about the target, noise processes in the scene, and a sample background of the input scene. Using iterative procedure the discrimination capability of the filter is improved until it reaches a desired level. Simulation results show that the proposed adaptive filters can detect a geometrically distorted target even if the reference images are corrupted by strong noise. An application of the proposed algorithm to tracking objects in video sequences in real-time is presented. The system uses a bank of filters designed to guarantee a desired location accuracy with the minimum the number of filters.

**Keywords:** *Correlation filters, pattern recognition, adaptive composite filters*

---

*A mi familia*  
*y a la memoria de mi madre*

---

# Agradecimientos

Quiero expresar mi más sincero agradecimiento:

A mi familia por su amor y su apoyo en todo.

A Nancy por su amor, comprensión y paciencia, y su apoyo incondicional.

Al Dr. Vitaly Kober por su paciencia, sus consejos y su dedicación para formarme como investigador.

A los Doctores Gilberto López, Hugo Hidalgo y Humberto Sossa por su paciencia, consejos y apoyo.

A los miembros del Departamento de Ciencias de la Computación por sus enseñanzas y consejos.

A todos los compañeros con los que me tocó compartir nuestro paso por CICESE, en especial a la generación 2006 por su amistad y compañerismo.

A Yee y a Saracho por su amistad y apoyo.

A la fundación Yee Romero por el hospedaje durante los primeros años en Ensenada.

A la Olimpiada Mexicana de Matemáticas de Baja California (OMMBC) por todo lo que con ellos he aprendido y que tanto me sirvió para sacar adelante este trabajo.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico para la realización de este proyecto.

Ensenada, México

Pablo Mario Aguilar González

Noviembre de 2011

---

# Contenido

	Página
Resumen en español	I
Resumen en inglés	II
Dedicatoria	III
Agradecimientos	IV
Contenido	V
Lista de Figuras	IX
<b>I. Introducción</b>	<b>1</b>
I.1. Definición del problema y objetivos . . . . .	2
I.1.1. Objetivo general . . . . .	3
I.1.2. Objetivos particulares . . . . .	3
I.2. Limitaciones y suposiciones . . . . .	3
I.3. Enfoques para el reconocimiento de patrones . . . . .	4
I.4. Trabajo previo . . . . .	7
I.4.1. Detección de objetos en escenas contaminadas por ruido aditivo	7
I.4.2. Detección de objetos distorsionados por ruido en escenas con fondos complejos . . . . .	8
I.4.3. Detección de objetos en escenas con fondo disjunto cuando no hay información estadística <i>a priori</i> . . . . .	9
I.4.4. Detección de objetos cuando la estructura de gris del objeto puede fluctuar . . . . .	9
I.4.5. Detección de objetos utilizando imágenes de entrenamiento con fondos complejos . . . . .	10
I.4.6. Detección de objetos con distorsiones geométricas . . . . .	10
I.5. Organización de la tesis . . . . .	11
<b>II. Fundamentos matemáticos</b>	<b>13</b>
II.1. Representación de imágenes . . . . .	13
II.2. Sistemas lineales . . . . .	14
II.2.1. Transformada de Fourier . . . . .	17
II.3. Teoría de probabilidad y procesos estocásticos . . . . .	21
II.3.1. Variables aleatorias . . . . .	21
II.3.2. Valor esperado . . . . .	23

# Contenido (continuación)

	Página
II.3.3. Momentos . . . . .	23
II.3.4. Dos variables aleatorias . . . . .	24
II.3.5. Procesos estocásticos . . . . .	25
II.3.6. Procesos estacionarios . . . . .	27
II.3.7. Densidad espectral . . . . .	28
II.3.8. Ergodicidad . . . . .	28
II.4. Optimización . . . . .	28
II.4.1. Desigualdad de Cauchy-Schwarz . . . . .	28
II.4.2. Multiplicadores de Lagrange . . . . .	29
II.5. Resumen . . . . .	29
<b>III. Teoría de detección</b>	<b>31</b>
III.1. Modelos de escena . . . . .	31
III.1.1. Modelo aditivo (traslapado) . . . . .	31
III.1.2. Modelo disjunto (no traslapado) . . . . .	32
III.2. Métricas de desempeño . . . . .	34
III.2.1. Razón señal a ruido (SNR) . . . . .	34
III.2.2. Razón pico a energía de correlación (POE) . . . . .	34
III.2.3. Error cuadrático medio (MSE) . . . . .	35
III.2.4. Capacidad de discriminación (DC) . . . . .	36
III.2.5. Errores de localización (LE) . . . . .	36
III.3. Distorsiones que degradan el desempeño de los filtros de correlación .	37
III.3.1. Modelos de ruido . . . . .	37
III.3.2. Distribuciones de ruido . . . . .	40
III.3.3. Distorsiones geométricas . . . . .	41
III.4. Teoría de estimación . . . . .	41
III.4.1. Estimación de parámetros . . . . .	42
III.5. Resumen . . . . .	46
<b>IV. Filtros clásicos de correlación</b>	<b>47</b>
IV.1. Filtros para el modelo aditivo . . . . .	47
IV.1.1. Filtro de correspondencia . . . . .	47
IV.1.2. Filtro óptimo . . . . .	49
IV.1.3. Filtro de sólo fase . . . . .	50
IV.2. Filtros para el modelo no traslapado . . . . .	51
IV.2.1. Filtro de correspondencia generalizado . . . . .	51
IV.2.2. Filtro óptimo generalizado . . . . .	53
IV.3. Filtros compuestos . . . . .	53

---

## Contenido (continuación)

	Página
IV.3.1. Funciones discriminantes sintéticas . . . . .	54
IV.3.2. Filtros de correlación de Mínima Energía Promedio de Correlación . . . . .	55
IV.3.3. Filtros de correlación de Máxima Altura Promedio de Correlación . . . . .	57
IV.4. Resumen . . . . .	59
<b>V. Diseño de filtros para el reconocimiento de objetos mediante imágenes ruidosas de referencia</b>	<b>60</b>
V.1. Modelo Aditivo-Aditivo (AA) . . . . .	61
V.1.1. Modelo de señales . . . . .	61
V.1.2. Derivación del filtro óptimo . . . . .	63
V.1.3. Estimación de parámetros . . . . .	66
V.1.4. Resultados experimentales . . . . .	67
V.2. Modelo Aditivo-No traslapado (AN) . . . . .	71
V.2.1. Estimación de parámetros . . . . .	75
V.2.2. Resultados experimentales . . . . .	78
V.3. Modelo No traslapado-No traslapado (NN) . . . . .	83
V.3.1. Estimación de parámetros . . . . .	85
V.3.2. Resultados experimentales . . . . .	90
V.4. Resumen . . . . .	97
<b>VI. Diseño de filtros adaptativos para el reconocimiento de objetos mediante imágenes ruidosas de referencia</b>	<b>99</b>
VI.1. Filtros compuestos de respuestas de impulso con adaptación a fondos típicos . . . . .	100
VI.1.1. Resultados experimentales . . . . .	103
VI.2. Aplicación a seguimiento de objetos (tracking) . . . . .	107
VI.2.1. Formulación del problema . . . . .	111
VI.2.2. Diseño de filtros compuestos para reconocimiento invariante a distorsiones geométricas . . . . .	112
VI.2.3. Sistema opto digital de rastreo . . . . .	115
VI.2.4. Resultados experimentales . . . . .	116
VI.3. Resumen . . . . .	121
<b>VII. Conclusiones</b>	<b>123</b>
VII.1. Trabajo futuro . . . . .	125
VII.2. Publicaciones derivadas de este trabajo . . . . .	125
<b>Bibliografía</b>	<b>128</b>

## Contenido (continuación)

	Página
Apéndice A	133
Apéndice B	136
Apéndice C	138

---

# Lista de Figuras

Figura	Página
1. Esquema general de un sistema de clasificación basado en características.	5
2. Ejemplo de escena traslapada . . . . .	32
3. Ejemplo de escena no traslapada . . . . .	33
4. (a) Objeto de interés y (b) realización de un fondo estocástico. . . . .	68
5. Desempeño de los filtros cuando el ruido es blanco y variando la desviación estándar del ruido en la imagen de referencia. (a) DC y (b) LE cuando $\zeta_s = 20$ ; (c) DC y (d) LE cuando $\zeta_s = 30$ . . . . .	69
6. Desempeño de los filtros cuando el ruido es correlacionado con coeficiente de correlación de 0.5, variando la desviación estándar del ruido en la imagen de referencia. (a) DC y (b) LE cuando $\zeta_s = 20$ ; (c) DC y (d) LE cuando $\zeta_s = 30$ . . . . .	70
7. Desempeño de los métodos de estimación de la apariencia del objeto de interés al variar $\mu_s$ , mientras que $\sigma_t = 10$ y $\zeta_r = 10$ . (a) MSE cuando $\mu_t = 100$ , (b) ME cuando $\mu_t = 100$ , (c) MSE cuando $\mu_t = 60$ y (d) ME cuando $\mu_t = 60$ . . . . .	80
8. Desempeño de los filtros de correlación al variar la desviación estándar del ruido en la imagen de referencia y $AN_1$ es mejor que $AN_2$ . $\mu_t = 100$ , $\sigma_t = 10$ y $\mu_s = 160$ . (a) DC (con intervalo de 95 % de confianza) y (b) LE.	81
9. Desempeño de los filtros de correlación al variar la desviación estándar del ruido en la imagen de referencia y $AN_2$ es mejor que $AN_1$ . $\mu_t = 140$ , $\sigma_t = 130$ y $\mu_s = 100$ . (a) DC (con intervalo de 95 % de confianza) y (b) LE. . . . .	82
10. Partición del espacio de parámetros en regiones en las que cada estimación tiene mejor desempeño que la otra con respecto a las métricas (a) DC, (b) ME, y (c) MSE. . . . .	83
11. Imágenes estocásticas usadas en los experimentos: (a) objeto con apariencia aleatoria, (b) fondo de la imagen de referencia, (c) fondo de la escena de entrada. . . . .	88

## Lista de Figuras (continuación)

Figura	Página
12. (a) Errores de estimación de forma al variar $\mu_r$ y $\sigma_r$ ; (b) errores de estimación de forma al variar $\mu_r$ y $\mu_s$ , y (c) errores mostrados en (b) después de amplificar por $ \mu_s - \mu_r $ . . . . .	88
13. Errores de correspondencia al variar $\mu_r$ y $\mu_s$ : (a) entre $\hat{t}_s^1(x)$ y $t_s(x)$ ; (b) entre $\hat{t}_s^2(x)$ y $t_s(x)$ con preprocesamiento, y (c) entre $\hat{t}_s^2(x)$ y $t_s(x)$ . . .	90
14. Imágenes reales usadas en los experimentos: (a) objeto de interés, (b) fondo de la imagen de referencia, (c) fondo de la escena de entrada. . .	91
15. Desempeño de los filtros al variar la desviación estándar del ruido aditivo en la imagen de referencia, en términos de (a) DC y (b) LE cuando $\mu_r = 80$ y $\mu_s = 60$ , y (c) DC y (d) LE cuando $\mu_r = 80$ y $\mu_s = 180$ . . . .	93
16. Desempeño de los filtros en términos de DC al variar $\mu_r$ y $\mu_s$ usando imágenes estocásticas: (a) $NN_1$ aplicado a la escena de entrada, (b) $NN_1$ aplicado a la escena de entrada preprocesada con el filtro Wiener y (c) $NN_2$ aplicado a la escena de entrada. . . . .	94
17. Desempeño de los filtros al variar $\mu_r$ y $\mu_s$ usando imágenes determinísticas: (a) $NN_1$ aplicado a la escena de entrada, (b) $NN_2$ aplicado a la escena de entrada preprocesada con el filtro Wiener y (c) $NN_2$ aplicado a la escena de entrada. . . . .	95
18. Desempeño de los filtros en términos de DC en el espacio de parámetros $(\mu_r, \mu_s)$ variando la desviación estándar del fondo en la imagen de referencia. En el renglón superior $NN_1$ y en el renglón inferior $NN_2$ . La desviación estándar del fondo de la imagen de referencia es (a), (d) 15; (b), (e) 30, y (c), (f) 45. . . . .	96
19. Diagrama del algoritmo para el diseño de filtros adaptativos. . . . .	103
20. Desempeño de los filtros compuestos, en términos de (a) DC y (b) LE cuando se varía la rotación del objeto. La desviación estándar del ruido en las imágenes de entrenamiento es 10. . . . .	104
21. Desempeño de los filtros propuestos en términos de (a) DC y (b) LE cuando se varía la rotación del objeto. El objeto aparece escalado a 0.7 veces su tamaño original y la desviación estándar del ruido en las imágenes de referencia es 15. . . . .	105

---

## Lista de Figuras (continuación)

Figura	Página
22. Desempeño de los filtros propuestos en términos de (a) DC y (b) LE cuando se varía la escala del objeto. La desviación estándar del ruido en las imágenes de entrenamiento es 10. . . . .	106
23. Desempeño de los filtros propuestos en términos de (a) DC y (b) LE cuando se varía la desviación estándar del ruido en las imágenes de entrenamiento. . . . .	107
24. Desempeño de los filtros propuestos en términos de (a) DC y (b) LE cuando se varía la desviación estándar del ruido en las imágenes de entrenamiento. El objeto aparece en la escena escalado a 1.15 veces su tamaño original y rotado 10.5°. . . . .	108
25. Diagrama del algoritmo de diseño del banco de filtros. . . . .	114
26. Diagrama del sistema opto digital de seguimiento. . . . .	116
27. Imágenes usadas en los experimentos: (a) Imagen de referencia ejemplo, (b) escena ejemplo. . . . .	118
28. Imágenes del objeto de interés distorsionadas por rotación y escalamiento. . . . .	118
29. Fondo usado como fondo ejemplo en el algoritmo adaptativo de diseño de filtros. . . . .	118
30. (a) Cuadro ejemplo y (b) su salida de correlación correspondiente. . .	119
31. Errores de localización en cada cuadro en la secuencia de imágenes para el sistema que utiliza uno, cuatro, seis y nueve filtros compuestos. . . .	120
32. Secuencia ejemplo con 25 cuadros. . . . .	120

# Capítulo I

---

## Introducción

---

Una capacidad fundamental de los seres humanos es la vista, la capacidad de detectar y clasificar objetos en su entorno de manera visual. Por tanto, un paso importante en el desarrollo de la inteligencia artificial es el desarrollo de la visión artificial, en la medida posible, similar a la de los humanos. El aumento en capacidades del equipo de cómputo disponible ha hecho posible el desarrollo, aunque limitado, de técnicas de *visión artificial*. Dos problemas importantes en el área de visión, son los de detección y localización de objetos. Por tanto son problemas clásicos en el área, y sumamente estudiados. Sin embargo, las técnicas de detección y localización desarrolladas hasta el momento, dependen, entre otras cosas, de la especificación precisa del objeto a detectar, esto es, proporcionando la representación visual del objeto a detectar sin un fondo, sobre un fondo controlado o proporcionando una región de soporte que delimite exactamente al objeto. La visión humana, por el contrario, rara vez requiere de una especificación de un objeto, a ese nivel de precisión, para poder detectarlo, incluso en fondos complejos. Una oportunidad, entonces, es el desarrollo de técnicas de detección y localización que se asemejen a la visión humana, en el sentido de poder funcionar utilizando definiciones de un objeto embebidas en una imagen con un fondo no controlado.

Existen diversas situaciones de la vida real en las que la apariencia y forma del objeto no se conocen en forma explícita, sino que se obtienen de una imagen de referencia por medio de algún preprocesamiento como segmentación manual. En este caso, la calidad de la referencia depende de la experiencia del operador humano del sistema. Por otro

---

lado, hay aplicaciones en las que se obtiene una referencia en línea, es decir, no hay tiempo para el preprocesamiento manual. Considérese, por ejemplo, aplicaciones de seguridad en las que se extrae la apariencia de un sujeto a partir de una secuencia de video y se le debe seguir a lo largo de la secuencia. Otro aplicación es la búsqueda basada en ejemplos, en la que se proporciona una imagen que se sabe que contiene al objeto de interés y se buscan objetos similares en un banco de imágenes, pero el usuario no segmenta explícitamente el objeto de interés en la imagen ejemplo.

Se han estudiado diversas técnicas en la literatura para el reconocimiento de patrones. Se puede ver un resumen de estrategias en Suetens *et al.* (1992). Algunas son simples heurísticas y otras tienen bases matemáticas bien sentadas como los clasificadores estadísticos (Duda *et al.*, 2000), el reconocimiento basado en correlación (Kumar *et al.*, 2005), las redes neuronales (Duda *et al.*, 2000), entre otros. El presente trabajo de investigación se enfoca al reconocimiento de objetos en dos dimensiones en imágenes utilizando técnicas basadas en correlación.

## I.1. Definición del problema y objetivos

Los problemas de detección y localización consisten en decidir si un objeto de interés está presente en una escena observada, y en caso de estarlo, determinar su localización. El enfoque de filtros de correlación supone que se conoce en forma explícita la apariencia del objeto de interés. Sin embargo, hay casos en los que no se puede obtener la apariencia y forma exacta del objeto, sino que se conoce únicamente de manera implícita en una imagen ruidosa de referencia. Para llevar a cabo el reconocimiento se toma en cuenta información estadística sobre el objeto de interés, los procesos de ruido que degradan la escena y los fondos sobre los que puede aparecer el objeto. El problema a tratar en

---

la presente investigación se resume en los objetivos que se presentan a continuación.

### **I.1.1. Objetivo general**

Desarrollar técnicas adaptativas de detección y localización de objetos en imágenes de dos dimensiones a partir de imágenes de referencia ruidosas, con las que se pueda garantizar un grado de confiabilidad respecto a distintas métricas de desempeño y diferentes modelos de señal, que presenten invariancia a distorsiones geométricas y distintas fuentes de ruido

### **I.1.2. Objetivos particulares**

- Obtener conocimiento teórico sobre las técnicas de detección actuales.
- Desarrollar técnicas de detección en dos dimensiones que sean capaces de detectar objetos definidos implícitamente.
- Desarrollar técnicas adaptativas para garantizar confiabilidad en el reconocimiento aún en presencia de distorsiones geométricas y distintos tipos de ruido.

## **I.2. Limitaciones y suposiciones**

El trabajo se centra en la detección y localización de objetos en imágenes de dos dimensiones. El enfoque que se sigue es el de reconocimiento de patrones mediante filtros de correlación. Se consideran dos modelos de escena: el modelo aditivo (traslapado) y el modelo disjunto (no traslapado). No se consideran degradaciones del objeto por cambios de iluminación ni oclusiones parciales. Se supone que se dispone de información *a priori* sobre los procesos de ruido presentes en las imágenes. Se suponen conocidas

---

las estadísticas del objeto de interés como su media, varianza y tamaño en relación al tamaño de la imagen. En caso de que algunas estadísticas no se conozcan se da por hecho que se pueden estimar tomando en cuenta la naturaleza de las imágenes utilizadas.

### I.3. Enfoques para el reconocimiento de patrones

En el reconocimiento de patrones se encuentran tres tareas principales: detección, localización y clasificación (Réfrégier y Goudail, 2003). La tarea de detección consiste en determinar si un objeto está presente en una escena o no. En caso de que se determine que el objeto está presente, la tarea de localización consiste en estimar la ubicación del objeto en la escena. Finalmente, la tarea de clasificación consiste en asignar una clase a un objeto de entre varias clases conocidas. El problema de reconocimiento de patrones aplica tanto para objetos o señales de una (audio), dos (imágenes) o varias dimensiones (imágenes multiespectrales, secuencias de video). El problema de detección, se puede tratar como un caso particular de clasificación en el que sólo hay dos clases conocidas, cuando el objeto de interés está presente y cuando no. Para llevar a cabo las tareas de reconocimiento de patrones se siguen dos enfoques principales: el basado en características y el basado en apariencia.

El enfoque estándar basado en características se ilustra en la figura 1 (Kumar *et al.*, 2005). El primer paso es realizar algún tipo de preprocesamiento sobre la imagen observada cuyo objetivo es hacer más fácil la extracción de características, ya sea eliminando ruido, mejorando contrastes, o algún otro tipo de proceso que generalmente depende de la aplicación. El siguiente paso es la extracción de características a partir de la imagen preprocesada. Las características que se extraen tienen como objetivo el capturar la esencia de los objetos presentes en una escena. El número de características extraídas

---

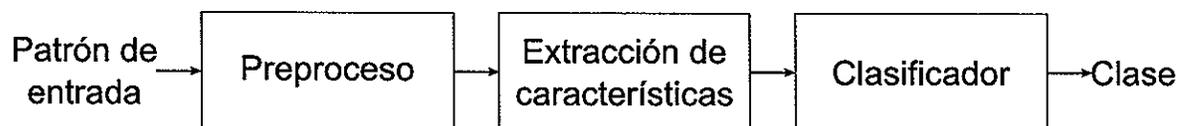


Figura 1: Esquema general de un sistema de clasificación basado en características.

normalmente es mucho menor que la cantidad de píxeles en la imagen. La selección de características a extraer es el proceso que requiere más cuidado en este enfoque. No existe un conjunto de características que sirvan para obtener buen desempeño para todos los problemas de reconocimiento de patrones, y por lo tanto, depende mucho de la experiencia e intuición de un experto en el área de la aplicación.

Las características extraídas de la imagen luego se usan como entrada para un clasificador. El objetivo del clasificador es separar las características como pertenecientes a una de varias clases. Un clasificador se diseña para optimizar algún criterio de desempeño, como minimizar la probabilidad de errores de clasificación (Webb, 2002). El entrenamiento del clasificador se realiza a partir de un conjunto de entrenamiento cuya clasificación es conocida<sup>1</sup>. Existen varios tipos de clasificadores como clasificadores estadísticos, clasificadores basados en redes neuronales artificiales, clasificadores basados en lógica difusa, entre otros (Duda *et al.*, 2000). La selección de un tipo de clasificador depende de la métrica de desempeño y de la información disponible *a priori*. Cuando se conocen las distribuciones de probabilidad de los datos a clasificar, un clasificador estadístico resulta ser lo más adecuado. Si no se conocen distribuciones de probabilidad de las características pero se dispone de ejemplos de diferentes clases se pueden usar Máquinas de Vectores de Soporte (Cortes y Vapnik, 1995) o Redes Neuronales Artificiales (Webb, 2002).

La principal desventaja del enfoque basado en características radica en la dificultad

---

<sup>1</sup>Conocido como "Ground truth" en inglés.

para determinar *a priori* qué características se deben utilizar para una aplicación dada. La selección del tipo de clasificador y de los valores para sus parámetros de ajuste dependen de la experiencia del experto que lo utiliza. Entonces, se propone un sistema de clasificación y se mide su desempeño *a posteriori*.

El enfoque alternativo para el reconocimiento de patrones se basa en apariencia (o imagen). En este enfoque se utilizan los datos de entrenamiento en forma directa y no características extraídas de ellos. La clasificación se hace mediante técnicas de correlación. El procedimiento general consiste en seleccionar una señal (o imagen) de referencia para después determinar el grado de similitud que tiene con la escena observada. La señal de referencia también se conoce como filtro. Los problemas de detección y localización se resuelven mediante correlación en dos pasos: se buscan los picos más altos en el plano de correlación tras aplicar el filtro, y posteriormente se utilizan las coordenadas de los picos como estimación de la posición de los objetos en la escena (Kumar *et al.*, 2005).

Los algoritmos de correlación tienen como objetivo el diseñar una señal de referencia que presente ciertas características deseables, como puede ser la tolerancia a cambios de apariencia del objeto de interés, la capacidad de discriminar entre el objeto deseado y otros objetos similares, tomar en cuenta la presencia de ruido, entre otras.

La principal ventaja del enfoque basado en correlación es que toma como base áreas de matemáticas bien desarrolladas. Se hace uso de teoría de probabilidad y estadística, de teoría de procesamiento de señales y técnicas analíticas de optimización. El diseño de filtros de correlación se realiza optimizando una métrica de desempeño para un modelo de escena seleccionado. El hecho de que la optimización se hace en forma analítica garantiza un buen desempeño de los filtros *a priori*.

La correlación se puede implementar tanto en el dominio espacial o en el dominio de

---

la frecuencia. De esta manera, la detección mediante correlación se puede implementar en forma eficiente tanto en sistemas digitales como en sistemas ópticos (Javidi, 1994). En el caso de implementaciones digitales, se hace uso de los algoritmos rápidos para el cálculo de la transformada de Fourier. Por otro lado, la implementación en correladores ópticos permite el reconocimiento basado en correlación en aplicaciones que requieren de alta velocidad de procesamiento, tales como el rastrear objetos en secuencias de video en tiempo real.

## **I.4. Trabajo previo**

El reconocimiento de patrones basado en correlación utiliza información de la apariencia y forma del objeto de interés. Por este motivo, si la apariencia del objeto en la escena difiere de la apariencia con la que se diseñó el filtro, el desempeño se ve afectado significativamente. Los cambios de apariencia del objeto se pueden deber a la presencia de ruido en la escena, a distorsiones geométricas como rotación, escalamiento o cambios de perspectiva, a cambios de iluminación, y finalmente, a la falta de información precisa sobre el objeto o el fondo. A continuación se presentan trabajos del área de reconocimiento de patrones mediante filtros de correlación que están cercanamente relacionados con el problema de detección y localización de objetos cuando la apariencia del objeto en la escena difiere de la apariencia usada en el diseño del filtro.

### **I.4.1. Detección de objetos en escenas contaminadas por ruido aditivo**

El reconocimiento de objetos en imágenes utilizando correlación se inicia con el trabajo de VanderLugt (1964). Propone un filtro para maximizar la razón señal a ruido

---

en imágenes que se representan con el modelo aditivo. Presenta la implementación del filtro propuesto utilizando correladores ópticos. Más adelante, en Kumar *et al.* (1992) se demostró que el filtro de correspondencia propuesto por VanderLugt también es óptimo en el sentido de que minimiza los errores de localización.

Para minimizar la probabilidad de errores anómalos o falsas alarmas, Yaroslavsky (1993) propuso el filtro óptimo. Este filtro está diseñado para maximizar la intensidad de la salida de correlación en presencia del objeto mientras que el fondo tiende a ser plano si el objeto no está presente.

#### **I.4.2. Detección de objetos distorsionados por ruido en escenas con fondos complejos**

Javidi y Wang (1992) mostraron que los filtros diseñados para detectar objetos en ruido aditivo presentan bajo desempeño cuando el objeto de interés se encuentra sobre un fondo espacialmente disjunto. Javidi y Wang (1994a) propusieron el filtro generalizado de correspondencia y el filtro óptimo generalizado para el modelo disjunto. Los filtros se obtienen mediante la optimización de criterios de desempeño que introdujeron para corregir las limitaciones de los criterios clásicos para el modelo aditivo. Kober y Campos (1996) demostraron que el filtro generalizado de correspondencia es el filtro óptimo para minimizar errores de localización para el modelo disjunto. Javidi *et al.* (1996) utilizaron la métrica de error cuadrático medio como alternativa para el diseño de filtros de correlación.

---

### **I.4.3. Detección de objetos en escenas con fondo disjunto cuando no hay información estadística *a priori***

En el proceso clásico de diseño de filtros de correlación se supone que se conoce información *a priori* sobre las estadísticas del fondo y del ruido presente en la escena. Towghi y Javidi (2001b) presentan un receptor basado en correlación que supone que el ruido tiene distribución Gaussiana pero con parámetros desconocidos. Posteriormente presentaron receptores que funcionan en el caso de ruido no Gaussiano (Towghi y Javidi, 2001a).

### **I.4.4. Detección de objetos cuando la estructura de gris del objeto puede fluctuar**

Goudail y Réfrégier (1996) propusieron un procesador basado en correlación que permite la detección de objetos cuando la apariencia del objeto en la escena puede variar. Utilizan el modelo de regiones estadísticamente independientes (SIR<sup>2</sup>). Bajo la suposición de que se conoce la forma del objeto y de que el ruido en la escena es blanco, el procesador para detección se diseña para encontrar las regiones en la imagen en la que las estadísticas dentro del área del objeto y fuera de ésta son lo más diferentes posible. Posteriormente propusieron un refinamiento de la técnica SIR para considerar las estadísticas de la región de transición entre el objeto y el fondo para el caso de fondos correlacionados (Guérault *et al.*, 1997). Finalmente, propusieron un marco de trabajo unificado en el que se pueden detectar objetos en escenas tanto con fondo blanco como con fondo correlacionado (Guérault y Réfrégier, 1998).

---

<sup>2</sup>Statistically Independent Regions

#### **I.4.5. Detección de objetos utilizando imágenes de entrenamiento con fondos complejos**

Papageorgiou y Poggio (2000) propusieron un sistema de reconocimiento de objetos basado en entrenamiento. Para una clase particular de objetos se forma un conjunto de imágenes de entrenamiento y se selecciona un conjunto de características que sean relevantes para la clase en cuestión. El sistema aprende una representación de la clase de objetos de interés que luego puede ser usada para reconocer objetos independiente de cambios de escala. El principal problema con el sistema propuesto es la cantidad de imágenes de entrenamiento que se deben utilizar para alcanzar buenas tasas de reconocimiento.

#### **I.4.6. Detección de objetos con distorsiones geométricas**

Para resolver el problema de distorsiones geométricas en la escena se utiliza la técnica de filtros compuestos que, usando un conjunto de imágenes de entrenamiento, toman en cuenta diversas apariencias posibles del objeto. El primer enfoque utilizado es el de funciones discriminantes sintéticas (Casasent, 1984) en el que un filtro compuesto se modela como una combinación lineal del conjunto de entrenamiento que satisface restricciones prefijadas para el valor del pico de correlación. La desventaja de este enfoque es que solo controla el pico de correlación y no el resto del plano, por lo que es posible la aparición de lóbulos laterales que produzcan falsas alarmas. Para compensar esto, Mahalanobis *et al.* (1987) propusieron una técnica de diseño de filtros compuestos que además de controlar el pico de correlación, minimiza la energía promedio en la salida del filtro, minimizando así la probabilidad de falsas alarmas. Una desventaja de estas dos técnicas es que no incorporan explícitamente tolerancia a distorsiones intra-

---

clase del objeto. Para esto, Mahalanobis *et al.* (1994) propusieron el diseño de filtros que maximizan la intensidad del pico de correlación mientras se minimiza la variación de la respuesta del filtro debida al ruido.

## I.5. Organización de la tesis

El resto del presente trabajo está organizado como se describe a continuación. En el Capítulo II se presentan los fundamentos matemáticos que se utilizan para el modelado y procesamiento de imágenes. Se presentan los conceptos de teoría de filtrado, teoría de probabilidad y conceptos de optimización que son esenciales para el análisis y diseño de filtros de correlación.

En el Capítulo III se presentan las definiciones formales de los modelos de escena utilizados en este trabajo. Se presentan también las métricas de desempeño que se utilizan en el reconocimiento de objetos basado en correlación. Los filtros de correlación se obtienen mediante la optimización de uno o más de los criterios presentados. Finalmente, se analizan los factores que degradan el desempeño de los filtros de correlación tales como la presencia de ruido y distorsiones geométricas.

En el Capítulo IV se presentan los principales filtros de correlación para los modelos aditivo y disjuncto de escena. Se discuten las características principales de cada filtro y se introducen las técnicas de diseño de filtros que se utilizan para derivar los filtros propuestos en capítulos posteriores. Se presenta también la derivación de filtros compuestos para el reconocimiento de objetos tolerante a distorsiones geométricas.

En el Capítulo V se presenta el diseño de filtros a partir de imágenes ruidosas de referencia. Se consideran tres modelos de señales en los que se toma en cuenta la presencia de ruido aditivo y fondos complejos en las imágenes de referencia. Los resultados

del modelo más sencillo se utilizan como base para derivar filtros para el siguiente modelo. Los filtros óptimos que se obtienen a partir de la optimización de un criterio de desempeño requieren de estimaciones que se hacen a partir de la imagen de referencia. Para cada uno de los tres modelos se presentan estimaciones alternativas y se evalúan en términos de capacidad de discriminación y errores de localización. El desempeño de los filtros se compara con filtros clásicos de correlación.

El Capítulo VI tiene como propósito presentar filtros tolerantes a distorsiones geométricas como rotación y escalamiento. Se utiliza un enfoque iterativo para diseñar filtros compuestos que garanticen un nivel de desempeño deseado en términos de capacidad de discriminación. Se presenta también la aplicación de los filtros propuestos para un sistema de rastreo de objetos en secuencias de video utilizando bancos de filtros compuestos. Se presenta una técnica iterativa para el diseño de bancos de filtros que garanticen un nivel de desempeño deseado en términos de errores de localización. Se discute también la implementación óptica digital del sistema propuesto.

En el Capítulo VII se presentan las conclusiones del presente trabajo de investigación así como áreas de oportunidad para trabajos futuros.

En el Apéndice A se muestra el procedimiento para la generación de ruido correlacionado que se utiliza en la realización de simulaciones por computadora. En el Apéndice B se muestra el análisis matemático para determinar el umbral teórico para separar dos distribuciones Gaussianas. Finalmente, en el Apéndice C se muestra el análisis teórico de las estimaciones de la apariencia del objeto de interés considerando el criterio de error cuadrático medio.

# Capítulo II

---

## Fundamentos matemáticos

---

### II.1. Representación de imágenes

Una imagen se puede representar como una función continua de de dos o tres variables. Para imágenes estáticas se utilizan funciones de dos variables cuyos argumentos representan coordenadas  $(x, y)$  en un plano. Para imágenes dinámicas se añade un tercer argumento que representa el tiempo y la imagen se denota por  $I(x, y, t)$ . La amplitud de la función se denomina *intensidad* o *nivel de gris*. Por simplicidad matemática, se trabaja bajo la suposición de que todas las imágenes son distintas de cero sólo dentro de una región rectangular finita del plano y que son observables solo en un intervalo finito de tiempo (Pratt, 2007).

Para el diseño y análisis de sistemas de procesamiento de imágenes se utilizan dos caracterizaciones matemáticas: la determinística y la estadística (Jain, 1989; Pratt, 2007). Al tratar las imágenes como funciones determinísticas o como realizaciones de procesos estocásticos, es necesario utilizar herramientas matemáticas provenientes de la teoría de sistemas lineales, de transformadas integrales, de matemáticas discretas y de la teoría de procesos estocásticos. Para las transformadas integrales, se supone que la función de la imagen es integrable y que existe su transformada de Fourier. Para imágenes discretas se considera que la transformada discreta de Fourier siempre existe.

## II.2. Sistemas lineales

La teoría de sistemas lineales es de gran importancia en el procesamiento digital de imágenes (Bovik, 2005). Muchas de las técnicas de mejoramiento o transformaciones de información visual digital se pueden expresar en términos de los conceptos de sistemas lineales. En el procesamiento de imágenes, los filtros lineales se aplican a la imagen en general, para tareas como el mejoramiento de contraste, eliminación de ruido, realce de contornos y para tareas más específicas sobre los objetos que componen una imagen, tales como el reconocimiento de patrones. Una clase importante de sistemas es la de sistemas lineales invariantes a desplazamiento (LSI<sup>1</sup>). Estos sistemas se pueden caracterizar completamente por la salida del sistema para una entrada particular, a la cual se le denomina entrada puntual en el origen y a la salida resultante se le denomina respuesta al impulso.

Un sistema se puede representar por un proceso de transformación para determinar una salida  $g(x)$  a partir de una entrada  $f(x)$ , es decir

$$\mathbf{T}[f(x)] = g(x). \quad (1)$$

El sistema  $\mathbf{T}$  es *lineal* si y sólo si para dos entradas arbitrarias  $f_1(x)$  y  $f_2(x)$  se tienen como salidas  $g_1(x)$  y  $g_2(x)$ , respectivamente, y para dos constantes  $a$  y  $b$  se tiene que

$$\mathbf{T}[af_1(x) + bf_2(x)] = a\mathbf{T}[f_1(x)] + b\mathbf{T}[f_2(x)], \quad (2)$$

para toda  $(x)$ . Esto se conoce como el principio de superposición en sistemas lineales.

El sistema  $\mathbf{T}$  es *invariante a desplazamiento* si para toda  $f(x)$  se cumple

$$\mathbf{T}[f(x - x')] = g(x - x'), \quad (3)$$

---

<sup>1</sup>Linear Shift Invariant

para cualquier desplazamiento  $x'$ . Esto significa que un desplazamiento en la señal de entrada produce un desplazamiento idéntico en la señal de salida.

La respuesta al impulso unitario  $\delta(x - x')$  de un sistema  $\mathbf{T}$  se denota por  $h(x)$ , es decir

$$\mathbf{T}[\delta(x - x')] = h(x; x'), \quad (4)$$

que es la respuesta del sistema  $\mathbf{T}$  en la posición espacial  $(x)$  debido a un impulso localizado en la posición  $(x')$ . Cuando el sistema  $\mathbf{T}$  es invariante a desplazamiento, si

$$\mathbf{T}[\delta(x)] = h(x), \quad (5)$$

representa la respuesta a un impulso en el origen, se tiene que

$$\mathbf{T}[\delta(x - x')] = h(x - x'). \quad (6)$$

Lo anterior significa que la respuesta a un impulso aplicado en cualquier posición espacial se puede encontrar a partir de la respuesta al impulso dada en la ecuación (5). En otras palabras,  $h(x)$  puede caracterizar completamente a un *sistema lineal invariante a desplazamiento*. La respuesta  $g(x)$  resultado de una entrada arbitraria  $f(x)$  se puede encontrar en términos de  $h(x)$  por medio de una convolución lineal (Kumar *et al.*, 2005), es decir

$$\begin{aligned} g(x) &= \mathbf{T}[f(x)] \\ &= \mathbf{T}\left[\int_{-\infty}^{\infty} f(x') \delta(x - x') dx'\right]. \end{aligned} \quad (7)$$

Si el sistema es lineal, entonces

$$\begin{aligned} g(x) &= \int_{-\infty}^{\infty} f(x') \mathbf{T}[\delta(x-x')] dx' \\ &= \int_{-\infty}^{\infty} f(x') h(x; x') dx', \end{aligned} \quad (8)$$

y si el sistema es invariante a desplazamiento, se tiene

$$\begin{aligned} g(x) &= \int_{-\infty}^{\infty} f(x') h(x-x') dx' \\ &= f(x) \bullet h(x), \end{aligned} \quad (9)$$

donde el símbolo  $\bullet$  denota la operación de convolución. De esta manera, la ecuación (9) representa una convolución entre las funciones  $f(x)$  y  $h(x)$ . Como la integral de convolución es simétrica, se tiene que  $f(x) \bullet h(x) = h(x) \bullet f(x)$ . La correlación es otra operación relacionada con la convolución, con la diferencia de que la correlación no es simétrica. La relación entre las operaciones de correlación y convolución está dada por

$$\begin{aligned} f(x) \circ h(x) &= \int_{-\infty}^{\infty} f^*(x') h(x+x') dx' \\ &= \int_{-\infty}^{\infty} f^*(-x') h(x-x') dx' \\ &= f^*(-x) \bullet h(x), \end{aligned} \quad (10)$$

donde  $\circ$  denota la operación de correlación y el superíndice\* denota el complejo conjugado.

La transformada de Fourier de la respuesta al impulso  $h(x)$  de un sistema lineal se conoce como *respuesta de frecuencia* o *función de transferencia* del sistema. Los filtros lineales en el procesamiento de imágenes se caracterizan en términos de su respuesta de frecuencia.

Las operaciones de convolución y correlación se llevan a cabo en el dominio de

frecuencia empleando los algoritmos de la transformada rápida de Fourier (FFT<sup>2</sup>). A continuación se analiza la transformada de Fourier y sus propiedades.

### II.2.1. Transformada de Fourier

Una forma de analizar una función imagen es tratándola como combinación lineal de funciones ortogonales. La transformada de Fourier de una función se define por la integral

$$F(\omega) = \int_{-\infty}^{\infty} f(x) e^{-i\omega x} dx. \quad (11)$$

Para el propósito del procesamiento de imágenes es razonable suponer que la transformada de Fourier es una función periódica que siempre existe. La transformada inversa de Fourier se define por la integral

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{i\omega x} d\omega. \quad (12)$$

La variable  $(x)$  denota la coordenada en el dominio espacial y la variable  $\omega$  denota la coordenada en el dominio de frecuencia. La función  $f(x)$  puede ser interpretada como una combinación lineal de funciones base periódicas  $e^{i\omega x}$ . Las componentes reales e imaginarias de estas funciones base son las funciones seno y coseno, y la función  $F(\omega)$  es una función de peso que representa la influencia de las funciones base en  $f(x)$ . En general, las componentes de la transformada de Fourier son cantidades complejas. A veces es conveniente expresar  $F(\omega)$  en forma polar:

$$F(\omega) = |F(\omega)| e^{i\phi(\omega)}, \quad (13)$$

---

<sup>2</sup>Fast Fourier Transform

donde

$$|F(\omega)| = \sqrt{[\operatorname{Re}\{F(\omega)\}]^2 + [\operatorname{Im}\{F(\omega)\}]^2}, \quad (14)$$

se le llama *magnitud* de la transformada de Fourier y

$$\phi(\omega) = \arctan \left[ \frac{\operatorname{Im}\{F(\omega)\}}{\operatorname{Re}\{F(\omega)\}} \right], \quad (15)$$

se le llama la *fase* de la transformada. En las ecuaciones (13) y (14),  $\operatorname{Re}\{F(\omega)\}$  e  $\operatorname{Im}\{F(\omega)\}$  representan las partes real e imaginaria de  $F(\omega)$ , respectivamente. Otra cantidad importante es el espectro de potencia, definido por el cuadrado de la magnitud:

$$\begin{aligned} P(\omega) &= |F(\omega)|^2 \\ &= F(\omega) F^*(\omega) \\ &= [\operatorname{Re}\{F(\omega)\}]^2 + [\operatorname{Im}\{F(\omega)\}]^2. \end{aligned} \quad (16)$$

Para simplificar la notación, se denota la transformada de Fourier por el operador  $\mathcal{F}$  y la transformada inversa por  $\mathcal{F}^{-1}$ . De esta manera, las ecuaciones (11) y (12) se pueden abreviar como

$$F(\omega) = \mathcal{F}\{f(x)\}, \quad (17)$$

y

$$f(x) = \mathcal{F}^{-1}\{F(\omega)\}, \quad (18)$$

respectivamente.

### II.2.1.1. Propiedades de la transformada de Fourier

Las siguientes propiedades de la transformada de Fourier resultan interesantes desde el punto de vista del procesamiento de imágenes:

### Linealidad

La transformada de Fourier es un operador lineal, por lo que para dos funciones  $f(x)$  y  $g(x)$  y dos constantes  $a$  y  $b$ , se cumple que

$$\mathcal{F}\{af(x) + bg(x)\} = aF(\omega) + bG(\omega). \quad (19)$$

### Desplazamiento en el dominio espacial

$$\mathcal{F}\{f(x - a)\} = F(\omega) e^{-i\omega a}. \quad (20)$$

### Desplazamiento en el dominio de la frecuencia

$$\mathcal{F}\{f(x) e^{iax}\} = F(\omega - a). \quad (21)$$

### Simetría conjugada

$$\mathcal{F}\{f^*(x)\} = F^*(-\omega). \quad (22)$$

Si la función es real, la transformada de Fourier es simétrica conjugada, esto es

$$F(\omega) = F^*(-\omega). \quad (23)$$


---

### II.2.1.2. Teorema de convolución

La convolución de dos funciones  $f(x)$  y  $h(x)$  se define en la ecuación (9). El teorema de convolución establece que

$$\mathcal{F}\{f(x) \bullet h(x)\} = F(\omega) H(\omega), \quad (24)$$

$$\mathcal{F}\{f(x) h(x)\} = F(\omega) \bullet H(\omega), \quad (25)$$

donde  $F(\omega)$  y  $H(\omega)$  son las transformadas de Fourier de  $f(x)$  y  $h(x)$ , respectivamente.

### II.2.1.3. Teorema de correlación

La correlación de dos funciones complejas  $f(x)$  y  $h(x)$  se define en la ecuación (10). El teorema de correlación establece que

$$\mathcal{F}\{f(x) \circ h(x)\} = F^*(\omega) H(\omega), \quad (26)$$

$$\mathcal{F}\{f^*(x) h(x)\} = F(\omega) \circ H(\omega), \quad (27)$$

donde  $F(\omega)$  y  $H(\omega)$  son las transformadas de Fourier de  $f(x)$  y  $h(x)$ , respectivamente.

### II.2.1.4. Teorema de Parseval

El teorema de Parseval establece la preservación de la energía de una señal  $f(x)$  en el dominio espacial y en el dominio de frecuencia. Esto es

$$\int_{-\infty}^{\infty} |f(x)|^2 dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} |F(\omega)|^2 d\omega. \quad (28)$$

En general, para dos funciones complejas, se cumple que

$$\int_{-\infty}^{\infty} f(x) g^*(x) dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) G^*(\omega) d\omega. \quad (29)$$

## II.3. Teoría de probabilidad y procesos estocásticos

Las imágenes sufren de perturbaciones impredecibles como el ruido, la oclusión, los cambios de iluminación, distorsiones geométricas, etc. Esta aleatoriedad sólo puede caracterizarse mediante modelos probabilísticos, cuya aplicación permite entender lo que le sucede a entradas aleatorias al pasar a través de un sistema de procesamiento de señales, tales como los filtros de correlación.

Para su caracterización estocástica, una imagen se considera una realización de un arreglo de variables aleatorias llamado *campo aleatorio*. Esta caracterización de un ensemble de imágenes es útil en técnicas de procesamiento de imágenes para validar una clase completa y no sólo una imagen individual. Para el manejo de esta caracterización se requieren conceptos de teoría de probabilidad, variables aleatorias y proceso estocásticos, que se describen a continuación.

### II.3.1. Variables aleatorias

La aleatoriedad de las medidas es inherente tanto a los fenómenos físicos como a la variabilidad dentro de los sistemas de observación y medición. Su descripción cuantitativa esta contenida en el concepto de variable aleatoria (Papoulis, 1991).

En teoría de probabilidad, el espacio  $S$  corresponde a todos los posible resultados que se pueden obtener de un experimento. Si a cada salida  $\xi$  del experimento se le asigna un número  $x(\xi)$ , entonces  $x$  queda definida como una variable aleatoria que indicará la regla de correspondencia entre cualquier elemento del conjunto  $S$  y el número asignado a éste (Papoulis, 1991). Los conceptos básicos relacionados con variables aleatorias se presentan a continuación.

Las probabilidades asociadas con una variable aleatoria pueden expresarse por medio

de la función de distribución acumulativa:

$$F(x) = P\{\mathbf{x} \leq x\}, \quad (30)$$

donde  $x$  indica el valor que toma la variable aleatoria  $\mathbf{x}$ . Una variable aleatoria se caracteriza por su función de densidad de probabilidad (PDF<sup>3</sup>) que es la derivada de la función de distribución acumulativa:

$$f(x) = \frac{d}{dx}F(x). \quad (31)$$

Integrando la PDF sobre un intervalo se puede determinar la probabilidad de que la variable aleatoria tome un valor dentro de ese intervalo, es decir

$$P\{x_1 < \mathbf{x} \leq x_2\} = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x) dx. \quad (32)$$

Dado que la función de distribución acumulada es no decreciente, la PDF nunca es negativa. El área total bajo la curva de la PDF debe ser 1. Existen muchas distribuciones de probabilidad, de las cuales la de mayor importancia es la función de densidad de probabilidad Gaussiana.

Una variable aleatoria continua  $\mathbf{x}$  tiene distribución *normal* o *Gaussiana* si su PDF tiene la siguiente forma:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}. \quad (33)$$

La PDF Gaussiana se caracteriza por dos parámetros, su media  $\mu$  y su varianza  $\sigma^2$ . La raíz cuadrada  $\sigma$  de la varianza se denomina *desviación estándar*.

---

<sup>3</sup>Probability Density Function

### II.3.2. Valor esperado

El *valor esperado* o *esperanza* de una función de una variable aleatoria  $\mathbf{x}$  con PDF  $f(x)$  se define como

$$E\{g(\mathbf{x})\} = \int_{-\infty}^{\infty} g(x) f(x) dx. \quad (34)$$

El valor esperado es un operador lineal, por lo que se satisface que

$$E\{ag(\mathbf{x}) + bh(\mathbf{x})\} = aE\{g(\mathbf{x})\} + bE\{h(\mathbf{x})\}, \quad (35)$$

para cualquier par de escalares  $a$  y  $b$ .

### II.3.3. Momentos

El  $n$ -ésimo momento de una variable aleatoria se define como sigue:

$$E\{\mathbf{x}^n\} = \int_{-\infty}^{\infty} x^n f(x) dx. \quad (36)$$

El primer momento de  $\mathbf{x}$  se llama valor promedio (o media) de la variable aleatoria y es la suma pesada de todos los valores posibles de la variable, donde el peso está dado por la función de densidad de probabilidad. Esto es

$$E\{\mathbf{x}\} = \mu = \int_{-\infty}^{\infty} x f(x) dx. \quad (37)$$

El segundo momento central se denomina *varianza*. La varianza de una variable aleatoria  $\mathbf{x}$ ,  $\text{Var}\{\mathbf{x}\}$ , se define como

$$\text{Var}\{\mathbf{x}\} = \sigma^2 = E\{(\mathbf{x} - E\{\mathbf{x}\})^2\} = E\{\mathbf{x}^2\} - \mu^2. \quad (38)$$

### II.3.4. Dos variables aleatorias

En aplicaciones de procesamiento de imágenes suelen encontrarse múltiples variables aleatorias, de manera que se requieren herramientas para describirlas y manejarlas (Kumar *et al.*, 2005). Los conceptos que se presentan a continuación para dos variables aleatorias son aplicables a cuando se tienen más variables aleatorias. La función de distribución acumulada conjunta de dos variables aleatorias se define como sigue:

$$F(x, y) = P\{\mathbf{x} \leq x, \mathbf{y} \leq y\}. \quad (39)$$

Esta función de distribución conjunta describe la conducta aleatoria conjunta de las dos variables aleatorias, no de cada variable por separado. La función de distribución acumulada es no negativa, no decreciente y debe tender a 1 cuando  $x$  y  $y$  tienden a infinito.

La función de densidad de probabilidad conjunta de dos variables aleatorias  $x$  y  $y$  se define de la siguiente manera:

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y). \quad (40)$$

La probabilidad de que dos variables aleatorias tomen un valor dentro de una región en el plano  $(x, y)$  se obtiene integrando la PDF conjunta sobre esa región. La integral de la PDF sobre todo el plano  $(x, y)$  es uno.

A partir de la PDF conjunta se obtiene la *PDF marginal* de cada variable, como se muestra a continuación:

$$f_x(x) = \int_{-\infty}^{\infty} f(x, y) dy, \quad (41)$$

$$f_y(y) = \int_{-\infty}^{\infty} f(x, y) dx. \quad (42)$$

Las PDFs marginales describen lo que le sucede a cada variable aleatoria por separado, sin importar lo que pasa con la otra variable. Por esta razón, las PDFs marginales  $f_x(x)$  y  $f_y(y)$  no son suficiente para conocer la PDF conjunta  $f(x, y)$ .

Dos variables aleatorias se dicen estadísticamente independientes si y sólo si su PDF conjunta es separable, es decir

$$f(x, y) = f_x(x) f_y(y). \quad (43)$$

Si dos variables aleatorias son independientes, el valor esperado de su producto cumple

$$E\{xy\} = E\{x\} E\{y\}. \quad (44)$$

La esperanza de una función de dos variables aleatorias  $g(\mathbf{x}, \mathbf{y})$  se define por

$$E\{g(\mathbf{x}, \mathbf{y})\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy. \quad (45)$$

La notación  $f_{\mathbf{x}|\mathbf{y}}(x|y)$  se usa para representar la función de densidad de probabilidad condicional de la variable  $\mathbf{x}$  con respecto a la variable  $\mathbf{y}$ . Otra forma de llamarle es la verosimilitud *a posteriori* de  $\mathbf{x}$  dado que  $\mathbf{y}$  se conoce, es decir, a partir de mediciones de  $\mathbf{y}$  se puede obtener información sobre qué valores de  $\mathbf{x}$  son más probables de ocurrir (Kumar *et al.*, 2005). Formalmente:

$$f_{\mathbf{x}|\mathbf{y}}(x|y) = \frac{f(x, y)}{f_y(y)}. \quad (46)$$

Si  $\mathbf{x}$  y  $\mathbf{y}$  son conjuntamente Gaussianas, también lo son sus PDFs condicionales.

### II.3.5. Procesos estocásticos

A veces es conveniente considerar una imagen como una muestra o *realización* de un proceso estocástico (Pratt, 2007). Un proceso estocástico (proceso aleatorio, campo

aleatorio) es una generalización del concepto de variable aleatoria, pues es una sucesión de éstas. Para imágenes continuas la función de imagen  $I(x, y, t)$  se supone miembro de un proceso estocástico continuo de tres dimensiones con coordenadas espaciales  $(x, y)$  y la coordenada de tiempo  $t$ . En caso de que las imágenes no varíen con el tiempo, se denota la función de imagen como  $I(x, y)$ .

El proceso estocástico  $I(x, y)$  se puede describir completamente por su función de probabilidad conjunta

$$f(I_1, I_2, \dots, I_j; x_1, y_1, x_2, y_2, \dots, x_j, y_j), \quad (47)$$

para todos los puntos de muestra  $j$  donde  $(x_j, y_j)$  representa la muestra espacial de la función imagen  $I_j(x_j, y_j)$ . En general, las densidades de probabilidad de alto orden de las imágenes no se conocen o no pueden ser modeladas fácilmente. La densidad de probabilidad de primer orden puede a veces modelarse exitosamente con base en la física del proceso, como es el caso del ruido aleatorio de un sensor electrónico, que comúnmente se modela mediante una función de densidad de probabilidad Gaussiana (Pratt, 2007):

$$f_I(x, y) = \frac{1}{\sqrt{2\pi}\sigma_I(x, y)} \exp\left\{-\frac{[I(x, y) - \mu_I(x, y)]^2}{2\sigma_I^2(x, y)}\right\}, \quad (48)$$

donde los parámetros  $\mu_I(x, y)$  y  $\sigma_I^2(x, y)$  denotan la media y varianza del proceso en la coordenada  $(x, y)$ .

La *función de autocorrelación* (ACF<sup>4</sup>) de un proceso estocástico es una caracterización de segundo orden y se define como

$$R_I(x_1, y_1; x_2, y_2) = E\{I(x_1, y_1) I(x_2, y_2)\}. \quad (49)$$

---

<sup>4</sup>Autocorrelation Function

Cabe destacar que la ACF es una función simétrica, es decir

$$R_I(x_1, y_1; x_2, y_2) = E\{I(x_1, y_1)I(x_2, y_2)\} = E\{I(x_2, y_2)I(x_1, y_1)\} = R_I(x_2, y_2; x_1, y_1). \quad (50)$$

Otra caracterización de segundo orden es la *autocovarianza* del proceso, definida de la siguiente forma:

$$\begin{aligned} C_I(x_1, y_1; x_2, y_2) &= E\{[I(x_1, y_1) - \mu_I(x_1, y_1)][I(x_2, y_2) - \mu_I(x_2, y_2)]\}, \quad (51) \\ &= R_I(x_1, y_1; x_2, y_2) - \mu_I(x_1, y_1)\mu_I(x_2, y_2). \end{aligned}$$

Para proceso de media cero, la autocorrelación y la autocovarianza coinciden.

### II.3.6. Procesos estacionarios

Un proceso estocástico se dice *estacionario en el sentido estricto* si sus momentos no cambian por desplazamientos en los orígenes espaciales. El proceso se dice que es *estacionario en el sentido amplio* si su media es constante y su autocorrelación depende sólo de las diferencias entre las coordenadas en la imagen  $(x_1 - x_2, y_1 - y_2)$ , y no de los valores particulares de  $(x_1, y_1)$  y  $(x_2, y_2)$ , es decir

$$R_I(x_1, y_1; x_2, y_2) = R_I(x_1 - x_2, y_1 - y_2). \quad (52)$$

Un proceso estacionario en sentido estricto es estacionario en sentido amplio, pero el converso no necesariamente es cierto. La excepción es en el caso de un proceso Gaussiano, que si es estacionario en sentido amplio, también lo es en sentido estricto.

Para un proceso de media cero estacionario en el sentido amplio, su ACF en el desplazamiento cero es igual a su varianza, es decir  $R_I(0, 0) = \sigma_I^2$ , donde  $\sigma_I$  no depende de las coordenadas.

### II.3.7. Densidad espectral

La densidad espectral de un proceso de imagen estacionario en el sentido amplio, se define como la transformada de Fourier de la función de autocorrelación (Kumar *et al.*, 2005), es decir:

$$S_I(\omega) = \int_{-\infty}^{\infty} R_I(x) e^{-i\omega x} dx. \quad (53)$$

Dado que la transformada de Fourier es invertible, tanto la función de densidad espectral como la función de autocorrelación son caracterizaciones equivalentes de segundo orden de un proceso estocástico estacionario en el sentido amplio. Como la autocorrelación es una función real y par, la densidad espectral también es real y par.

### II.3.8. Ergodicidad

Otro concepto importante relacionado con procesos estocásticos es el de ergodicidad. Un proceso estacionario se dice *ergódico* si el promedio espacial es igual al promedio del ensamble. Esto es de suma utilidad porque en la mayoría de los casos, no se cuenta con la colección completa de todos los posibles resultados del proceso. Entonces si un proceso es ergódico, se puede calcular su media y su correlación partir de una sola muestra (Pratt, 2007; Kumar *et al.*, 2005).

## II.4. Optimización

### II.4.1. Desigualdad de Cauchy-Schwarz

La desigualdad de Cauchy-Schwarz establece lo siguiente:

$$\left| \int_{-\infty}^{\infty} f(x) g^*(x) dx \right|^2 \leq \left[ \int_{-\infty}^{\infty} |f(x)|^2 dx \right] \left[ \int_{-\infty}^{\infty} |g(x)|^2 dx \right], \quad (54)$$

donde la igualdad se cumple si y sólo si  $f(x) = \alpha g^*(x)$ , siendo  $\alpha$  una constante arbitraria.

## II.4.2. Multiplicadores de Lagrange

El método de multiplicadores de Lagrange es útil para minimizar una función cuadrática sujeta a un conjunto de restricciones. Para el caso de dos variables, el problema de optimización se puede formular como

$$\begin{aligned} &\text{maximizar } f(x, y) \\ &\text{sujeto a } g(x, y) = c. \end{aligned}$$

Se introduce una nueva variable  $\lambda$  y se forma el funcional

$$\Lambda(x, y, \lambda) = f(x, y) + \lambda(g(x, y) - c). \quad (55)$$

Las soluciones del problema de optimización ocurren en los puntos  $(x, y)$  en los que se satisface la siguiente ecuación:

$$\nabla_{x,y,\lambda} \Lambda(x, y, \lambda) = 0. \quad (56)$$

## II.5. Resumen

En este capítulo se presentan los fundamentos matemáticos para la representación de imágenes y filtrado lineal. Se presentan las propiedades básicas de la transformada de Fourier y su aplicación para filtrado lineal eficiente. Se presentan los conceptos de teoría de probabilidad y procesos estocásticos útiles para la representación estocástica de imágenes. Finalmente se presentan las herramientas de optimización que se utilizan

comúnmente en la derivación de filtros de correlación.

# Capítulo III

---

## Teoría de detección

---

Los filtros de correlación comúnmente se diseñan optimizando una métrica de desempeño para un modelo de escena dado. La optimización se puede realizar en forma analítica (Kumar *et al.*, 2005) o en forma iterativa (Díaz-Ramírez y Kober, 2007; Ramos-Michel y Kober, 2008; González-Fraga *et al.*, 2006). En este capítulo se presentan los modelo de escena y las métricas de desempeño más utilizadas en el diseño de filtros de correlación.

### III.1. Modelos de escena

La exactitud de la localización del objeto en la escena depende del uso de un modelo que represente la escena adecuadamente. Hay dos modelos de escena que son los más utilizados para la derivación de filtros de correlación: el modelo aditivo y el modelo disjunto.

#### III.1.1. Modelo aditivo (traslapado)

En este modelo, la escena de entrada  $s(x)$  contiene un objeto de interés  $t(x)$ , centrado en la posición desconocida  $x_s$  y degradado por ruido aditivo  $n_s(x)$ :

$$s(x) = t(x - x_s) + n_s(x). \quad (57)$$

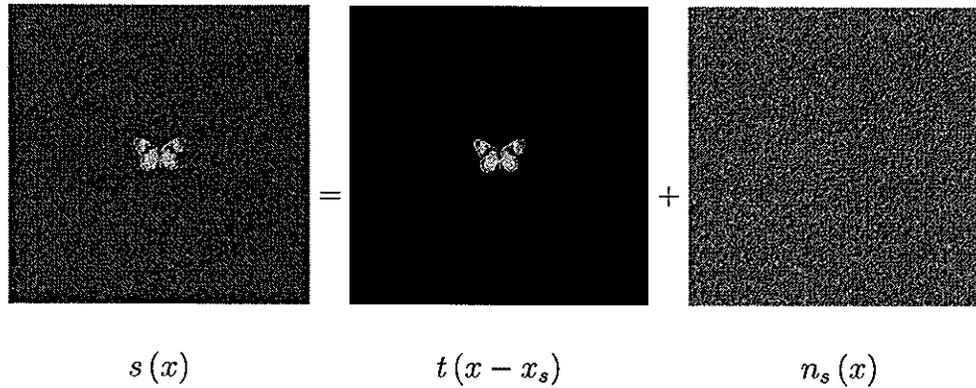


Figura 2: Ejemplo de escena traslapada

En la figura 2 se ilustra un ejemplo de escena aditiva donde se puede observar que la imagen del objeto de interés se encuentra traslapada con el ruido, es decir, en un pixel de la escena donde está presente el objeto contiene información tanto del objeto como del ruido. El modelo aditivo se puede utilizar para representar señales de audio o para representar imágenes de infrarrojo (Milton *et al.*, 1985).

### III.1.2. Modelo disjunto (no traslapado)

Para imágenes de intensidad, como las capturadas con cámaras comunes, se utiliza el modelo disjunto o no traslapado. En este modelo, se considera que un objeto opaco se encuentra sobre un fondo espacialmente disjunto y que toda la escena se corrompe con ruido aditivo debido a las imperfecciones del dispositivo de captura. Formalmente, el modelo se define como sigue:

$$s(x) = t(x - x_s) + b_s(x) \bar{w}(x - x_s) + n_s(x), \quad (58)$$

donde  $t(x - x_s)$ , y  $n_s(x)$  son el objeto de interés y el ruido aditivo como en el modelo traslapado,  $b_s(x)$  representa el fondo disjunto y  $\bar{w}(x)$  es la región inversa de soporte del

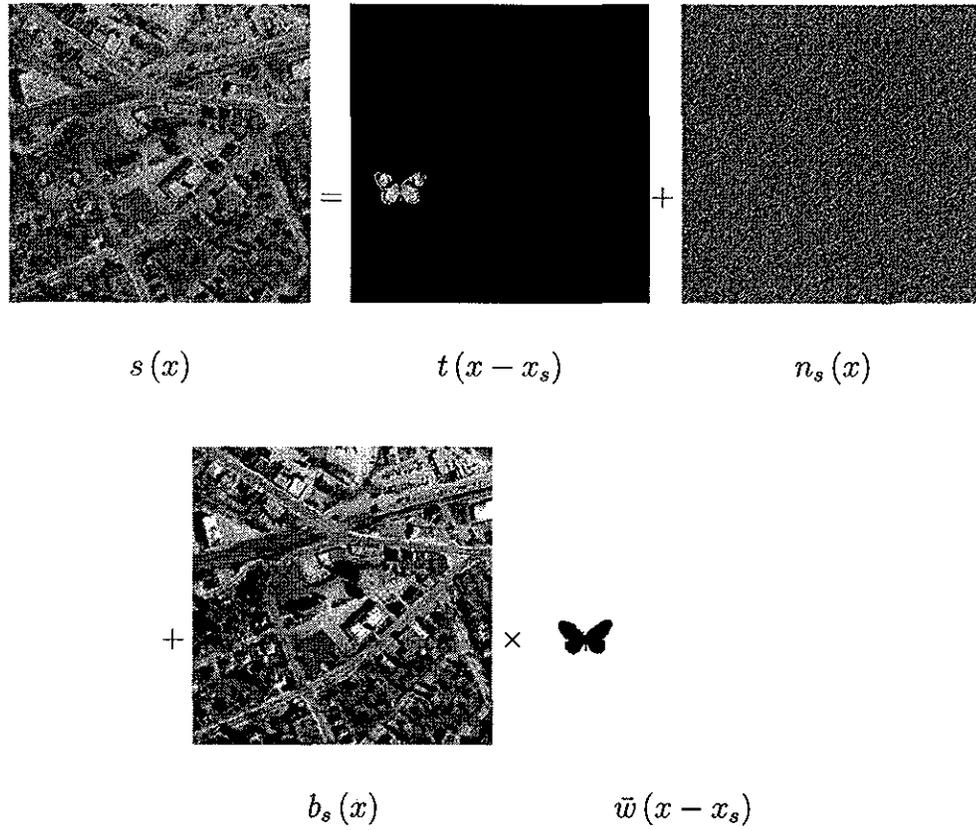


Figura 3: Ejemplo de escena no traslapada

objeto. La región de soporte del objeto,  $w(x)$  se define como 1 en las coordenadas de la señal en las que el objeto está presente y cero en caso contrario. La región inversa de soporte se define como  $\bar{w}(x) = 1 - w(x)$ , es decir, toma un valor de cero en presencia del objeto y uno en caso contrario. En la figura 3 se muestra un ejemplo de escena real como se representa por el modelo no traslapado. Para facilitar el tratamiento matemático, el fondo en una escena disjunta se modela como realización de un proceso estocástico estacionario (con valor esperado  $\mu_s$ ).

## III.2. Métricas de desempeño

### III.2.1. Razón señal a ruido (SNR<sup>1</sup>)

Caracteriza la variación del pico de correlación deseado en la salida del filtro con respecto al ruido en la entrada. Se define formalmente como

$$\text{SNR} = \frac{|\text{E}\{y(x_s)\}|^2}{\text{Var}\{y(x_s)\}}, \quad (59)$$

donde  $y(x)$  denota la salida del filtro; en particular,  $y(x_s)$  denota la salida del filtro en la posición del objeto en la escena. Mayores valores de SNR indican una mayor tolerancia al ruido y por lo tanto una menor probabilidad de error en el proceso de detección. La principal desventaja de esta métrica es que sólo considera el efecto que tiene el ruido sobre pico de la salida de correlación y no sobre el resto del plano de correlación. En el caso del modelo disjunto, se utiliza una variante de SNR definida como la razón del pico a la varianza promedio de la salida de correlación (SÑR):

$$\text{SÑR} = \frac{|\text{E}\{y(x_s)\}|^2}{\text{Var}\{y(x)\}}. \quad (60)$$

### III.2.2. Razón pico a energía de correlación (POE<sup>2</sup>)

Para minimizar la probabilidad de falsas alarmas, es conveniente que en la salida del filtro las regiones del objeto y del fondo sean los más diferentes posible. Para lograr esta seaparación, se propuso el criterio POE (Javidi y Wang, 1994a), definido formalmente como

$$\text{POE} = \frac{|\text{E}\{y(x_s)\}|^2}{\text{E}\{|y(x)|^2\}}, \quad (61)$$

---

<sup>1</sup>Signal-to-noise ratio

<sup>2</sup>Peak-to-output energy ratio

donde la sobrebarra indica promedio espacial. El denominador del criterio POE denota la energía promedio de la salida del filtro. Un filtro obtenido al optimizar el criterio POE produce un pico intenso de correlación en la posición del objeto en la escena y valores cercanos a cero en la región del fondo.

### III.2.3. Error cuadrático medio (MSE<sup>3</sup>)

Una manera alternativa de diseñar filtros de correlación consiste en especificar explícitamente la salida deseada y diseñar un filtro que minimice el error cuadrático promedio sobre un ensamble de imágenes. La salida de correlación ideal es un pico de muy alta intensidad en la localización del objeto sobre un fondo plano, lo que se modela analíticamente como una función delta de Dirac:

$$y^i(x) = \delta(x - x_s), \quad (62)$$

donde  $y^i(x)$  denota la salida ideal. La diferencia entre la salida ideal y la salida real del filtro está dada por

$$\text{MSE} = \text{E} \left\{ \frac{1}{\mathcal{L}} \int_{-\infty}^{\infty} |y^i(x) - y(x)|^2 dx \right\}, \quad (63)$$

donde  $\mathcal{L}$  es la duración espacial de la salida.

Los filtros obtenidos mediante la optimización del criterio MSE tienden a ser de la misma forma que los filtros derivados del criterio POE (Réfrégier *et al.*, 1993; Javidi *et al.*, 1996).

---

<sup>3</sup>Mean square error

### III.2.4. Capacidad de discriminación (DC<sup>4</sup>)

Indica la capacidad del filtro de discriminar entre el objeto deseado y estructuras similares en la escena de entrada

$$DC = 1 - \left| \frac{y_{\text{máx}}^b}{y_{\text{máx}}^t} \right|^2, \quad (64)$$

donde  $y_{\text{máx}}^b$  y  $y_{\text{máx}}^t$  denotan los valores de mayor intensidad en las regiones del fondo y del objeto, respectivamente. Las regiones del fondo y del objeto son complementarias y regularmente se considera la región del objeto como una región rectangular de dimensiones similares a la región de soporte del objeto. Valores de la DC cercanos a uno indican una buena capacidad del filtro de discriminar objeto no deseados de los objetos de la clase verdadera. Si el reconocimiento falla, es decir, ocurre un error anómalo, el valor de la DC pasa a ser negativo. Dado que no es sencillo optimizar la DC en forma analítica, se utilizan métodos iterativos para lograrlo (Díaz-Ramírez y Kober, 2007; Ramos-Michel y Kober, 2008; González-Fraga *et al.*, 2006).

### III.2.5. Errores de localización (LE<sup>5</sup>)

Se calcula como la desviación entre la localización estimada del objeto y su localización real en la escena

$$LE = \|x_s - \hat{x}_s\|_2, \quad (65)$$

donde  $\|\cdot\|_2$  indica la distancia euclidiana entre la posición estimada y la posición real.

---

<sup>4</sup>Discrimination capability

<sup>5</sup>Location errors

### III.3. Distorsiones que degradan el desempeño de los filtros de correlación

Las imágenes que se obtienen para mostrar o registrar información útil se ven afectadas por imperfecciones en los procesos de captura, por lo que la imagen obtenida en realidad representa una versión degradada de la escena real. El tratar con estas imperfecciones es de suma importancia para aplicaciones de procesamiento de imágenes, en particular para detección y localización de objetos en imágenes. En el presente trabajo de investigación se considera el diseño de filtros tolerantes a la presencia de ruido aditivo, fondos complejos y distorsiones geométricas.

Dado que los filtros de correlación utilizan la información de la apariencia y forma del objeto, su desempeño se ve afectado por degradaciones que afectan la apariencia del objeto. A continuación se presentan algunas distorsiones que se consideran para el diseño de los filtros propuestos en este trabajo.

#### III.3.1. Modelos de ruido

Una imagen contiene información degradada debido al proceso de captura de la imagen. Se denomina ruido a todas aquellas degradaciones que alteran los niveles de intensidad de la imagen capturada. Dada la naturaleza aleatoria del ruido, este se modela como un proceso estocástico y se caracteriza por sus propiedades estadísticas. La principal fuente de ruido es el proceso de adquisición de imágenes que convierte una señal óptica en una señal eléctrica continua que posteriormente se muestrea para formar la imagen digital. El ruido se debe a las imperfecciones de fabricación de los sensores, a variaciones de los voltajes con los que operan los sensores, y a errores de cuantización

---

(Gonzalez y Woods, 1992).

La visibilidad del ruido en las imágenes depende de varios factores: del tipo de ruido, de la intensidad del ruido, de su grado de correlación y del fondo en la imagen. Generalmente el ruido correlacionado es más notorio que el ruido blanco con la misma varianza.

### Ruido aditivo

Se utilizan distintos tipos de ruido para modelar las degradaciones que pueden afectar a las imágenes. El ruido aditivo de media cero se utiliza para modelar el ruido debido a las imperfecciones de sensores y a los errores de cuantización (Pratt, 2007). Una señal contaminada por ruido aditivo se modela como

$$s(x) = t(x) + n_s(x), \quad (66)$$

donde  $t(x)$  es la señal original sin perturbar y  $s(x)$  es la señal observada contaminada por el ruido  $n_s(x)$ . Cuando las muestras de ruido en cada posición son estadísticamente independientes se denomina *ruido blanco*. Su función de autocorrelación  $R_n(x)$  tiene la forma

$$R_n(x) = \sigma^2 \delta(x), \quad (67)$$

donde  $\sigma^2$  es la varianza y  $\delta(x)$  es la función delta de Dirac. Su densidad espectral, dada por la transformada de Fourier de la función de autocorrelación, es una constante:

$$S_n(\omega) = N_0 \sigma^2,$$

donde  $N_0$  es una constante de normalización. La denominación de ruido blanco se debe a que su densidad espectral es no nula en todas las frecuencias.

La autocorrelación de la señal observada está dada por

$$R_s(x) = R_t(x) + R_n(x). \quad (68)$$

En particular para el desplazamiento cero se cumple que

$$\sigma^2 = R_s(0) - R_t(0).$$

Como generalmente las funciones de autocorrelación de las señales de interés son suaves, se puede estimar la varianza del ruido al calcular la función de autocorrelación de la escena observada y extrapolando el valor de  $R_t(0)$ . La diferencia entre el valor calculado para la escena observada y el valor extrapolado para la señal de interés proporciona la estimación de la varianza (Yaroslavsky, 1993, 2003).

### Ruido multiplicativo

El ruido multiplicativo se utiliza para modelar situaciones en las que la iluminación varía a lo largo de la imagen. El modelo formal es el siguiente:

$$s(x) = t(x)n_m(x), \quad (69)$$

donde  $n_m(x)$  denota el ruido multiplicativo. Este modelo frecuentemente se trata como el modelo aditivo aplicando la operación de logaritmo a la escena de entrada. En el caso de iluminación homogénea,  $n_m(x)$  es una constante y se modela como una variable aleatoria (Javidi y Wang, 1994b).

### Ruido impulsivo

Se caracteriza porque se pierde completamente la información de la señal original. Este tipo de ruido es producto de fallas en los sensores del dispositivo de captura o por

errores en el canal de transmisión. Se modela formalmente como

$$s(x) = (1 - e)t(x) + e \cdot n_i(x), \text{ donde } e = \begin{cases} 1 & \text{con probabilidad } P_e \\ 0 & \text{en caso contrario} \end{cases}. \quad (70)$$

Un caso particular de ruido impulsivo es el ruido sal y pimienta, donde  $n_i(x)$  toma valores extremos de 0 o 255 (Gonzalez y Woods, 1992).

### III.3.2. Distribuciones de ruido

En muchas imágenes el ruido se puede modelar con una distribución Gaussiana, uniforme, sal y pimienta, entre otras. La distribución de un proceso de ruido depende de la naturaleza física de la fuente de ruido. A continuación se describen algunas distribuciones de ruido.

**Ruido Gaussiano:** Se utiliza para modelar procesos naturales de ruido como el ruido de sensores, errores de cuantización, etc.

**Ruido con distribución uniforme:** Se utiliza como base para otras distribuciones y para degradar imágenes que luego se emplean para evaluar algoritmos de restauración.

**Ruido impulsivo:** La probabilidad de que ocurra un error generalmente es pequeña (menor a 0.1). Los valores que puede tomar un impulso de error pueden tener cualquier distribución. En el caso de ruido sal y pimienta los valores sólo pueden ser 0 y 255.

**Ruido con distribución Rayleigh:** Se utiliza para modelar el ruido presente en imágenes digitales que se obtienen mediante un radar.

### III.3.3. Distorsiones geométricas

Una distorsión geométrica es un error sobre la imagen que se determina entre las coordenadas de la imagen actual y las coordenadas de la imagen ideal. Se clasifican en dos tipos: distorsiones internas (debidas a la geometría del sensor) y distorsiones externas (debidas a la posición del sensor o a la forma del objeto).

**Rotación** La distorsión por rotación se debe a cambios de posición del objeto en relación a la posición del dispositivo de captura. Las rotaciones se clasifican en dos tipos: rotaciones en el plano y rotaciones fuera del plano. Los objetos de dos dimensiones sólo se afectan por rotaciones en el plano, mientras que objetos del mundo real pueden presentar rotaciones en más ejes. En el presente trabajo de investigación solo se consideran rotaciones en el plano.

**Escalamiento** La distorsión por escalamiento se presenta cuando cambia la distancia entre el objeto y el sistema de captura, de manera que cambian las dimensiones aparentes del objeto en la escena.

### III.4. Teoría de estimación

En esta sección se presentan las herramientas de teoría de probabilidad que se utilizan como base para la localización de objetos en presencia de ruido. Existen distintos tipos de ruido que afectan el desempeño de las técnicas de reconocimiento basadas en correlación; sin embargo, en este capítulo nos enfocamos exclusivamente a la localización en presencia de ruido aditivo.

---

### III.4.1. Estimación de parámetros

La estimación de parámetros en imágenes es uno de los problemas fundamentales en el reconocimiento de patrones y el procesamiento de imágenes. El problema se formula de la siguiente manera: Dadas muestras observadas de una señal de entrada  $s = \{s_k\}$  que implícitamente dependen de un parámetro  $\theta$ , el problema consiste en determinar el valor de  $\theta$  con la mejor precisión posible. Se supone que la dependencia entre la señal observada y el parámetro  $\theta$  se ve afectada por factores externos tales como ruido aleatorio o el proceso de captura de la señal. El parámetro a estimar puede tomar valores aleatorios. Se puede suponer que se conoce la distribución de probabilidad de los valores de  $\theta$ . La probabilidad se denota por  $P(\theta)$  en el caso de que  $\theta$  tome valores discretos y en el caso de valores continuos se utiliza la función de densidad de probabilidad  $p(\theta)$ . Debido a la presencia de factores aleatorios en la formulación del problema, la calidad de la estimación sólo se puede evaluar en términos estadísticos. Esta evaluación depende de la probabilidad de error (para parámetros discretos) o de la densidad de probabilidad de la estimación del error  $p(\theta - \hat{\theta})$ , donde  $\hat{\theta}$  denota la estimación del parámetro  $\theta$ . Se conoce que el estimador que tiene menor probabilidad de error es aquél que tiene máxima probabilidad *a posteriori*  $P(\theta|s)$  (Duda *et al.*, 2000), es decir, la probabilidad de que el parámetro tome el valor  $\theta$  dada la señal observada  $s$ . Esto se expresa como

$$\hat{\theta} = \underbrace{\arg \max}_{\theta} P(\theta|s). \quad (71)$$

El estimador de máxima probabilidad a posterior se conoce como estimador MAP<sup>6</sup>. Dada la señal observada, el dispositivo óptimo para la estimación de  $\theta$  debe ser capaz de medir la probabilidad *a posteriori* de los valores del parámetro y de determinar el

---

<sup>6</sup>Maximum *a posteriori* probability

valor que maximiza esa probabilidad. Esto significa que se debe especificar en forma precisa la dependencia de la señal observada sobre el parámetro  $\theta$ .

### Estimación en presencia de ruido aditivo blanco Gaussiano

El modelo más sencillo para la representación de señales es el modelo aditivo blanco Gaussiano. Se considera que la señal observada contiene a la señal de interés contaminada por ruido estadísticamente independiente de la señal. Además se supone que el ruido tiene una distribución Gaussiana y las distintas muestras de ruido son independientes entre sí. El modelo se representa como

$$\{s_k = t_k(\theta) + n_k\}, \quad (72)$$

que es la forma discreta de representar el modelo aditivo que se explica en la sección III.1. Los valores de la señal de interés  $t_k(\theta)$  dependen del parámetro  $\theta$  y  $\{n_k\}$  son las muestras de ruido en cada punto de la señal observada. El ruido se considera de media cero y desviación estándar  $\sigma$ , por lo que su densidad de probabilidad está dada por

$$p_n(\nu) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{\nu^2}{2\sigma^2}\right]. \quad (73)$$

Para un valor de  $\theta$  dado, se puede estimar la probabilidad de la observación como

$$P(s|\theta) = P(\{n_k = s_k - t_k(\theta)\}), \quad (74)$$

y la probabilidad *a posteriori* se puede calcular a partir del teorema de Bayes:

$$P(\theta|\{s_k\}) = \frac{P(\{s_k\}|\theta)p(\theta)}{p(\{s_k\})}. \quad (75)$$

Dado que  $p(\{s_k\})$  no depende de  $\theta$ , no contribuye a discriminar entre qué valores de  $\theta$  son más probables que otros. De manera que se puede expresar la probabilidad *a*

*posteriori* como:

$$P(\theta | \{s_k\}) \cong P(\{s_k\} | \theta) p(\theta) = P(\{n_k = s_k - t_k(\theta)\}) p(\theta). \quad (76)$$

Por la suposición de que el ruido es estadísticamente independiente, la probabilidad conjunta de  $N$  muestras de la observación se puede calcular como un producto. Sustituyendo la distribución de probabilidad de la ecuación (73) en la ecuación (76) se obtiene

$$\begin{aligned} P(\theta | \{s_k\}) &\cong \prod_{k=0}^{N-1} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2} (s_k - t_k(\theta))^2\right] p(\theta), \\ &= (2\pi\sigma^2)^{-N/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{k=0}^{N-1} (s_k - t_k(\theta))^2\right] p(\theta). \end{aligned} \quad (77)$$

El factor  $(2\pi\sigma^2)^{-N/2}$  es común para todos los valores de  $\theta$ , por lo que se puede descartar sin pérdida de optimización. Desarrollando el binomio en el interior de la sumatoria se obtiene

$$P(\theta | \{s_k\}) \cong \exp\left[-\frac{1}{2\sigma^2} \sum_{k=0}^{N-1} s_k + \frac{1}{\sigma^2} \sum_{k=0}^{N-1} s_k t_k(\theta) - \frac{1}{2\sigma^2} \sum_{k=0}^{N-1} t_k^2(\theta) + \ln p(\theta)\right] \quad (78)$$

Nuevamente el primer término no depende de  $\theta$ , por lo que se puede eliminar sin pérdida de optimización. Por la monotonidad de la función logaritmo, se puede calcular el logaritmo a la ecuación (78) sin cambiar el valor de  $\theta$  que la optimiza. De esta manera, el estimador MAP se obtiene al maximizar el logaritmo de  $P(\theta | \{s_k\})$ , es decir

$$\begin{aligned} \hat{\theta}_{\text{MAP}}^{\text{opt}} &= \underbrace{\arg \max}_{\theta} \{\ln P(\theta | \{s_k\})\}, \\ &= \underbrace{\arg \max}_{\theta} \left\{ \frac{1}{\sigma^2} \sum_{k=0}^{N-1} s_k t_k(\theta) - \frac{1}{2\sigma^2} \sum_{k=0}^{N-1} t_k^2(\theta) + \ln p(\theta) \right\}. \end{aligned} \quad (79)$$

Cuando el parámetro  $\theta$  se supone uniformemente distribuido, el último término en

la ecuación (79) no afecta la determinación del valor de  $\theta$  y se puede descartar. Esto se conoce como el estimador de máxima verosimilitud (o estimador ML<sup>7</sup>), es decir

$$\hat{\theta}_{\text{ML}}^{\text{opt}} = \underbrace{\arg \max}_{\theta} \left\{ \frac{1}{\sigma^2} \sum_{k=0}^{N-1} s_k t_k(\theta) - \frac{1}{2\sigma^2} \sum_{k=0}^{N-1} t_k^2(\theta) \right\}. \quad (80)$$

El estimador ML también se utiliza cuando no se conoce la distribución *a priori* del parámetro  $\theta$ .

Para la estimación de la ubicación de un objeto en ruido aditivo blanco Gaussiano se utiliza el estimador ML donde el parámetro a estimar es la localización. El objetivo es determinar la coordenada  $x_s$  en la que se ubica el objeto. En primer lugar, se observa que la energía del objeto es independiente de la ubicación del mismo, por lo que el segundo término de la ecuación (80) se puede descartar sin pérdida de optimización. El estimador resultante es

$$(\hat{x}_s)_{\text{ML}}^{\text{opt}} = \underbrace{\arg \max}_{x_s} \left\{ \frac{1}{\sigma^2} \sum_{k=0}^{N-1} s_k t_k(x_s) \right\}. \quad (81)$$

Esto es la *correlación discreta* entre la señal observada y la señal de interés con un desplazamiento  $x_s$ . La correlación para encontrar el estimador óptimo  $\hat{x}_s$  se puede calcular como un filtro lineal invariante a desplazamiento con respuesta de impulso  $t(x)$ . En el dominio de frecuencia esto se expresa como

$$H_{\text{opt}}(\omega) = T^*(\omega). \quad (82)$$

Este filtro se conoce como el filtro de correspondencia. En el capítulo IV se muestra que este filtro también maximiza el criterio SNR.

---

<sup>7</sup>Maximum Likelihood

### III.5. Resumen

En este capítulo se presentan los modelos de escena que se utilizan en este trabajo de tesis, así como las métricas de desempeño comúnmente utilizadas en el diseño de filtros de correlación. Se presentan también los distintos factores que afectan la calidad del reconocimiento con filtros de correlación como el ruido y distorsiones geométricas.

Finalmente, se presenta la detección y localización de objetos mediante estimadores de máxima probabilidad *a posteriori* y de máxima verosimilitud. Se muestra que el filtro óptimo para localizar objetos en presencia de ruido aditivo blanco Gaussiano es el filtro de correspondencia.

# Capítulo IV

---

## Filtros clásicos de correlación

---

En 1964 VanderLugt propuso un método óptico para el reconocimiento de objetos utilizando técnicas de correlación. Aunque la técnica presenta algunas deficiencias, la simplicidad del proceso propuesto ha motivado que este trabajo sirva como base para el diseño de diversos tipos de filtros basados en correlación. Muchas de las mejoras propuestas buscan incrementar la eficiencia de luz e incorporar tolerancia a distorsiones y distintos tipos de ruido.

Generalmente, los filtros se diseñan optimizando algún criterio de desempeño. Un filtro puede utilizarse exitosamente en casos que se apeguen al modelo utilizado para su diseño. En la literatura se puede encontrar un gran número de filtros basados en correlación; sin embargo, en este capítulo se describen los más importantes que sirvieron como base para el desarrollo de esta tesis.

### IV.1. Filtros para el modelo aditivo

En esta sección se presenta la derivación analítica de los principales filtros para el modelo traslapado de escena.

#### IV.1.1. Filtro de correspondencia

El filtro de correspondencia (MF<sup>1</sup>) propuesto por VanderLugt (1964) optimiza el criterio SNR para el modelo aditivo. La optimización del criterio SNR requiere de

---

<sup>1</sup>Matched Filter

maximizar el numerador, que indica la intensidad esperada del pico de correlación en presencia del objeto, y minimizar el denominador que corresponde a la varianza del pico causada por el ruido. Bajo la suposición de que el ruido aditivo tiene media cero el valor del pico de correlación está dado por

$$E \{y(x_s)\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} T(\omega) H(\omega) d\omega. \quad (83)$$

La varianza del pico de correlación se calcula a partir de la densidad espectral del ruido después de ser procesado con el filtro:

$$\text{Var} \{y(x_s)\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} N_s(\omega) |H(\omega)|^2 d\omega. \quad (84)$$

Sustituyendo las ecuaciones (83) y (84) en la definición de la SNR (ecuación (59)) se obtiene

$$\text{SNR} = \frac{\left| \frac{1}{2\pi} \int_{-\infty}^{\infty} T(\omega) H(\omega) d\omega \right|^2}{\frac{1}{2\pi} \int_{-\infty}^{\infty} N_s(\omega) |H(\omega)|^2 d\omega}. \quad (85)$$

La optimización de la SNR se hace aplicando la desigualdad Cauchy-Schwarz. Para esto, se reescribe la ecuación (85) como

$$\text{SNR} = \frac{\left| \int_{-\infty}^{\infty} \frac{T(\omega)}{\sqrt{N_s(\omega)}} \left[ H(\omega) \sqrt{N_s(\omega)} \right] d\omega \right|^2}{2\pi \int_{-\infty}^{\infty} N_s(\omega) |H(\omega)|^2 d\omega}. \quad (86)$$

Es posible reescribir la función de esta manera dado que la densidad espectral del ruido es no negativa. Aplicando la desigualdad se obtiene

$$\text{SNR} \leq \frac{\left[ \int_{-\infty}^{\infty} \frac{|T(\omega)|^2}{N_s(\omega)} d\omega \right] \left[ \int_{-\infty}^{\infty} |H(\omega)|^2 N_s(\omega) d\omega \right]}{2\pi \int_{-\infty}^{\infty} N_s(\omega) |H(\omega)|^2 d\omega}, \quad (87)$$

donde la igualdad se cumple cuando

$$H(\omega) \sqrt{N_s(\omega)} = \alpha \frac{T^*(\omega)}{\sqrt{N_s(\omega)}}, \quad (88)$$

con  $\alpha$  una constante arbitraria. Fijando el valor de  $\alpha = 1$  y despejando se obtiene la respuesta de frecuencia del filtro de correspondencia:

$$H(\omega) = \frac{T^*(\omega)}{N_s(\omega)}. \quad (89)$$

En el caso de ruido blanco, la densidad espectral de potencia  $N_s(\omega)$  es una constante; la respuesta de impulso del filtro es una versión invertida del objeto de interés  $t(-x)$ . De aquí el nombre de filtro de correspondencia.

#### IV.1.2. Filtro óptimo

El filtro óptimo (OF<sup>2</sup>) propuesto por Yaroslavsky (1993) para minimizar la probabilidad de falsas alarmas para el modelo de escena aditivo. El filtro se obtiene al optimizar el criterio POE. El numerador del criterio POE se calcula como en la ecuación (83). La energía promedio en la salida de correlación se calcula como

$$E \left\{ \overline{|y(x)|^2} \right\} = \frac{1}{\mathcal{L}} \int_{-\infty}^{\infty} |y(x)|^2 dx, \quad (90)$$

donde  $\mathcal{L}$  es la extensión espacial de la salida. Por el teorema de Parseval, la energía de salida se puede calcular en el dominio de frecuencia como

$$E \left\{ \overline{|y(x)|^2} \right\} = \frac{1}{2\pi\mathcal{L}} \int_{-\infty}^{\infty} E \{ |S(\omega)|^2 \} |H(\omega)|^2 d\omega. \quad (91)$$

Sustituyendo las ecuaciones (83) y (91) en la definición del criterio POE (ver ecuación (61)), se obtiene

$$\text{POE} = \frac{\left| \frac{1}{2\pi} \int_{-\infty}^{\infty} T(\omega) H(\omega) d\omega \right|^2}{\frac{1}{2\pi\mathcal{L}} \int_{-\infty}^{\infty} E \{ |S(\omega)|^2 \} |H(\omega)|^2 d\omega}. \quad (92)$$

Como el ruido se considera de media cero, el valor esperado de la densidad espectral

---

<sup>2</sup>Optimum Filter

de la escena está dado por

$$E \{ |S(\omega)|^2 \} = |T(\omega)|^2 + N_s(\omega). \quad (93)$$

La POE se puede reescribir de la forma

$$\text{POE} = \frac{\mathcal{L} \left| \int_{-\infty}^{\infty} \frac{T(\omega)}{\sqrt{|T(\omega)|^2 + N_s(\omega)}} \left[ H(\omega) \sqrt{|T(\omega)|^2 + N_s(\omega)} \right] d\omega \right|^2}{2\pi \int_{-\infty}^{\infty} [ |T(\omega)|^2 + N_s(\omega) ] |H(\omega)|^2 d\omega}. \quad (94)$$

Esta operación está bien definida dado que la densidad espectral de la escena es no negativa. Aplicando la desigualdad Cauchy-Schwarz, el filtro que maximiza el valor del criterio POE es

$$H(\omega) = \alpha \frac{T^*(\omega)}{|T(\omega)|^2 + N_s(\omega)}, \quad (95)$$

donde  $\alpha$  es una constante arbitraria.

### IV.1.3. Filtro de sólo fase

Para la aplicación de filtros de correlación en sistemas ópticos, un criterio de desempeño importante es el de la eficiencia de luz. Este criterio se define como (Kumar y Hassebrook, 1990):

$$\eta = \frac{\text{Energía de salida}}{\text{Energía de entrada}} = \frac{\int |y(x)|^2 dx}{\int |s(x)|^2 dx}. \quad (96)$$

Por el teorema de Parseval, la ecuación (96) se puede reescribir en el dominio de frecuencia como

$$\eta = \frac{\int |S(\omega) H(\omega)|^2 d\omega}{\int |S(\omega)|^2 d\omega}. \quad (97)$$

Este valor es máximo cuando  $|H(\omega)| = 1$ , por lo que el filtro resultante es (Horner y Gianino, 1984):

$$H(\omega) = \frac{T^*(\omega)}{|T(\omega)|}. \quad (98)$$

Este filtro se conoce como el filtro de solo fase (POF<sup>3</sup>) y dado que su magnitud es igual a uno no absorbe luz. Aunque maximiza la eficiencia de luz, el hecho de que deje pasar todas las frecuencias conlleva a que sea sumamente sensible a la presencia de ruido en la escena.

## IV.2. Filtros para el modelo no traslapado

El MF está diseñado para detectar objetos en escenas con fondo traslapado y su desempeño, así como el de otros filtros diseñados para ese modelo, se degrada significativamente en presencia de ruido no traslapado (Javidi y Wang, 1992). La importancia del modelo no traslapado radica en que se utiliza para representar imágenes del mundo real en las que objetos opacos se encuentran incrustados en fondos complejos. Se han propuesto diversos filtros para el modelo no traslapado (Javidi y Wang, 1994a,b; Javidi *et al.*, 1996; Kober y Campos, 1996), de los cuales, en esta sección se presentan los propuestos por Javidi y Wang (1994a): el filtro generalizado de correspondencia y el filtro generalizado óptimo.

### IV.2.1. Filtro de correspondencia generalizado

El filtro generalizado de correspondencia (GMF<sup>4</sup>) fue propuesto por Javidi y Wang (1994a) para optimizar el criterio SÑR para el modelo disjuncto. La diferencia entre el

---

<sup>3</sup>Phase-only filter

<sup>4</sup>Generalized Matched Filter

criterio SNR y el criterio SÑR se debe a que el fondo no traslapado hace que el ruido en la escena no pueda ser modelado como proceso estacionario. Dada la naturaleza no estacionaria del ruido, la varianza en la región del pico de correlación no se puede suponer igual a la varianza fuera del área del objeto. Para compensar esto, se modela la posición del objeto como una variable aleatoria uniformemente distribuida, y se calcula la varianza promedio sobre el ensamble de imágenes que considera las distintas posibles ubicaciones del objeto. Al optimizar respecto al desempeño promedio en el ensamble, el ruido se puede tratar nuevamente como estacionario y se aplican técnicas de optimización similares a las usadas para el modelo de escena traslapada.

El denominador del criterio SÑR está dado por  $\overline{\text{Var}\{y(x)\}}$ , opuesto al denominador  $\text{Var}\{y(x_s)\}$  para el criterio SNR. De esta manera, se considera la varianza promedio de la salida del filtro y no solo la varianza en la ubicación del objeto.

La respuesta de frecuencia del GMF está dada por (Javidi y Wang, 1994a):

$$\text{GMF} = \frac{[T(\omega) + \mu_s \bar{W}(\omega)]^*}{\frac{1}{2\pi} B_s^0(\omega) \bullet |\bar{W}(\omega)|^2 + N_s(\omega)}. \quad (99)$$

Se puede observar que el filtro está diseñado para realizar la correspondencia con un nuevo objeto. El nuevo objeto está compuesto por el objeto original a detectar más la región inversa de soporte del objeto pesada con el valor promedio del fondo. Es decir, para el modelo no traslapado el fondo es tan importante como el objeto para realizar adecuadamente la correspondencia.

Además de optimizar el criterio SÑR, Kober y Campos (1996) demostraron que el GMF es óptimo en el sentido de minimizar la varianza de errores de localización. La principal desventaja del GMF es que, aunque produce planos de correlación de mínima varianza, el pico de correlación puede perderse en un fondo de muy alto promedio (Javidi y Wang, 1994a). Debido a esto, el filtro GMF presenta bajo desempeño en términos del

criterio DC.

### IV.2.2. Filtro óptimo generalizado

El filtro óptimo generalizado (GOF<sup>5</sup>) optimiza el criterio POE para el modelo de escena disjunto. Maximiza la intensidad del pico de correlación en la localización del objeto de interés, mientras minimiza la energía promedio del plano de correlación. La respuesta de frecuencia del GOF es la siguiente (Javidi y Wang, 1994a):

$$\text{GOF} = \frac{[T(\omega) + \mu_s \bar{W}(\omega)]^*}{|T(\omega) + \mu_s \bar{W}(\omega)|^2 + \frac{1}{2\pi} B_s^0(\omega) \bullet |\bar{W}(\omega)|^2 + N_s(\omega)}. \quad (100)$$

Cuando no hay un fondo disjunto, la respuesta de frecuencia en la ecuación (100) coincide exactamente con la respuesta de frecuencia del OF.

## IV.3. Filtros compuestos

Los filtros de correlación se diseñan utilizando información de la apariencia y de la forma del objeto de interés. Por este motivo, el desempeño decrece rápidamente cuando la apariencia del objeto en la escena de entrada no coincide con la apariencia conocida al momento de realizar el diseño del filtro. Para considerar cambios de apariencia, se introdujeron los filtros compuestos de correlación. La idea consiste en diseñar un filtro que pueda tomar en cuenta distintas vistas del objeto que se proporcionan en un conjunto de imágenes de entrenamiento. De manera opcional, se pueden incluir imágenes en el conjunto que se desea que el filtro rechace. A continuación se describen las tres técnicas más comunes para el diseño de filtros compuestos.

---

<sup>5</sup>Generalized Optimum Filter

### IV.3.1. Funciones discriminantes sintéticas

La técnica de funciones discriminantes sintéticas (SDF<sup>6</sup>) propuesta por Casasent (1984), consiste en diseñar un filtro compuesto como combinación lineal de las imágenes que constituyen el conjunto de entrenamiento. Los pesos en la combinación lineal asignados a cada imagen se escogen de manera que la salida de correlación en la coordenada del objeto tome un valor deseado. Típicamente se escogen valores de uno para las imágenes del objeto de interés y cero para imágenes de objetos a rechazar. Para cubrir el espacio de apariencias posibles del objeto de interés, es necesario tener un conjunto de imágenes suficientemente completo. Aunque no hay una regla para escoger la cantidad ideal de imágenes en el conjunto, se han propuesto algunas estrategias como se menciona en (Kumar, 1992).

Es importante señalar que el número de imágenes de entrenamiento afecta la calidad de la salida del filtro. Esto fue analizado por Kumar y Pochapsky (1986), quienes concluyeron que la SNR en el peor caso se puede mejorar con un mayor número de imágenes de entrenamiento, pero al precio de empeorar la SNR del mejor caso.

La principal desventaja del enfoque SDF es que sólo se tiene control sobre el pico de correlación y no sobre el plano completo, por lo que pueden aparecer lóbulos laterales de mayor intensidad que en la región del objeto causando falsas alarmas.

El diseño de filtros SDF se detalla a continuación. Sea  $\mathbf{S}$  una matriz de tamaño  $d \times n$ , donde  $n$  es la cantidad de imágenes del conjunto de entrenamiento y  $d$  es el número de pixeles de las imágenes. Las columnas de  $\mathbf{S}$  son las imágenes de entrenamiento en forma de vector. Para imágenes de dos dimensiones, cada vector se forma tomando los elementos de la imagen de izquierda a derecha en cada renglón y los renglones de arriba hacia abajo. Sea  $\mathbf{a}$  el vector de pesos que se le asignan a cada imagen y sea  $\mathbf{u}$  el vector

---

<sup>6</sup>Synthetic Discriminant Functions

de salidas deseadas para cada filtro. El filtro en forma de vector se denota como  $\mathbf{h}$  y se calcula como

$$\mathbf{h} = \mathbf{S}\mathbf{a}. \quad (101)$$

Las restricciones de salida se expresan como sigue:

$$\mathbf{u} = \mathbf{S}^T \mathbf{h}, \quad (102)$$

donde el superíndice  $T$  denota la matriz transpuesta. Sustituyendo la ecuación (101) en la ecuación (102), se obtiene

$$\mathbf{u} = \mathbf{S}^T \mathbf{S} \mathbf{a}, \quad (103)$$

de donde se puede despejar  $\mathbf{a}$  y sustituirla en la ecuación (101) para obtener el filtro SDF:

$$\mathbf{h} = \mathbf{S}(\mathbf{S}^T \mathbf{S})^{-1} \mathbf{a}. \quad (104)$$

Las restricciones de la salida se cumplen de forma exacta cuando en la escena aparece una de las imágenes de entrenamiento. Para apariencias del objeto que no se contemplan en el conjunto de entrenamiento se espera que la salida del filtro sea similar a aquellas con las que se diseñó, es decir, se espera que el filtro sea tolerante a variaciones intracase.

### IV.3.2. Filtros de correlación de Mínima Energía Promedio de Correlación

Como los filtros SDF no controlan el plano completo de correlación, Mahalanobis *et al.* (1987) propusieron una alternativa para diseñar filtros compuestos en la que se controla tanto el pico de correlación en la localización del objeto como la energía en el plano de correlación completo. El objetivo es minimizar la energía en toda la salida

del filtro de manera que el fondo tienda a ser plano mientras que el pico de correlación sea lo más afilado posible. De aquí el nombre de filtros de mínima energía promedio de correlación (MACE<sup>7</sup>). Al minimizar la energía en todo el plano, se suprimen los lóbulos laterales y por lo tanto se disminuye la probabilidad de falsas alarmas.

Para el diseño de filtros MACE se minimiza la energía promedio del plano de salida sujeto a restricciones de la salida  $\mathbf{u}$  para cada una de las imágenes de entrenamiento. Como en el caso de los filtros SDF, las restricciones se fijan como uno para las imágenes del objeto de interés y cero para objetos a rechazar. Sea  $\mathbf{S}$  la matriz cuyas columnas son las transformadas discretas de Fourier de las imágenes de entrenamiento y sea  $\mathbf{h}$  la respuesta de frecuencia del filtro en forma de vector.

$$d \cdot \mathbf{u} = \mathbf{S}^+ \mathbf{h}, \quad (105)$$

donde el superíndice  $+$  indica transpuesta conjugada. La energía promedio del plano de salida (ACE<sup>8</sup>) se calcula como

$$\text{ACE} = \mathbf{h}^+ \mathbf{D} \mathbf{h}. \quad (106)$$

$\mathbf{D}$  es una matriz de  $d \times d$  elementos de la siguiente forma:

$$\mathbf{D} = \frac{1}{d \cdot n} \sum_{i=1}^n |\mathbf{X}_i|^2, \quad (107)$$

donde  $\mathbf{X}_i$  es una matriz diagonal de  $d \times d$  elementos, donde los elementos en su diagonal son los coeficientes de la transformada discreta de Fourier de la  $i$ -ésima imagen de entrenamiento. El problema es la minimización del funcional cuadrático ACE sujeto a las restricciones de la ecuación (105). Esto se puede resolver utilizando la optimización

---

<sup>7</sup>Minimum Average Correlation Energy

<sup>8</sup>Average Correlation Energy

con multiplicadores de Lagrange. La solución para el filtro  $\mathbf{h}$  está dada por

$$\mathbf{h} = \mathbf{D}^{-1}\mathbf{S} (\mathbf{S} + \mathbf{D}^{-1}\mathbf{S})^{-1} \mathbf{u}. \quad (108)$$

Los filtros MACE presentan dos desventajas principales. La primera es que al no tomar en cuenta el ruido en el proceso de diseño, su desempeño es altamente sensible a la presencia de ruido. La segunda es que son altamente sensibles a variaciones intra-clase. Es decir, mientras que un filtro MACE cumple exactamente las restricciones de salida para las imágenes de entrenamiento, no generaliza adecuadamente para apariencias del objeto que no se hayan considerado como parte del conjunto de entrenamiento.

### IV.3.3. Filtros de correlación de Máxima Altura Promedio de Correlación

En el caso de las técnicas SDF y MACE para el diseño de filtros compuestos, los filtros se diseñan para cumplir un conjunto de restricciones en forma exacta. Sin embargo, en estas técnicas no se incorpora explícitamente la tolerancia a variaciones intra-clase. Con este fin, Mahalanobis *et al.* (1994) propusieron los filtros de máxima altura promedio de correlación (MACH<sup>9</sup>). Dado que el valor del pico de correlación para vistas del objeto no pertenecientes al conjunto de entrenamiento difiere del valor especificado para la clase verdadera, el diseño de filtros MACH se hace relajando las restricciones de salida para las imágenes entrenadas.

La primer idea que se utiliza en su diseño es que la salida completa de correlación es una transformación lineal de la imagen de entrada. Por lo que se debe tomar en cuenta no solo el pico de correlación sino el resto de la superficie. Entonces se establece

---

<sup>9</sup>Maximum Average Correlation Height

la métrica de similitud promedio (ASM<sup>10</sup>) entre los planos de salida como

$$\text{ASM} = \frac{1}{d \cdot n} \sum_{i=1}^n |y_i - \bar{y}|^2, \quad (109)$$

donde  $y_i$  es la salida de correlación para la  $i$ -ésima imagen de entrenamiento, representada como vector en el dominio de frecuencia, y  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  es la salida promedio para todas las imágenes de entrenamiento. Si  $\mathbf{X}_i$  denota la matriz diagonal con los coeficientes de la transformada discreta de Fourier de la  $i$ -ésima imagen en su diagonal y  $\mathbf{h}$  el filtro en forma de vector,  $y_i$  se puede calcular como

$$y_i = \mathbf{X}_i^* \mathbf{h}, \quad (110)$$

y la salida promedio como

$$\bar{y} = \mathbf{M}^* \mathbf{h}, \quad (111)$$

donde  $\mathbf{M} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i$ . Con esta notación, se puede reescribir la ASM como

$$\text{ASM} = \mathbf{h}^+ \mathbf{S} \mathbf{h}, \quad (112)$$

con

$$\mathbf{S} = \frac{d}{n} \sum_{i=1}^n (\mathbf{X}_i - \mathbf{M}) (\mathbf{X}_i - \mathbf{M})^*. \quad (113)$$

La otra idea que se utiliza en el diseño de filtros MACH es la de maximizar la altura del pico de correlación para las imágenes de la clase verdadera. Esto se logra maximizando la altura promedio de correlación (ACH<sup>11</sup>), que se define a continuación:

$$\text{ACH} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^+ \mathbf{h} = \mathbf{m}^+ \mathbf{h}, \quad (114)$$

donde  $\mathbf{x}_i$  y  $\mathbf{m}$  son vectores cuyos elementos son los elementos de las diagonales de  $\mathbf{X}_i$  y

---

<sup>10</sup>Average Similarity Metric

<sup>11</sup>Average Correlation Height

$\mathbf{M}$ , respectivamente.

Para dotar al filtro de tolerancia a ruido, se minimiza la varianza debida al ruido (ONV<sup>12</sup>) que se calcula como (Kumar *et al.*, 2005):

$$\text{ONV} = \mathbf{h}^+ \mathbf{C} \mathbf{h}, \quad (115)$$

donde  $\mathbf{C}$  es una matriz diagonal cuyos elementos son los elementos del espectro de potencia del ruido.

Finalmente, el filtro es aquél que maximiza el criterio

$$J(\mathbf{h}) = \frac{|\text{ACH}|^2}{\text{ASM} + \text{ONV}} = \frac{\mathbf{h}^+ \mathbf{m} \mathbf{m}^+ \mathbf{h}}{\mathbf{h}^+ (\mathbf{S} + \mathbf{C}) \mathbf{h}}. \quad (116)$$

La solución del problema de optimización de la ecuación (116) es el eigenvector dominante de  $(\mathbf{S} + \mathbf{C})^{-1} \mathbf{m} \mathbf{m}^+$ , es decir

$$\mathbf{h} = \gamma (\mathbf{S} + \mathbf{C})^{-1} \mathbf{m}, \quad (117)$$

donde  $\gamma$  es una constante de normalización.

## IV.4. Resumen

En este capítulo se presentan los filtros de correlación más comunes en la literatura y con los que se compara el desempeño de los filtros propuestos en esta tesis. Se presentan la derivación de filtros para dos modelos de escena, traslapada y no traslapada, así como sus respectivas características. Se presentan también los principales filtros compuestos para lograr tolerancia a distorsiones geométricas. Además de los filtros, se presentan las técnicas de diseño que se utilizan en el desarrollo de esta tesis.

---

<sup>12</sup>Output Noise Variance

## Capítulo V

---

# Diseño de filtros para el reconocimiento de objetos mediante imágenes ruidosas de referencia

---

El diseño de filtros de correlación, generalmente se realiza bajo la suposición de que se conocen la apariencia y forma del objeto de interés en forma explícita. Sin embargo, en la práctica la apariencia y forma del objeto se obtienen a partir de imágenes de referencia que inevitablemente se encuentran contaminadas por ruido. La extracción de la apariencia del objeto se hace mediante un proceso manual de segmentación y mejora de la imagen de referencia. Puede ser el caso también, que la aplicación de detección no permita el tiempo necesario para la realización del proceso manual de segmentación. Por ejemplo, en una aplicación de vigilancia, en la que se podría tomar una imagen de ejemplo que contenga al objeto de interés y se requiere detectar en tiempo real en una secuencia de video. Otra aplicación puede ser la búsqueda en bancos de imágenes mediante ejemplos, donde un ejemplo puede ser una imagen natural que contenga al objeto de interés incrustado en un fondo complejo.

Un enfoque popular para la detección de objetos mediante ejemplos es el de aprendizaje de máquina. Esto se realiza entrenando un clasificador estadístico, como una máquina de vectores de soporte, mediante muchos ejemplos que contienen al objeto y ejemplos que no. Mientras que este enfoque puede dar buenos resultados, la principal desventaja es que la cantidad de imágenes de entrenamiento puede ser grande.

El objetivo de este capítulo es presentar la derivación de filtros óptimos para el

---

reconocimiento de objetos utilizando imágenes ruidosas de referencia. La suposición fundamental es que al considerar la presencia del ruido como parte del modelo de señales se obtienen mejores resultados que considerando la segmentación y extracción de la apariencia del objeto como actividades separadas del proceso de diseño de filtros. Se proponen tres modelos de señales que representan situaciones progresivamente más complejas para el reconocimiento de objetos.

Además, en este capítulo se analiza el desempeño de los filtros derivados, junto con los métodos de estimación propuestos, en términos de capacidad de discriminación y errores de localización. Se compara el desempeño con el de filtros clásicos de correlación usando tanto información ideal sobre la apariencia del objeto de interés como imágenes ruidosas de referencia.

## V.1. Modelo Aditivo-Aditivo (AA)

Como primer aproximación al problema, consideremos un modelo en el que las imágenes de referencia están contaminadas por ruido aditivo y que los filtros obtenidos se utilizarán para detectar objetos en escenas que también se encuentran contaminadas sólo por ruido aditivo. Este modelo representa el problema de detección de objetos en imágenes de infrarrojo (Milton *et al.*, 1985) o en señales de audio.

### V.1.1. Modelo de señales

El modelo de señales está dado formalmente por

$$r(x) = t(x - x_r) + n_r(x), \quad (118)$$

$$s(x) = t(x - x_s) + n_s(x), \quad (119)$$

donde  $r(x)$  y  $s(x)$  son la imagen de referencia y la escena, respectivamente;  $t(x)$  representa al objeto de entrada;  $x_r$  y  $x_s$  son las coordenadas del objeto en la imagen de referencia y en la escena, respectivamente;  $n_r(x)$  y  $n_s(x)$  son los ruidos aditivos en la imagen de referencia y en la escena, respectivamente.

La derivación de los filtros se hace bajo las suposiciones siguientes:

1. Los procesos de ruido en las imágenes se pueden modelar como realizaciones de procesos estocásticos estacionarios.
2. Se conocen las densidades espectrales de los ruidos  $n_r(x)$  y  $n_s(x)$  y se representan con  $N_r(\omega)$  y  $N_s(\omega)$ , respectivamente.
3. Las medias y varianzas de los ruidos son conocidas. Sean  $\nu_r$  y  $\zeta_r$  la media y varianza de  $n_r(x)$ , respectivamente. Análogamente,  $\nu_s$  y  $\zeta_s$  son la media y varianza de  $n_s(x)$ , respectivamente.
4. Las coordenadas del objeto en las imágenes  $x_r$  y  $x_s$  se consideran variables aleatorias.
5. Todas las variables aleatorias y procesos estocásticos del modelo se tratan como estadísticamente independientes.

La salida del filtro  $y(x)$  se calcula como

$$y(x) = \mathcal{F}^{-1} \{S(\omega) H(\omega)\}, \quad (120)$$

donde  $S(\omega)$  es la transformada de Fourier de la escena de entrada y  $H(\omega)$  es la respuesta de frecuencia un filtro lineal invariante a desplazamiento. La forma deseada del filtro es la siguiente:

$$H(\omega) = A(\omega) R^*(\omega), \quad (121)$$

donde  $A(\omega)$  es una función determinística real y  $R(\omega)$  es la transformada de Fourier de la imagen de referencia dada en la ecuación (118). Es importante notar que la función de transferencia del filtro en realidad representa un banco de funciones, y una realización particular del filtro está determinada por la realización particular del ruido en la imagen de referencia. La forma de la respuesta de frecuencia del filtro debe ser la dada en la ecuación (121), ya que si se permite que el filtro sea una función compleja arbitraria, el proceso de optimización produce como resultado los filtros clásicos conocidos en la literatura.

Dado que la ubicación del objeto en la imagen de referencia es  $x_r$  y no necesariamente el origen, el pico de correlación debe ocurrir en la coordenada  $x_0 = x_s - x_r$  en la salida del filtro. Sin embargo, si el objeto se encuentra razonablemente centrado en la imagen de referencia, la estimación de la posición del objeto debe estar en la región cercana de su verdadera posición en la escena. Aún y cuando la posición exacta del objeto no se pueda determinar, hay aplicaciones como el seguimiento de objetos (Yilmaz *et al.*, 2006), donde lo que importa es el desplazamiento relativo del objeto en una secuencia de imágenes.

### V.1.2. Derivación del filtro óptimo

El filtro propuesto se obtiene al maximizar la razón pico a energía de salida (Javidi y Wang, 1994a). El criterio POE se define formalmente como

$$\text{POE} = \frac{|\mathbb{E}\{y(x_0)\}|^2}{\mathbb{E}\{|y(x)|^2\}}. \quad (122)$$

La barra en el denominador denota promedio espacial, es decir, si  $L$  es la duración espacial de la señal  $y(x)$  su promedio espacial es

$$\frac{1}{L} \int_{-\infty}^{\infty} y(x) dx. \quad (123)$$

Cabe destacar que los filtros obtenidos a partir de este criterio también presentan buen desempeño en términos de capacidad de discriminación, ya que tienden a producir picos muy marcados en la posición del objeto, sobre un fondo plano en el resto de la salida de correlación.

El numerador de la ecuación (122) representa el valor esperado del pico de correlación. Esto se puede calcular de la siguiente forma:

$$E\{y(x_0)\} = E\left\{\frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) H(\omega) e^{i\omega x_0} d\omega\right\}. \quad (124)$$

El denominador del criterio POE es el valor esperado de la energía promedio de la salida del filtro. En el dominio espacial la energía de salida del filtro se puede calcular como

$$E\left\{\overline{|y(x)|^2}\right\} = E\left\{\frac{1}{L} \int_{-\infty}^{\infty} |y(x)|^2 dx\right\}. \quad (125)$$

Por el teorema de Parseval, la ecuación (125) se puede calcular en el espacio de frecuencia como sigue:

$$E\left\{\overline{|y(x)|^2}\right\} = E\left\{\frac{1}{2\pi L} \int_{-\infty}^{\infty} |S(\omega) H(\omega)|^2 d\omega\right\}. \quad (126)$$

Sustituyendo la forma del filtro y utilizando la linealidad del valor esperado, las ecuaciones (124) y (126) se pueden reescribir como

$$E\{y(x_0)\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} A(\omega) E\{S(\omega) R^*(\omega) e^{i\omega x_0}\} d\omega, \quad (127)$$

y

$$E \left\{ \overline{|y(x)|^2} \right\} = \frac{1}{2\pi L} \int_{-\infty}^{\infty} |A(\omega)|^2 E \left\{ |S(\omega) R^*(\omega)|^2 \right\} d\omega, \quad (128)$$

respectivamente. Sustituyendo las ecuaciones (127) y (128) en (122), el criterio POE se puede reescribir como

$$\text{POE} = \frac{L \left| \int_{-\infty}^{\infty} A(\omega) E \left\{ S(\omega) R^*(\omega) e^{i\omega x_0} \right\} d\omega \right|^2}{2\pi \int_{-\infty}^{\infty} |A(\omega)|^2 E \left\{ |S(\omega) R^*(\omega)|^2 \right\} d\omega}, \quad (129)$$

y aplicando la desigualdad de Cauchy-Schwarz, el valor del criterio POE es máximo cuando  $A(\omega)$  toma la siguiente forma:

$$A(\omega) = \alpha \frac{E \left\{ S(\omega) R^*(\omega) e^{i\omega x_0} \right\}}{E \left\{ |S(\omega) R^*(\omega)|^2 \right\}}, \quad (130)$$

donde  $\alpha$  es una constante. Bajo la suposición de que la escena de entrada y la imagen de referencia son estadísticamente independientes, y sustituyendo  $x_0$  en la ecuación (130), podemos calcular  $A(\omega)$  de la siguiente forma:

$$A(\omega) = \alpha \frac{E \left\{ S(\omega) e^{i\omega x_s} \right\} [E \left\{ R(\omega) e^{i\omega x_r} \right\}]^*}{E \left\{ |S(\omega)|^2 \right\} E \left\{ |R(\omega)|^2 \right\}}. \quad (131)$$

Finalmente, sustituyendo las ecuaciones (118), (119) y (131) en la ecuación (121), obtenemos la respuesta de frecuencia del filtro óptimo respecto al criterio POE:

$$H_{AA}(\omega) = \frac{|T(\omega)|^2 R^*(\omega)}{\left[ (|T(\omega)|^2 + N_r(\omega)) (|T(\omega)|^2 + N_s(\omega)) \right]}, \quad (132)$$

donde  $T(\omega)$  denota la transformada de Fourier del objeto de interés. Cabe mencionar que cuando la imagen de referencia contiene al objeto libre de ruido, el filtro obtenido es exactamente el filtro OF (Yaroslavsky, 1993) para el modelo aditivo.

### V.1.3. Estimación de parámetros

El filtro dado en la ecuación (132) requiere conocer el espectro de potencia del objeto de interés. Como el espectro de potencia es un dato que no se conoce bajo el modelo utilizado, es necesario estimarlo a partir de la imagen de referencia. Bajo la suposición de independencia estadística del objeto y el ruido en la imagen de referencia, el espectro de potencia del objeto se puede estimar de la siguiente forma (McDonough y Whalen, 1995)

$$|T(\omega)|^2 \approx |R(\omega)|^2 - N_r(\omega). \quad (133)$$

Usando esta estimación, se obtiene una aproximación del filtro óptimo, denotada por  $AA_1$ :

$$AA_1 = \frac{|R(\omega)|^2 - N_r(\omega)}{|R(\omega)|^2} \times \frac{R^*(\omega)}{|R(\omega)|^2 - N_r(\omega) + N_s(\omega)}. \quad (134)$$

El filtro resultante en (134) consiste en la aplicación de dos filtros en cascada. El primero es un filtro Wiener empírico que se aplica a la imagen de referencia y la imagen después del filtrado se usa para construir una aproximación del OF para detectar el objeto en en la escena de entrada. De hecho, el filtro obtenido en la ecuación (132) es equivalente a aplicar un filtro de Wiener ideal a la imagen de referencia para luego formar un filtro OF. De esto se puede concluir que de todos los filtros de la forma dada en la ecuación (121), el enfoque óptimo respecto al criterio POE consiste en limpiar la imagen de referencia con un filtro Wiener. De acuerdo a este resultado, se puede obtener una aproximación distinta del OF de la siguiente manera: se estima el objeto a partir de la imagen de referencia y se usa esa estimación para diseñar una instancia del OF. Denotando esta aproximación como  $AA_2$ , se tiene que

$$AA_2 = \frac{\hat{T}^*(\omega)}{|\hat{T}(\omega)|^2 + N_s(\omega)}, \quad (135)$$

donde  $\hat{T}(\omega)$  denota la transformada de Fourier del objeto restaurado con el filtro Wiener.

#### V.1.4. Resultados experimentales

En esta sección se presentan los resultados de simulaciones por computadora del filtro propuesto y se compara con el desempeño del OF (Yaroslavsky, 1993) en términos de capacidad de discriminación y errores de localización. Por las suposiciones del modelo, la localización exacta del objeto en la escena no se puede determinar al no conocer la localización del objeto en la imagen de referencia. Sin embargo, si el objeto en la referencia se encuentra razonablemente centrado, la estimación de la posición del objeto debe estar cerca de su verdadera localización. En general, el objeto puede aparecer en cualquier ubicación en la imagen de referencia, sin embargo, para las simulaciones que aquí se presentan la ubicación del objeto se fija en el centro de la imagen, es decir  $x_r = 0$ .

Las imágenes que se utilizan todas son de  $256 \times 256$  píxeles. Los valores de intensidad de las imágenes utilizadas están en el intervalo  $[0-255]$ . En las figuras 4(a) y 4(b) se muestra el objeto de interés (de tamaño  $44 \times 28$  píxeles) y una realización de un fondo correlacionado. El ejemplo del fondo proviene de un proceso estocástico de media  $\mu_s = 100$ , coeficientes de correlación vertical y horizontal  $\rho_v = \rho_h = 0.95$  y desviación estándar de 30. El fondo se genera a partir de instancias de ruido blanco aplicando el proceso de correlación que se describe en el Apéndice A.

Para garantizar resultados estadísticamente correctos, se realizaron 30 repeticiones de cada experimento para diferentes ubicaciones del objeto en la escena y distintas realizaciones de procesos de ruido. Todos los filtros se implementaron utilizando la

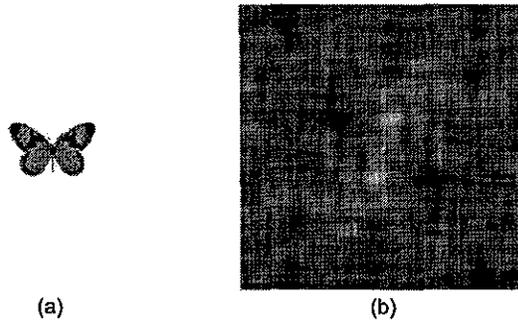


Figura 4: (a) Objeto de interés y (b) realización de un fondo estocástico.

transformada discreta de Fourier.

Para fines de comparación, se realizaron experimentos con el OF (Yaroslavsky, 1993) ideal y con una variante del OF diseñada a partir de la imagen de referencia ruidosa. Para el OF ideal, se conoce la apariencia del objeto de forma exacta. La variante del OF que utiliza la imagen ruidosa se denota por OF-R. Para componer una instancia del OF-R, se utiliza la imagen de referencia exactamente como se obtiene como si contuviera al objeto sin presencia de ruido.

Primero se analiza el caso en el que el ruido aditivo es blanco Gaussiano de media cero, tanto en la imagen de referencia como en la escena de entrada. En las figuras 5(a) y 5(b) se muestra el desempeño de los filtros en términos de DC y LE, respectivamente, cuando se varia la desviación estándar del ruido en la referencia de 5 a 40. La desviación estándar del ruido en la escena de entrada se mantiene fijo en  $\zeta_s = 20$ . En las figuras 5(c) y 5(d) se muestra el desempeño de los filtros cuando a desviación estándar del ruido en la escena es 30.

Se puede observar que el filtro OF logra el mejor desempeño de reconocimiento. Se utiliza para determinar una cota superior del desempeño alcanzable por los filtros propuestos. Su alto desempeño es de esperarse, dado que dispone de información ideal sobre la apariencia del objeto. Por otro lado, el filtro OF-R que ignora la presencia de

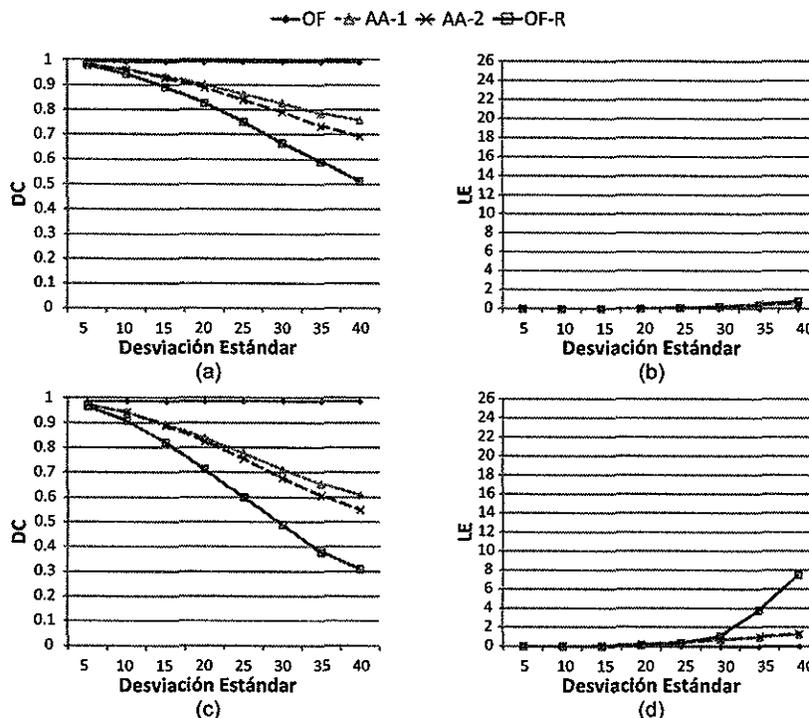


Figura 5: Desempeño de los filtros cuando el ruido es blanco y variando la desviación estándar del ruido en la imagen de referencia. (a) DC y (b) LE cuando  $\zeta_s = 20$ ; (c) DC y (d) LE cuando  $\zeta_s = 30$ .

ruido en la imagen de referencia es consistentemente peor que los filtros propuestos. Es claro que, a excepción del OF que conoce la apariencia exacta del objeto, el desempeño de los filtros baja en función de la desviación estándar del ruido en la referencia. Al aumentar el nivel de ruido presente en la escena, el desempeño de los filtros baja más rápido conforme aumenta  $\zeta_r$ . Sin embargo, el desempeño relativo es consistente en ambos casos. El desempeño de los filtros propuestos se encuentra entre el desempeño ideal del OF y por encima del desempeño del filtro OF-R. Para los casos considerados, todos los filtros detectan correctamente al objeto.

Los resultados son similares cuando la escena de entrada está contaminada por ruido correlacionado. Para los resultados presentados, el fondo proviene de un proceso

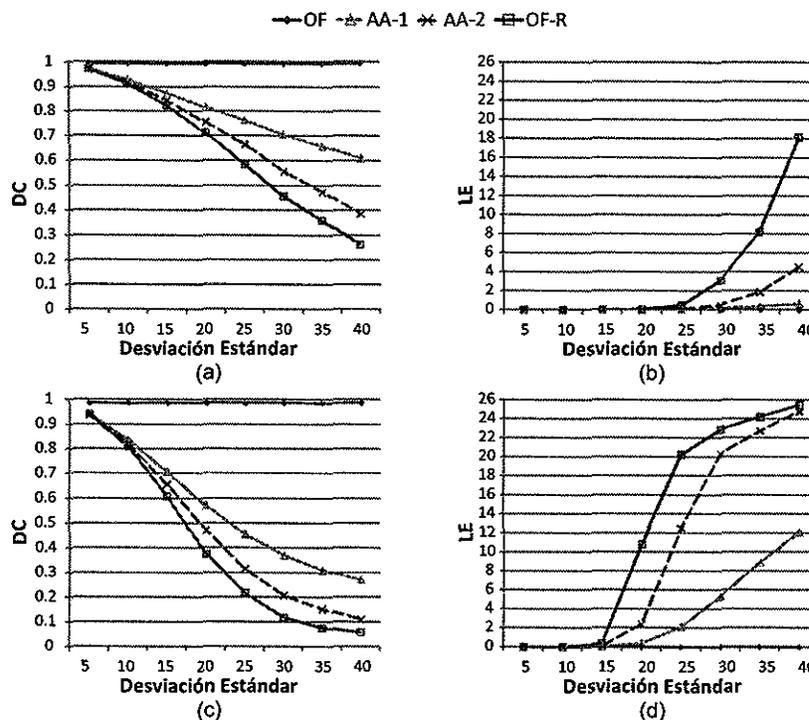


Figura 6: Desempeño de los filtros cuando el ruido es correlacionado con coeficiente de correlación de 0.5, variando la desviación estándar del ruido en la imagen de referencia. (a) DC y (b) LE cuando  $\zeta_s = 20$ ; (c) DC y (d) LE cuando  $\zeta_s = 30$ .

estacionario con coeficientes de correlación 0.5. En las figuras 6(a) y 6(b) se muestra el desempeño de los filtros probados, en términos de DC y LE, respectivamente. La desviación estándar del ruido en la imagen de referencia se varía de 5 a 40 mientras la desviación estándar del ruido en la escena de entrada se mantiene fija en 20. Es claro que el desempeño de los filtros se degrada más rápido en este caso que en el caso en que ambos ruidos eran blancos. Esto se debe a que al aumentar el coeficiente de correlación del ruido, se forman estructuras en el fondo que tienen apariencia similar a la del objeto.

El desempeño del OF-R es el más afectado por al aumento de la intensidad del ruido en la imagen de referencia. Los filtros propuestos, sin embargo, son capaces de detectar al objeto aún en presencia de ruido intenso en la referencia, aunque con pequeños errores

de localización. En las figuras 6(c) y 6(d) se muestra el desempeño en términos de DC y LE, respectivamente, cuando la desviación estándar del ruido en la escena es 30. En este caso la diferencia en desempeño entre los distintos filtros estudiados es más notoria. Se puede observar que tanto el  $AA_1$  y el  $AA_2$  pueden detectar al objeto aunque con errores de localización.

Mediante numerosas simulaciones por computadora, se puede concluir que el desempeño del filtro  $AA_1$  es consistentemente mejor que el del filtro  $AA_2$ . Dada que la diferencia entre ambos filtros radica en la estimación del espectro de potencia del objeto, se puede concluir que para el reconocimiento de patrones es mejor estimar el espectro de potencia del objeto directamente de la referencia ruidosa en lugar de utilizar el objeto restaurado (véanse las ecuaciones (133) y (135)).

## V.2. Modelo Aditivo-No traslapado (AN)

El modelo AN constituye la segunda parte de este trabajo. En este modelo se considera que la imagen de referencia se encuentra contaminada por ruido aditivo, mientras que el objeto se va a detectar en escenas que contienen al objeto de interés incrustado en un fondo complejo. Este modelo puede representar casos en los que la imagen de referencia se captura en un ambiente controlado y se usa para diseñar un filtro para detección en imágenes naturales sin realizar la segmentación manual para indicar explícitamente la apariencia del objeto.

$$r(x) = t(x - x_r) + n_r(x), \quad (136)$$

$$s(x) = t(x - x_s) + b_s(x) \bar{w}(x - x_s) + n_s(x), \quad (137)$$

donde  $b_s(x)$  denota el fondo complejo y  $\bar{w}(x)$  denota la región de soporte inversa del objeto, que toma valores de cero y uno en las regiones de adentro y afuera del objeto, respectivamente. Se denota por  $\bar{W}(\omega)$  la transformada de Fourier de la región de soporte inversa. Para facilitar el tratamiento matemático, se supone que el fondo es una realización de un proceso estocástico estacionario de media  $\mu_s$  y varianza  $\sigma_s^2$ . Además, la densidad espectral del proceso de media cero  $b_s^0(x) = b_s(x) - \mu_s$ , se supone conocida y se denota por  $B_s^0(\omega)$ . Para el resto de las señales en el modelo se usan las mismas suposiciones que en la sección V.1. En el caso de que el fondo no se pueda modelar adecuadamente como un proceso estacionario, se puede partir la escena en regiones más pequeñas que sí sean estacionarias. De esta manera, en cada región se puede aplicar el filtro diseñado bajo la suposición de que el fondo es estacionario.

La densidad espectral del fondo se supone conocida. Sin embargo, si no se conoce exactamente se puede modelar mediante el modelo exponencial (Jain, 1989). La longitud de correlación se puede estimar al analizar una pequeña región homogénea.

A partir del modelo de la escena de entrada, el reconocimiento debe realizarse al encontrar la correspondencia entre la escena de entrada y el objeto de referencia junto con su región de soporte inversa  $\mu_s \bar{w}(x)$ . Si la apariencia y forma del objeto se conocen en forma explícita, el filtro óptimo para el criterio POE y el modelo de escena disjunto es el GOF (Javidi y Wang, 1994a). La respuesta de frecuencia del GOF está dada por

$$\text{GOF} = \frac{T_s^*(\omega)}{|T_s(\omega)|^2 + \tilde{N}_s(\omega)}, \quad (138)$$

donde

$$T_s(\omega) = T(\omega) + \mu_s \bar{W}(\omega), \quad (139)$$

y

$$\tilde{N}_s(\omega) = \frac{1}{2\pi} B_s^0(\omega) \bullet |\bar{W}(\omega)|^2 + N_s(\omega), \quad (140)$$

es decir,  $T_s(\omega)$  es la transformada de Fourier un nuevo objeto compuesto  $t_s(x)$  formado por el objeto de referencia y el promedio del fondo en la escena, y  $\tilde{N}_s(\omega)$  es la densidad espectral de potencia del ruido en la escena  $\tilde{n}_s(x; x_s) = b_s^0(x) \bar{w}(x - x_s) + n_s(x)$  formado por el fondo complejo y el ruido aditivo.

Dado que en el modelo de señales considerado, ni el objeto de interés ni su forma se conocen explícitamente, es necesario diseñar una aproximación del GOF usando sólo la información disponible en la imagen de referencia. Sea  $\hat{w}(x)$  una estimación de la región de soporte del objeto. Sustituyendo esta estimación en la ecuación (140) se obtiene una estimación de la densidad espectral del ruido. Un enfoque sencillo para estimar  $T_s(\omega)$  es restaurar el objeto aplicando un filtro lineal a la imagen de referencia y sumarle la estimación de la región de soporte:

$$\hat{t}_1(x) = g_1(x) \bullet r(x) + \mu_s \hat{w}(x), \quad (141)$$

donde  $g_1(x)$  es la respuesta al impulso del filtro lineal para eliminar el ruido en la imagen de referencia. Bajo el criterio de mínimo error cuadrático medio, el filtro óptimo es el filtro de Wiener (Pratt, 2007). De esta manera, una aproximación del filtro GOF para el modelo AN se obtiene como sigue:

$$\text{AN}_1(\omega) = \frac{T_1^*(\omega)}{|T_1(\omega)|^2 + \hat{N}_s(\omega)}, \quad (142)$$

siendo  $\hat{T}_1(\omega)$  la transformada de Fourier de  $\hat{t}_1(x)$ .

Otra manera de obtener un filtro para el modelo AN consiste en considerar una

imagen de referencia sintetizada como sigue:

$$\tilde{r}(x) = r(x) + \mu_s \bar{w}(x - x_r). \quad (143)$$

De esta manera, podemos reescribir el modelo de señales como

$$\tilde{r}(x) = t_s(x - x_r) + n_r(x), \quad (144)$$

y

$$\tilde{s}(x) = t_s(x - x_s) + \tilde{n}_s(x; x_s). \quad (145)$$

Esto es, reducir un problema del modelo AN a modelo AA, considerando la parte determinística de la escena de entrada como un nuevo objeto a detectar, mientras que las partes no determinísticas se tratan como ruido aditivo. Usando el filtro óptimo para el modelo AA dado en la ecuación (132), obtenemos el filtro óptimo para el modelo AN:

$$H_{AN}(\omega) = \frac{|T_s(\omega)|^2}{|T_s(\omega) + N_r(\omega)|} \times \frac{\tilde{R}(\omega)}{|T_s(\omega)|^2 + \tilde{N}_s(\omega)}, \quad (146)$$

donde  $\tilde{R}(\omega)$  denota la transformada de Fourier de  $\tilde{r}(x)$ . El filtro resultando consiste en un proceso de dos pasos: primero se aplica un filtro lineal a  $\tilde{r}(x)$  para restaurar  $t_s(x)$ , y segundo, el objeto restaurado resultante se utiliza para realizar la detección en la escena de entrada. Es importante notar que si la imagen de referencia no contiene ruido, el filtro obtenido es exactamente el filtro GOF. Sea  $\hat{t}_2(x - x_r)$  una estimación de  $t_s(x - x_r)$  obtenida de esta manera, es decir

$$\hat{t}_2(x - x_r) = g_2(x) \bullet [r(x) + \mu_s \hat{w}(x - x_r)], \quad (147)$$

donde  $g_2(x)$  es la respuesta al impulso del filtro óptimo para restaurar el objeto compuesto a partir de  $\tilde{r}(x)$ . Si  $\hat{T}_2(\omega)$  denota la transformada de Fourier de  $\hat{t}_2(x - x_r)$ , se puede obtener una estimación del filtro dado en la ecuación (147) de la siguiente

manera:

$$\text{AN}_2(\omega) = \frac{T_2^*(\omega)}{|T_2(\omega)|^2 + \hat{N}_s(\omega)}. \quad (148)$$

### V.2.1. Estimación de parámetros

El diseño de filtros para el modelo AN requiere conocer la región inversa de soporte del objeto. Esta se puede estimar utilizando un umbral (Duda *et al.*, 2000; Mardia y Hainsworth, 1988) a la imagen de referencia después de filtrar el ruido:

$$\hat{w}(x - x_r) = \begin{cases} 1 & \hat{t}(x - x_r) \leq \tau(\mu_t, \sigma_t, \zeta_r, A_t) \\ 0 & \text{de lo contrario} \end{cases}, \quad (149)$$

donde  $\hat{t}(x)$  es el objeto restaurado a partir de filtrar la imagen de referencia con el filtro Wiener y  $\tau(\mu_t, \sigma_t, \zeta_r, A_t)$  denota el umbral óptimo que depende de las estadísticas del objeto y del ruido en la imagen de referencia;  $\mu_t$  y  $\sigma_t$  denotan la media y desviación estándar del objeto, respectivamente y  $A_t$  denota la porción de la imagen de referencia ocupada por el objeto. El umbral óptimo se puede determinar suponiendo que tanto el objeto como el ruido aditivo tienen distribuciones Gaussianas. De hecho, el ruido residual tras filtrar con Wiener tiende a una distribución Gaussiana independientemente de la distribución del ruido original. Esto se explica mediante el Teorema del Límite Central (Papoulis, 1991): cada valor en la imagen filtrada con Wiener es la suma pesada de variables aleatorias idénticamente distribuidas. De manera que el umbral óptimo para separar distribuciones Gaussianas puede ser usado aún y cuando el ruido original sea no Gaussiano.

Después de filtrar, el objeto tiene media y varianza  $\mu_t$  y  $\tilde{\sigma}_t^2$ , respectivamente. El ruido después de filtrar tiene media cero y varianza  $\tilde{\zeta}_r^2$ . Bajo la suposición de que el objeto y el ruido son estadísticamente independientes, la varianza dentro de la región

del objeto es  $\hat{\sigma}_t^2 = \tilde{\sigma}_t^2 + \tilde{\zeta}_r^2$ . El umbral que se utiliza es el de mínima tasa de errores de clasificación (Duda *et al.*, 2000):

$$\tau = -\frac{\mu_t \tilde{\zeta}_r^2}{\tilde{\sigma}_t^2} + \frac{\hat{\sigma}_t \tilde{\zeta}_r}{\tilde{\sigma}_t^2} \sqrt{\mu_t^2 + 2\tilde{\sigma}_t^2 \log \left[ \frac{(1 - A_t) \hat{\sigma}_t}{A_t \tilde{\zeta}_r} \right]}. \quad (150)$$

La derivación del umbral se muestra en forma detallada en el Apéndice B.

El desempeño de los filtros se ve significativamente afectado por errores en la estimación de la región inversa de soporte. Es interesante que el mismo error en la estimación de la región inversa de soporte produce diferentes errores al estimar el objeto compuesto con las ecuaciones (141) y (147). En el Apéndice C se muestra la derivación de una condición sencilla para determinar cuándo  $\hat{t}_1(x)$  es mejor estimación que  $\hat{t}_2(x)$  en términos de error cuadrático medio. El umbral está dado por

$$\mu_s \geq 2 \left( \mu_t + k \frac{\sigma_t}{\sqrt{s_t}} \right), \quad (151)$$

donde  $s_t$  es la cantidad de pixeles que componen el objeto y  $k$  es una constante que depende del nivel de confianza con que se desea calcular el umbral. Sin embargo, el criterio MSE no refleja directamente el desempeño final de los filtros  $AN_1$  y  $AN_2$  en términos de capacidad de discriminación y errores de localización. Para caracterizar mejor la influencia que tienen de los errores de estimación del objeto compuesto sobre el desempeño de los filtros, se puede utilizar el criterio de error de correspondencia ( $ME^1$ ).

La detección y localización ideal se pueden lograr dado que se satisfagan dos condiciones: que las señales del objeto y de la escena de entrada estén adecuadamente normalizadas, y que la señal del objeto de referencia y de la escena de entrada que se usan en el proceso de reconocimiento sean idénticas. La primer condición se satisface

---

<sup>1</sup>Matching Error

al diseñar un filtro de correlación óptimo para el modelo de señales utilizado. Esencialmente tanto la señal de referencia como la escena de entrada se procesan con un filtro lineal para blanquearlas antes de buscar la correspondencia. La segunda condición se puede caracterizar por el ME, que mide en forma cuantitativa el qué tan parecidas son las señales. Nótese que en el dominio de frecuencia la correlación de dos señales idénticas es una función real dado que la fase se cancela completamente. Sean  $\{\hat{T}_i(\omega), i = 1, 2\}$ , las transformadas de Fourier de las estimaciones del objeto compuesto y  $T_s(\omega)$  la transformada de Fourier del objeto compuesto verdadero. El valor de la correlación de una estimación y el objeto verdadero está dado por

$$y_i(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{T_s(\omega)}{|T_s(\omega)|} \times \frac{T_i^*(\omega)}{|T_i(\omega)|} d\omega. \quad (152)$$

La cancelación completa de fase hace que  $y_i(0) = 1$ . Los errores de cancelación de fase en la frecuencia  $\omega$  se denotan por

$$\Delta\omega^{(i)} = \phi_s(\omega) - \phi_s^{(i)}(\omega), \quad (153)$$

donde  $i = 1, 2$ . Dado que las señales originales se consideran reales, los errores de cancelación de fase se pueden escribir como

$$y_i(0) = \frac{1}{\pi} \int_{-\infty}^{\infty} \cos(\Delta\omega^{(i)}) d\omega. \quad (154)$$

Finalmente, se define ME como la diferencia entre la salida ideal y la salida obtenida con las estimaciones, es decir:

$$ME_i = 1 - \frac{1}{\pi} \int_{-\infty}^{\infty} \cos(\Delta\omega^{(i)}) d\omega. \quad (155)$$

Los valores de la métrica ME cerca de cero indican una buena correspondencia entre las dos señales. Dado que las dos estimaciones propuestas del objeto compuesto tienen

diferente desempeño dependiendo de las estadísticas de las señales de entrada, se puede construir un mapa de decisión para señales de entrada típicas que indique qué estimación es mejor en términos de ME. Finalmente, el algoritmo para diseñar el filtro usando las estimaciones  $AN_1$  y  $AN_2$  es el siguiente:

1. Restaurar el objeto a partir de la imagen de referencia para obtener  $\tilde{t}(x - x_r) = g_1(x) \bullet r(x)$ .
2. Calcular el umbral óptimo  $\tau$  usando las estadísticas de las señales conocidas *a priori*.
3. Estimar  $\hat{w}(x - x_r)$  usando  $\tilde{t}(x - x_r)$  y  $\tau$ .
4. Si se espera que la estimación  $\hat{t}_1(x - x_r)$  funcione mejor que  $\hat{t}_2(x - x_r)$  en términos de ME:
  - a) Estimar el objeto compuesto como  $\hat{t}_1(x - x_r) = \tilde{t}(x - x_r) + \mu_s \hat{w}(x - x_r)$ .
  - b) Diseñar el filtro  $AN_1$ .
5. De lo contrario:
  - a) Sintetizar el objeto compuesto ruidoso  $\tilde{r}(x) = r(x) + \mu_s \hat{w}(x - x_r)$ .
  - b) Estimar el objeto compuesto como  $\hat{t}_2(x - x_r) = g_2(x) \bullet \tilde{r}(x)$ .
  - c) Diseñar el filtro  $AN_2$ .

## V.2.2. Resultados experimentales

En esta sección se analiza el desempeño de las dos estimaciones propuestas del filtro óptimo para el modelo AN. Primero se analizan el desempeño de las estimaciones del

objeto compuesto. Una cota en el espacio de los parámetros del modelo de señales a partir de la cual la estimación  $\hat{t}_1(x - x_r)$  es mejor que la estimación  $\hat{t}_2(x - x_r)$ , en términos del criterio MSE, se da en la ecuación (151). En la figura 7 se muestra el desempeño de ambas estimaciones en términos de MSE y ME como función de la media del fondo en la escena de entrada. Para estos experimentos se fijan la desviación estándar del objeto y la del ruido en la referencia en  $\sigma_t = 10$  y  $\zeta_r = 10$ , mientras se varía la media del fondo en la escena. La media del objeto es  $\mu_t = 100$  para las figuras 7(a) y 7(b) y  $\mu_t = 60$  para las figuras 7(c) y 7(d). Hay un punto de cruce trivial del desempeño de ambas estimaciones cuando la media del fondo es cero  $\mu_s = 0$ . El segundo punto de cruce para el criterio MSE se calcula a partir de la ecuación (151). Se puede apreciar en las figuras 7(a) y 7(c) que las curvas de desempeño se cortan cuando  $\mu_s$  se acerca a  $2\mu_t$ . Sin embargo, la capacidad de reconocimiento de los filtros tiene mejor correlación con el criterio ME. Esto se verificó con numerosas simulaciones por computadora. En las figuras 7(b) y 7(d) se puede apreciar que el desempeño de las estimaciones propuestas se cruza cerca de la región  $\mu_s \approx \mu_t$ . De aquí se puede concluir que la mejor estimación del objeto compuesto en términos de MSE no necesariamente es mejor cuando el objetivo final es el reconocimiento de objetos. Dependiendo de los valores de los parámetros del modelo, tanto la estimación  $AN_1$  puede ser mejor que la estimación  $AN_2$  y viceversa. Por lo tanto, para el modelo AN se deben utilizar ambas estimaciones para lograr el mejor desempeño de reconocimiento posible.

Ahora se presenta el desempeño de los filtros en términos de DC y LE. En la figura (8) se muestra el desempeño de los filtros en un punto del espacio de parámetros en el que la estimación  $AN_1$  es mejor que la estimación  $AN_2$ . Para estos experimentos se fijan  $\sigma_t = 10$ ,  $\mu_t = 100$ ,  $\mu_s = 160$  y  $\zeta_r$  se varía de 5 a 50. La DC de los filtros se muestra en la figura 8(a) con un intervalo de confianza de 95%. Se puede observar que

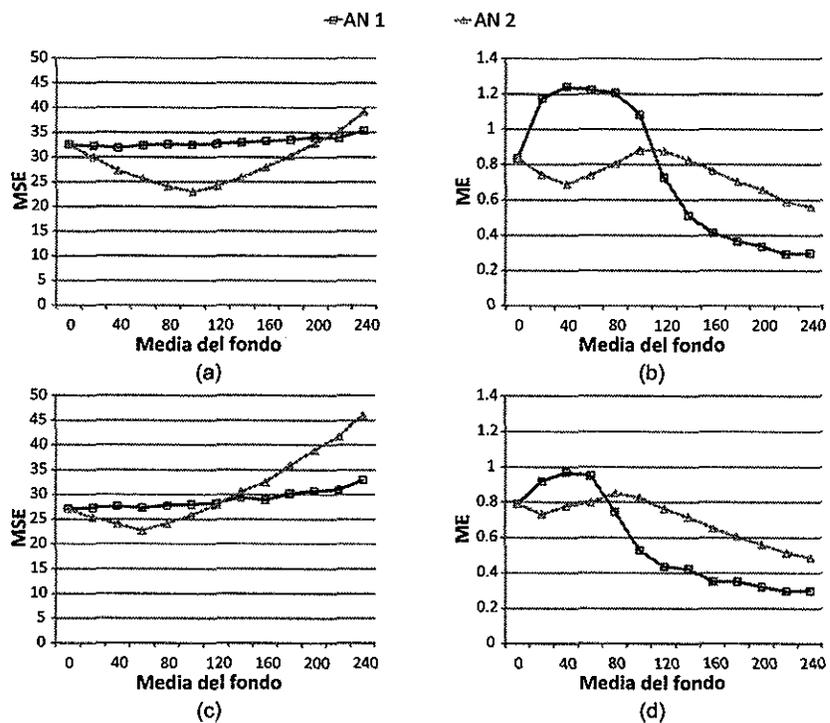


Figura 7: Desempeño de los métodos de estimación de la apariencia del objeto de interés al variar  $\mu_s$ , mientras que  $\sigma_t = 10$  y  $\zeta_r = 10$ . (a) MSE cuando  $\mu_t = 100$ , (b) ME cuando  $\mu_t = 100$ , (c) MSE cuando  $\mu_t = 60$  y (d) ME cuando  $\mu_t = 60$ .

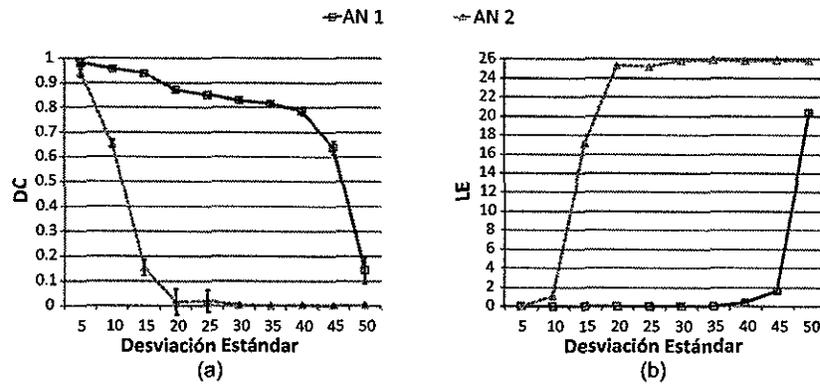


Figura 8: Desempeño de los filtros de correlación al variar la desviación estándar del ruido en la imagen de referencia y AN<sub>1</sub> es mejor que AN<sub>2</sub>.  $\mu_t = 100$ ,  $\sigma_t = 10$  y  $\mu_s = 160$ . (a) DC (con intervalo de 95 % de confianza) y (b) LE.

el desempeño de la estimación AN<sub>2</sub> baja rápidamente como función de la desviación estándar del ruido en la referencia. Por el contrario, el desempeño de la estimación AN<sub>1</sub> es significativamente más tolerante al ruido. Se puede observar en la figura 8(a) que el filtro AN<sub>1</sub> no presenta falsas alarmas para los valores de parámetros probados, con un 95 % de confianza. El filtro AN<sub>2</sub> empieza a fallar cuando la desviación estándar del ruido en la imagen de referencia supera el valor de 15. Las falsas alarmas ocurren cuando el intervalo de confianza cruza la línea DC = 0. Es interesante notar que para los parámetros del modelo que se muestran en estos experimentos la estimación  $\hat{t}_2(x - x_r)$  es mejor en términos de MSE que la estimación  $\hat{t}_1(x - x_r)$ , tal como se muestra en la figura 7(a), mientras que la DC del filtro AN<sub>2</sub> es peor que el desempeño del filtro AN<sub>1</sub>. Esto coincide con lo que indica el criterio ME en la figura 7(b).

Ahora se analiza un conjunto distinto de valores de parámetros del modelo, para los cuales el desempeño del filtro AN<sub>2</sub> es mejor que el del filtro AN<sub>1</sub>. En la figura 9 se muestran los resultados de estas simulaciones. Para este experimento se fijan los valores  $\sigma_t = 30$ ,  $\mu_t = 140$  y  $\mu_s = 100$ . La desviación estándar del ruido en la imagen de referencia se varía de 5 a 50. Se muestra la DC de los filtros en la figura 9(a) con

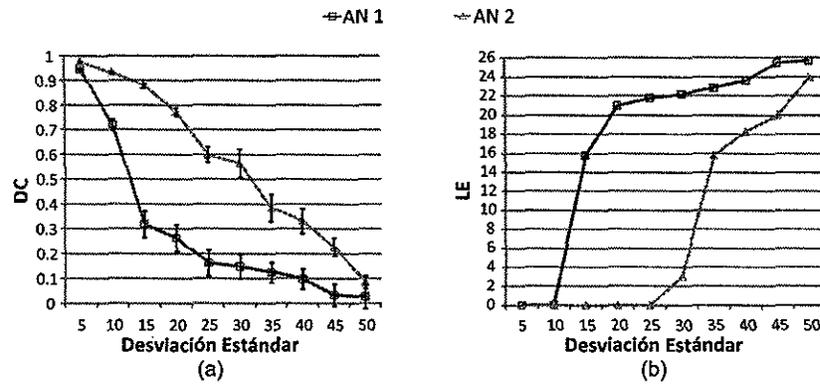


Figura 9: Desempeño de los filtros de correlación al variar la desviación estándar del ruido en la imagen de referencia y AN<sub>2</sub> es mejor que AN<sub>1</sub>.  $\mu_t = 140$ ,  $\sigma_t = 130$  y  $\mu_s = 100$ . (a) DC (con intervalo de 95 % de confianza) y (b) LE.

un intervalo de 95 % de confianza. Se puede observar que el desempeño del filtro AN<sub>2</sub> es superior al del AN<sub>1</sub>, tanto en términos de DC como en términos de LE. Como se muestra en la figura 9(a) el filtro AN<sub>2</sub> no presenta falsas alarmas para las intensidades de ruido consideradas en los experimentos, con 95 % de confianza. Por el contrario, el filtro AN<sub>1</sub> presenta falsas alarmas cuando la  $\zeta_r$  supera el valor de 40.

Para caracterizar mejor el comportamiento de los filtros, se realizaron numerosas simulaciones por computadora. Se varían los parámetros del modelo como sigue:  $\mu_t$  de 10 a 200;  $\mu_s$  de 10 a 240;  $\sigma_t$  de 0 a 100;  $\zeta_r$  de 5 a 50. Dado que los parámetros restantes  $\sigma_s$  y  $\zeta_s$  afectan el desempeño de ambos filtros de la misma manera, el valor de estos se fija en 10 para todos los experimentos. En la figura 10(a) se ilustran las regiones del espacio de parámetros donde cada estimación tiene un mejor desempeño que la otra en términos de DC. En las figuras 10(b) y 10(c) se muestra la separación aproximada del mismo espacio de parámetros donde cada estimación es mejor que la otra en términos de ME y MSE, respectivamente. Las separaciones se muestran con un nivel de 90 % de confianza. Se puede observar que las regiones mostradas en las figuras 10(a) y 10(b) son muy similares. De esta manera, se puede concluir que el desempeño

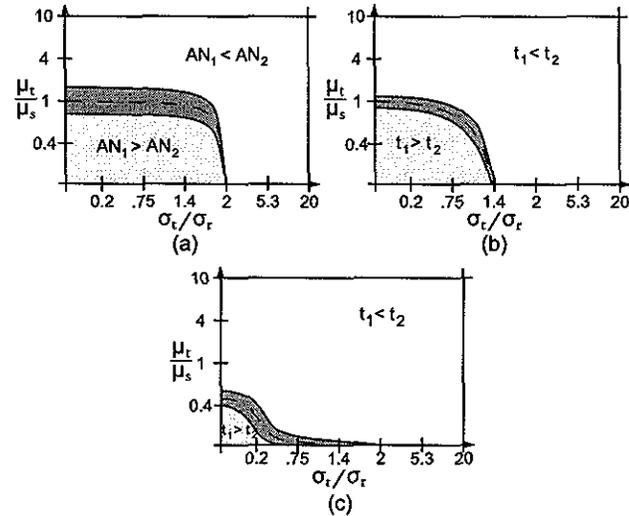


Figura 10: Partición del espacio de parámetros en regiones en las que cada estimación tiene mejor desempeño que la otra con respecto a las métricas (a) DC, (b) ME, y (c) MSE.

de las estimaciones en términos de ME se puede utilizar para predecir el desempeño esperado de los filtros en términos de calidad de detección. Se observan dos tendencias claras:

- incrementar las razones  $\mu_t/\mu_s$  o  $\sigma_t/\zeta_r$  tiende a favorecer a la variante  $AN_2$ , lo que explica el por qué es mejor esta estimación en la mayor parte del espacio de parámetros,
- y que la transición entre las regiones de mejor desempeño de cada filtro tiende a estar cerca del cruce  $\mu_t \approx \mu_s$ .

### V.3. Modelo No traslapado-No traslapado (NN)

El modelo NN se utiliza para representar la detección de objetos en imágenes naturales cuando lo único que se tiene de conocimiento sobre el objeto de interés es una

imagen ejemplo. Se supone que el objeto aparece en la imagen de referencia exactamente una vez, pero en una posición desconocida. El modelo de señales está dado formalmente por

$$r(x) = t(x - x_r) + b_r(x) \bar{w}(x - x_r) + n_r(x), \quad (156)$$

$$s(x) = t(x - x_s) + b_s(x) \bar{w}(x - x_s) + n_s(x), \quad (157)$$

donde  $b_r(x)$  representa el fondo complejo de la imagen de referencia. Se considera a  $b_r(x)$  como la realización de un proceso estocástico de media  $\mu_r$  y varianza  $\sigma_r^2$ . La densidad espectral del proceso  $b_r^0(x) = b_r(x) - \mu_r$  se supone conocida y se denota por  $B_r^0(\omega)$ .

Sea  $t_r(x - x_r)$  el valor esperado de la imagen de referencia, es decir

$$t_r(x - x_r) = t(x - x_r) + \mu_r \bar{w}(x - x_r), \quad (158)$$

y sea  $T_r(\omega)$  la transformada de Fourier de  $t_r(x)$ . El filtro óptimo para el modelo NN respecto al criterio POE de la forma del filtro dada en la ecuación (121), es el siguiente:

$$\begin{aligned} H_{\text{NN}}(\omega) &= \frac{T_r(\omega) R^*(\omega)}{|T_r(\omega)|^2 + \frac{1}{2\pi} B_r^0(\omega) \bullet |\bar{W}(\omega)|^2 + N_r(\omega)} \\ &\times \frac{T_s^*(\omega)}{|T_s(\omega)|^2 + \frac{1}{2\pi} B_s^0(\omega) \bullet |\bar{W}(\omega)|^2 + N_s(\omega)}. \end{aligned} \quad (159)$$

Cuando la imagen de referencia contiene al objeto sin la presencia de ruido, la respuesta de frecuencia del filtro es exactamente la del GOF para el modelo disjunto (Javidi y Wang, 1994a). Por este motivo, se le llama al filtro dado en la ecuación (159) el Filtro Óptimo Generalizado para el modelo NN ( $\text{GOF}_{\text{NN}}$ ).

### V.3.1. Estimación de parámetros

El  $\text{GOF}_{\text{NN}}$  requiere conocimiento tanto de la transformada de Fourier del objeto de interés como de su región de soporte. Como estos datos no están disponibles bajo las suposiciones del modelo NN, es necesario estimarlas a partir de la imagen de referencia. Aplicando un filtro Wiener de suavizado (Pratt, 2007) se pueden atenuar los efectos del ruido aditivo y del fondo disjunto en la imagen de referencia. Si  $\tilde{r}(x)$  denota la imagen de referencia después de filtrar con Wiener, se puede estimar la región de soporte del objeto de la siguiente manera:

$$\hat{w}(x - x_r) = \begin{cases} 1 & \mu_r \geq \mu_t \text{ y } \tilde{r}(x) > \tau(\tilde{t}, \tilde{b}_r, \tilde{n}_r) \\ 1 & \mu_r < \mu_t \text{ y } \tilde{r}(x) < \tau(\tilde{t}, \tilde{b}_r, \tilde{n}_r) \\ 0 & \text{de lo contrario} \end{cases}, \quad (160)$$

donde  $\tilde{t}(x)$ ,  $\tilde{b}_r(x)$  y  $\tilde{n}_r(x)$  son las señales que componen la imagen de referencia después de filtrar con Wiener y  $\tau$  es el umbral óptimo que depende de las estadísticas de estas señales. Se utiliza el umbral óptimo para separar las distribuciones de  $\tilde{t}(x) + \tilde{n}_r(x)$  y  $\tilde{b}_r(x) + \tilde{n}_r(x)$ , es decir, de las áreas dentro y fuera del objeto, respectivamente, después de filtrar.

Suponiendo que los procesos de ruido son aproximadamente normales después de filtrar, el umbral óptimo para determinar la región inversa de soporte del objeto se puede determinar como el óptimo para separar dos distribuciones Gaussianas. Las densidades espectrales de  $\tilde{b}_r^0(x)$  y  $\tilde{n}_r(x)$  se pueden calcular a partir de las densidades espectrales conocidas *a priori* de  $b_r^0(x)$  y  $n_r(x)$  y la respuesta de frecuencia  $G_1(\omega)$  del filtro Wiener

(Kumar *et al.*, 2005). Esto es

$$\tilde{B}_r^0(\omega) = |G_1(\omega)|^2 B_r^0(\omega), \quad (161)$$

$$\tilde{N}_r(\omega) = |G_1(\omega)|^2 N_r(\omega). \quad (162)$$

Una vez obtenida la estimación de la región inversa de soporte, se puede estimar la apariencia del valor esperado de la escena como

$$t_s(x - x_r) \approx \tilde{r}(x) + (\mu_s - \mu_r) \hat{w}(x - x_r). \quad (163)$$

Esta estimación se usa para aproximar la respuesta de frecuencia del  $\text{GOF}_{\text{NN}}$  y se denota por  $\text{NN}_1$  la estimación construida de esta manera.

La estimación de la región inversa de soporte del objeto aplicando un umbral se basa en la suposición de que después de filtrar con Wiener las intensidades dentro y fuera de la región del objeto en la imagen de referencia tienen diferentes distribuciones estadísticas. Sin embargo, si las estadísticas originales del objeto son muy similares a las del fondo, se vuelve difícil obtener una estimación confiable mediante la aplicación de un umbral. A continuación se ilustran los errores de clasificación usando el objeto en la figura 11(a) y realizaciones del fondo estocástico similares al de la figura 11(b). Se utilizan imágenes estocásticas para asegurar que se cumple la suposición del modelo de que los procesos son estacionarios. La figura 12(a) ilustra el número de píxeles mal clasificados al variar la media y la desviación estándar del fondo en la imagen de referencia. Para un valor fijo de la desviación estándar, en la figura 12(b) se muestra el número de píxeles mal clasificados al variar las medias de los fondos tanto de la imagen de referencia como de la escena de entrada. De hecho, una cota inferior sobre la probabilidad de clasificar mal

un pixel se puede calcular como sigue (Avi-Itzhak y Diep, 1996; Duda *et al.*, 2000):

$$p_\epsilon = \sup_x \left\{ \int_{-\infty}^{\infty} \frac{A_t f_t(\nu) + (1 - A_t) f_b(\nu)}{\beta} \times \ln \left[ \frac{1 + \exp(-\beta)}{\exp(-\beta P(x, \nu)) + \exp(-\beta P(1 - P(x, \nu)))} \right] d\nu \right\}, \quad (164)$$

donde  $A_t$  es la razón del tamaño del objeto al tamaño de la escena;  $f_t(\nu)$  y  $f_b(\nu)$  son las funciones de densidad de probabilidad de los valores de intensidad en las regiones del objeto y el fondo, respectivamente;  $\beta$  es una constante positiva arbitraria, y  $P(x, \nu) = \Pr\{\hat{w}(x) = 0 | \tilde{r}(x) = \nu\}$  es la probabilidad condicional de clasificar un pixel como parte del objeto dado que su intensidad después de filtrar es  $\nu$ . Además, al estimar  $t_s(x - x_r)$  usando la ecuación (163) los errores en la estimación de la región de soporte se amplifican por un factor  $\Delta\mu = \mu_s - \mu_r$ . La figura 12(c) muestra una representación de los errores de la estimación de la forma del objeto después de multiplicar por la diferencia de las medias.

En la figura 13(a) se muestran los valores de error de correspondencia entre el objeto compuesto ideal  $t_s(x)$  y la estimación (163) mientras se varían las medias de los fondos en la imagen de referencia y la escena de entrada y se mantienen fijas las estadísticas del objeto. Las regiones con un sombreado claro y oscuro indican valores bajos y altos de ME, respectivamente. Se puede observar que el error aumenta repentinamente al cambiar el signo de  $\Delta\mu$ . Esto se debe a la inercia lineal del filtro Wiener. Después de filtrar, los valores del objeto cerca de los bordes tienden hacia la media del fondo de la imagen de referencia. Los errores en la estimación de la región inversa de soporte también tienden a incrementarse cerca de los bordes. Por lo tanto, el término  $\Delta\mu\hat{w}(x - x_r)$  en la estimación del objeto compuesto produce muchos errores cerca de los bordes, mismos que son la parte más importante para la correspondencia después de blanquear las

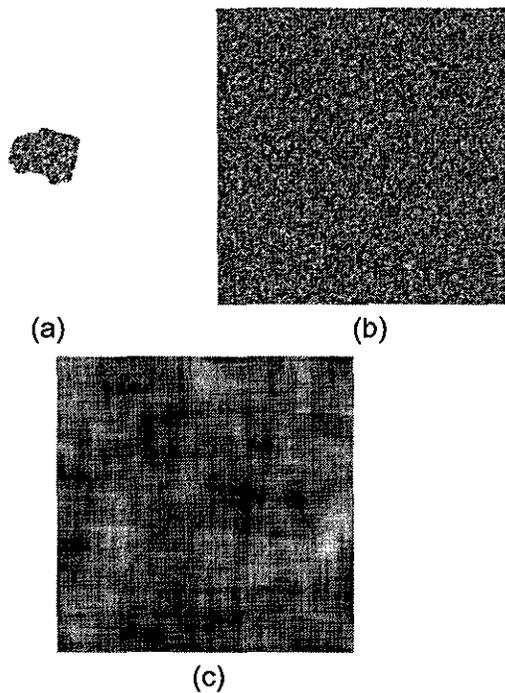


Figura 11: Imágenes estocásticas usadas en los experimentos: (a) objeto con apariencia aleatoria, (b) fondo de la imagen de referencia, (c) fondo de la escena de entrada.

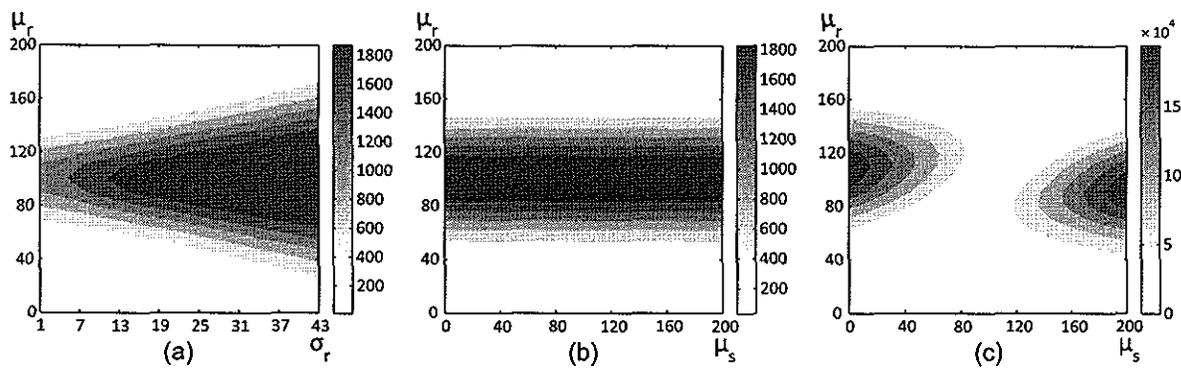


Figura 12: (a) Errores de estimación de forma al variar  $\mu_r$  y  $\sigma_r$ ; (b) errores de estimación de forma al variar  $\mu_r$  y  $\mu_s$ , y (c) errores mostrados en (b) después de amplificar por  $|\mu_s - \mu_r|$ .

señales. Hay dos maneras de abordar este problema. La primer manera es procesar la escena de entrada con la misma instancia del filtro Wiener que se use para la imagen de referencia. El propósito de este procedimiento es el suavizar los bordes del objeto en la escena de entrada en una forma similar a como sucede en la imagen de referencia. En la figura 13(b) se muestran los valores de ME entre la estimación del objeto compuesto y el valor esperado de la escena después de filtrar con Wiener. Se puede observar una mejora significativa con respecto a los errores mostrados en figura 13(a), en particular alrededor de la línea  $\mu_s = \mu_r$ . Otra forma de abordar el problema es el posponer el filtrado Wiener de la imagen de referencia hasta después de formar la estimación de la escena de entrada. Esto es,  $t_s(x)$  se puede estimar como  $t_s(x - x_r) \approx g_2(x) \bullet [r(x) + \hat{w}(x - x_r)]$ . En la figura 13(c) se muestra el valor de ME entre esta estimación y el valor esperado de la escena de entrada. Es claro que los errores de correspondencia son menores que usando la estimación original. Aunque el filtro Wiener es lineal, la construcción de una instancia del filtro Wiener empírico (Yaroslavsky, 2003) requiere de una estimación no lineal del espectro de potencia de la imagen de referencia, es decir

$$G_2(\omega) = \text{máx} \left\{ 0, 1 - \frac{\frac{1}{2\pi} B_r^0(\omega) \bullet |\hat{W}(\omega)|^2 + N_r(\omega)}{|R(\omega)|^2} \right\}, \quad (165)$$

de manera que se suprimen las frecuencias en las que el ruido es más fuerte que la señal de interés. Cuando se realiza la estimación del objeto compuesto la aplicación del filtro Wiener después de corregir el fondo suprime estas frecuencias de la salida del filtro. Si se suma la corrección del fondo después de aplicar el filtro Wiener, la información en las frecuencias ruidosas está presente también en la salida del filtro. En conclusión, debido a la naturaleza no lineal del filtro Wiener empírico el orden en que se aplican las operaciones se vuelve importante.

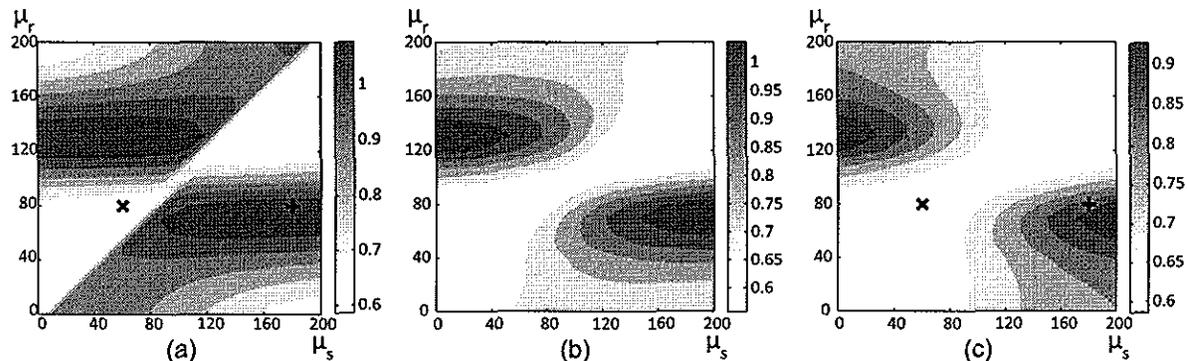


Figura 13: Errores de correspondencia al variar  $\mu_r$  y  $\mu_s$ : (a) entre  $\hat{t}_s^1(x)$  y  $t_s(x)$ ; (b) entre  $\hat{t}_s^2(x)$  y  $t_s(x)$  con preprocesamiento, y (c) entre  $\hat{t}_s^2(x)$  y  $t_s(x)$ .

### V.3.2. Resultados experimentales

En esta sección se evalúa el desempeño del filtro óptimo y estimaciones propuestas para detección de objetos en un fondo no traslapado mediante simulaciones por computadora. Los criterios de desempeño son el de capacidad de discriminación y errores de localización. Se utilizan imágenes de  $256 \times 256$  píxeles con intensidades en el intervalo  $[0-255]$ . Se consideran dos tipos de imágenes: imágenes estocásticas en las que se cumple perfectamente la suposición de distribuciones estacionarias, e imágenes de la vida real. Las imágenes estocásticas son realizaciones de proceso estacionarios correlacionados. El objeto de interés estocástico se muestra en la figura 11(a), tiene media 100 y desviación estándar 25. Ejemplos de fondos de la imagen de referencia y de la escena de entrada en las figuras 11(b) y 11(c). El fondo en la imagen de referencia tiene un factor de correlación de 0.5, mientras que el fondo en la escena de entrada tiene factor de 0.95. Las imágenes determinísticas que se utilizan se muestran en la figura 14. El objeto de interés se muestra en la figura 14(a). Las imágenes que se muestran en las figuras 14(b) y 14(c) se usan como fondo en la imagen de referencia y en la escena de entrada, respectivamente.

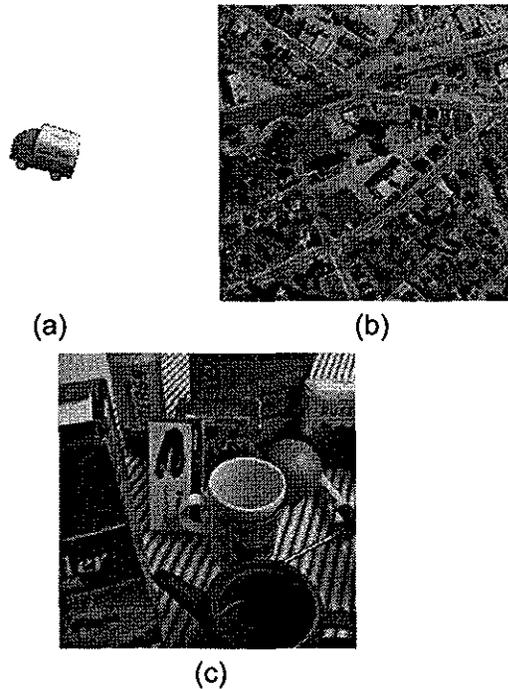


Figura 14: Imágenes reales usadas en los experimentos: (a) objeto de interés, (b) fondo de la imagen de referencia, (c) fondo de la escena de entrada.

### Efectos del ruido aditivo en la imagen de referencia

Primero se analiza la influencia que tiene ruido aditivo en la imagen de referencia sobre el desempeño de los filtros en términos de DC y LE. Las estadísticas del objeto de y los fondos se mantienen fijas mientras se varía la desviación estándar del ruido aditivo en la referencia desde 5 hasta 50. Las simulaciones ilustran dos situaciones; cuando cada una de las estimaciones presenta mejor desempeño que la otra. El desempeño de los filtros en términos de DC y LE se muestra en las figuras 15(a) y 15(b), respectivamente, cuando los valores de las medias son  $\mu_r = 80$  y  $\mu_s = 60$ . Este punto del espacio de parámetros se muestra en las figuras 13(a) y 13(c) con una marca en forma de  $\times$ . Los errores de correspondencia para las estimaciones  $NN_1$  y  $NN_2$  son 0.6 y 0.65, respectivamente. Se puede observar que ambas estimaciones del filtro óptimo son capaces de detectar al objeto en presencia de ruido aditivo con desviación estándar

de hasta 60. El  $AN_1$  presenta un desempeño ligeramente mejor que el  $AN_2$ , aunque el efecto del ruido es el mismo sobre ambos filtros. Para otro conjunto de simulaciones las medias de los fondos se escogen como  $\mu_r = 80$  y  $\mu_s = 180$ . Este punto del espacio de parámetros se muestra con una marca en forma de + en las figuras 13(a) y 13(c). El desempeño del filtro  $NN_1$  y del filtro  $NN_2$  es de 1.1 y 0.85, respectivamente. Los criterios DC y LE para este caso se muestran en las figuras 15(c) y 15(d), respectivamente. El filtrado Wiener produce errores cerca de los bordes del objeto, por lo que degrada el desempeño del  $NN_1$ . Se puede observar en la figura 15 que para valores de 10 o mayores de la desviación estándar del ruido, el filtro  $NN_1$  es incapaz de localizar adecuadamente al objeto. Es interesante notar que el cambio de medias no afecta significativamente el desempeño del filtro  $NN_2$  y su desempeño es consistente con el del caso anterior.

### **Efectos de las medias de los fondos en la imagen de referencia y en la escena de entrada**

Ahora se analiza el efecto de variar las medias de los fondos de la imagen de referencia y de la escena de entrada. Es claro que la calidad de la restauración de la forma del objeto empeora cuando el valor de  $|\mu_r - \mu_t|$  disminuye. Primero se evalúa el desempeño de los filtros utilizando imágenes estocásticas que cumplen exactamente las suposiciones de que los procesos en las imágenes son estacionarios. Se fijan las desviaciones estándar del ruido y de los fondos, mientras que se varían las medias de los fondos desde 0 hasta 200. El desempeño de los filtros en términos de DC se muestra en la figura 16. El  $NN_1$ ; el  $NN_1$  aplicado a la escena de entrada procesada con el filtro Wiener, y el  $NN_2$ , se muestran en las figuras 16(a), 16(b) y 16(c), respectivamente. Las regiones del espacio de parámetros que se muestran con sombreado claro indican un buen desempeño de detección, mientras que las regiones con sombreado oscuro indican errores de detección.

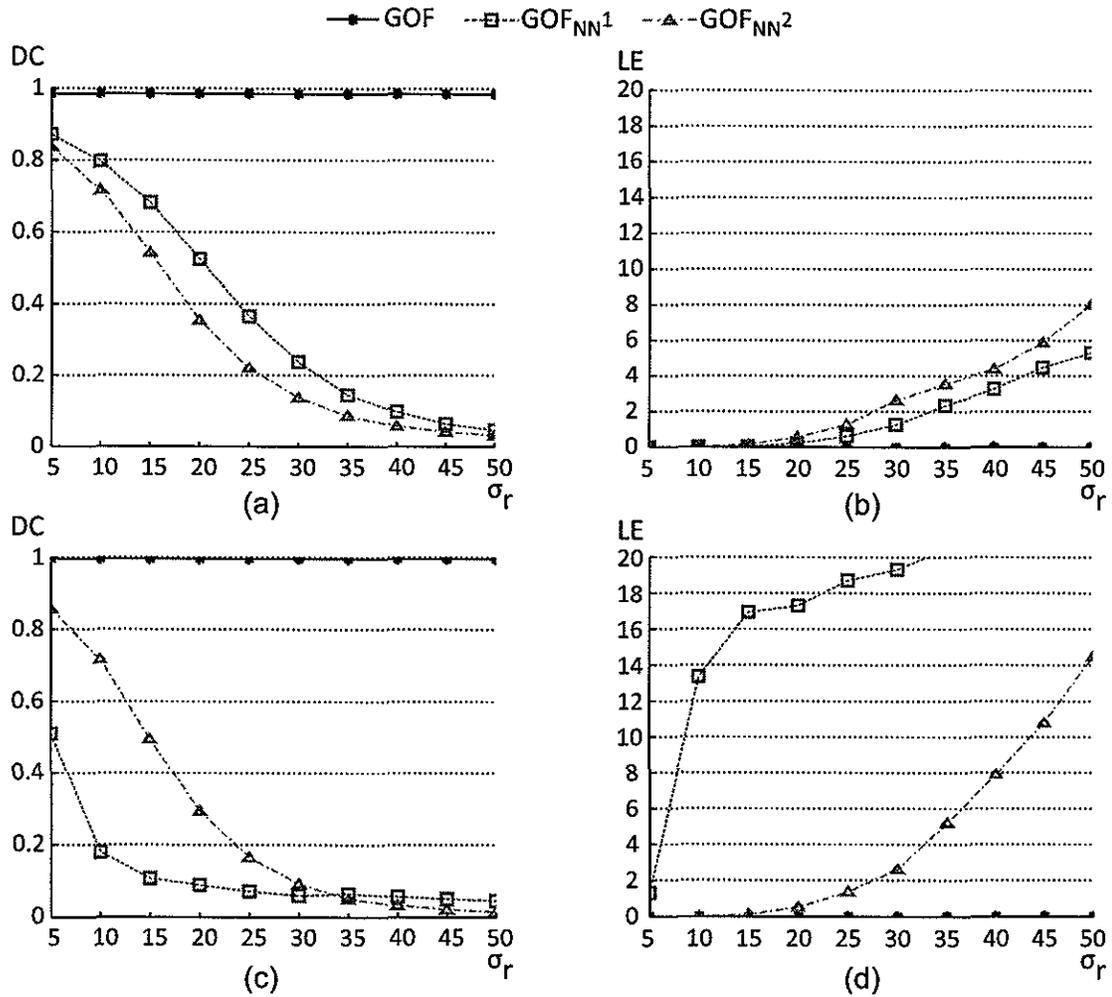


Figura 15: Desempeño de los filtros al variar la desviación estándar del ruido aditivo en la imagen de referencia, en términos de (a) DC y (b) LE cuando  $\mu_r = 80$  y  $\mu_s = 60$ , y (c) DC y (d) LE cuando  $\mu_r = 80$  y  $\mu_s = 180$ .

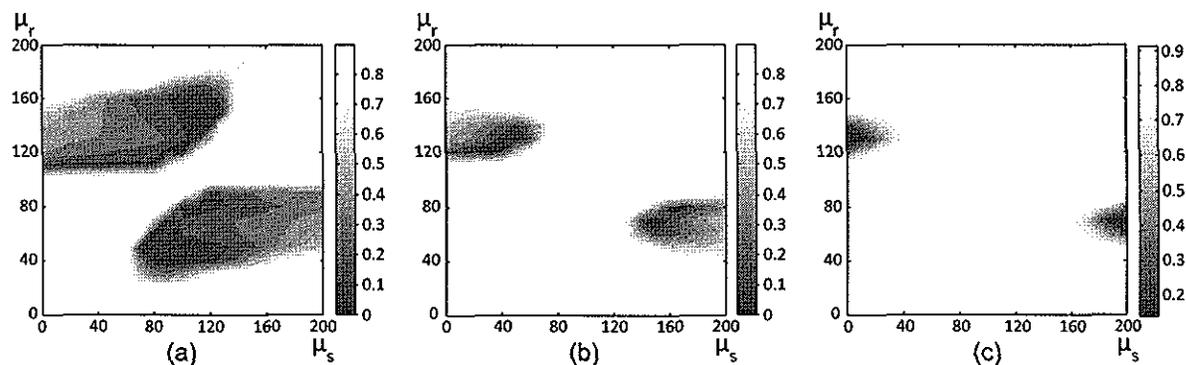


Figura 16: Desempeño de los filtros en términos de DC al variar  $\mu_r$  y  $\mu_s$  usando imágenes estocásticas: (a)  $NN_1$  aplicado a la escena de entrada, (b)  $NN_1$  aplicado a la escena de entrada preprocesada con el filtro Wiener y (c)  $NN_2$  aplicado a la escena de entrada.

Es interesante notar que las regiones en las que los filtros fallan corresponden a las regiones en las que los errores de correspondencia son altos (véase la figura 13). Se puede observar que el filtro  $NN_2$  logra un mejor desempeño en una mayor región del espacio de parámetros.

El siguiente experimento se realiza con imágenes determinísticas. Los resultados de las simulaciones se muestran en las figuras 17(a), 17(b) y 17(c) para los filtros  $NN_1$ ,  $NN_1$  con preprocesamiento Wiener y  $NN_2$ , respectivamente. Se puede observar que las regiones del espacio de parámetros donde los filtros fallan al reconocer el objeto es mayor que para el caso de imágenes estocásticas. Esto se debe a la naturaleza no estacionaria de las imágenes que limita la efectividad del filtro Wiener. Se puede observar también que el desempeño del filtro  $NN_1$  con preprocesamiento es casi igual al del filtro  $NN_2$ , y definitivamente superior al filtro  $NN_1$  aplicado directamente. Se puede concluir que el desempeño de detección de los filtros propuestos con imágenes no homogéneas depende mucho de la calidad del filtrado. Esta se puede mejorar utilizando una técnica más apropiada (como filtrado localmente adaptativo). en lugar de filtrado Wiener, que falla en estos casos.

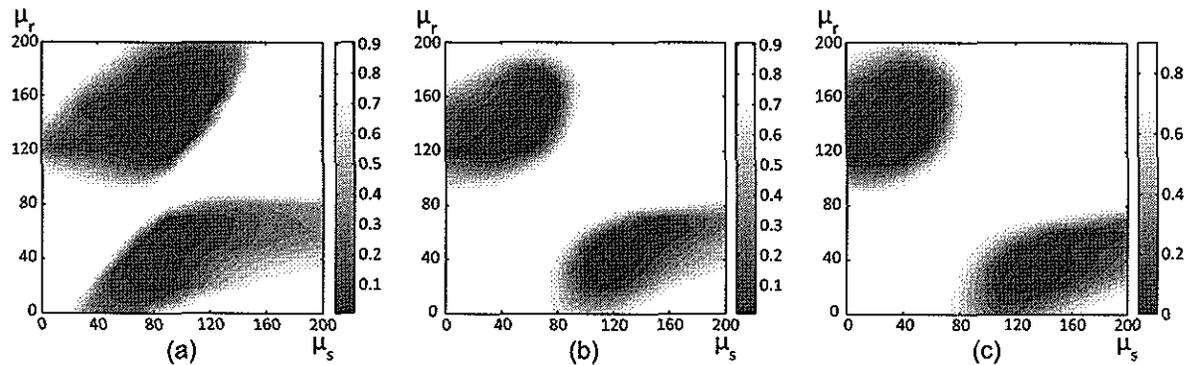


Figura 17: Desempeño de los filtros al variar  $\mu_r$  y  $\mu_s$  usando imágenes determinísticas: (a)  $NN_1$  aplicado a la escena de entrada, (b)  $NN_2$  aplicado a la escena de entrada preprocesada con el filtro Wiener y (c)  $NN_2$  aplicado a la escena de entrada.

### Efecto de la desviación estándar del fondo en la imagen de referencia

Finalmente se analiza el desempeño de los filtros cuando varía la desviación estándar del fondo en la imagen de referencia. El desempeño de los filtros  $NN_1$  y  $NN_2$  se muestra respectivamente, en las filas superior e inferior de la figura 18. La desviación estándar del fondo toma valores de 15, 30 y 45 en la primer, segunda y tercer columna de la figura 18, respectivamente. Para cada imagen, las medias de los fondos se varían desde 0 hasta 200. La desviación estándar del ruido aditivo en la imagen de referencia y en la escena de entrada se mantienen fijas. Se puede observar que el  $NN_2$  es más tolerante a las variaciones en el fondo que el  $NN_1$ . Como es de esperarse, mayores niveles de ruido hacen que crezca la región del espacio de parámetros donde los filtros fallan al reconocer el objeto. Sin embargo, es interesante notar que las tendencias no cambian. Es decir, las regiones del espacio en las que la detección falla se determina por los valores de las medias de los fondos, mientras que la desviación estándar del fondo degrada el desempeño de igual manera para las dos estimaciones propuestas.

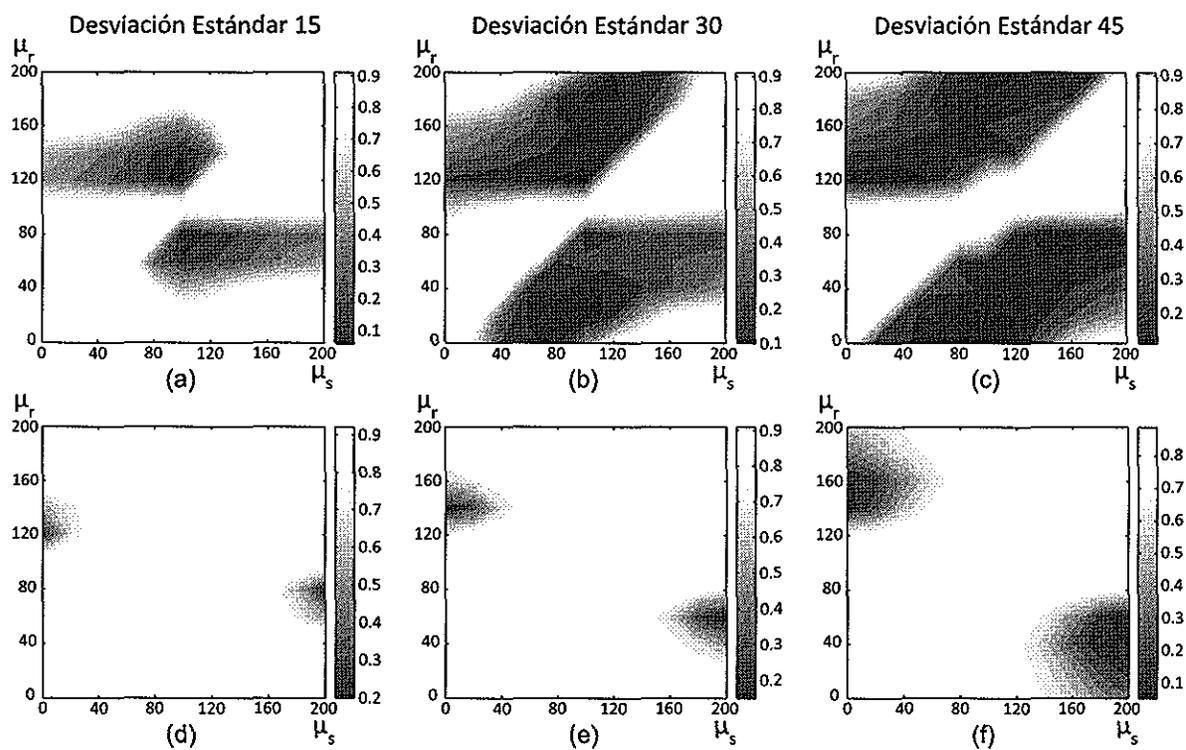


Figura 18: Desempeño de los filtros en términos de DC en el espacio de parámetros  $(\mu_r, \mu_s)$  variando la desviación estándar del fondo en la imagen de referencia. En el renglón superior  $NN_1$  y en el renglón inferior  $NN_2$ . La desviación estándar del fondo de la imagen de referencia es (a), (d) 15; (b), (e) 30, y (c), (f) 45.

## V.4. Resumen

En este capítulo se presenta el desarrollo de filtros de correlación para detectar y localizar objetos en escenas con fondo disjunto. El diseño de filtros se hace a partir de un nuevo modelo de señales que considera que el objeto de interés se conoce únicamente en una escena de referencia contaminada con ruido. Se utilizan modelos para la imagen de referencia, el aditivo y no traslapado. Su supone que se conoce información estadística sobre los procesos de ruido y los fondos disjuntos en las escenas. Los resultados para los modelos sencillos de escenas contaminadas con ruido aditivo se utilizan en el desarrollo de los filtros para escenas con fondo disjunto. Para cada modelo utilizado se proponen dos formas de estimar los filtros óptimos derivados. Las técnicas propuestas permiten la detección y localización automática de objetos sin necesidad de procesamiento manual de imágenes de referencia por parte de un operador humano.

Se muestra mediante simulaciones por computadora que los filtros propuestos y usando las estimaciones sugeridas, son capaces de detectar el objeto en presencia de niveles moderados de ruido. Se presenta la métrica de errores de correspondencia que tiene buena correlación con el desempeño de los filtros en cuanto a capacidad de discriminación. Se muestra que esta métrica se puede utilizar para predecir adecuadamente qué estimación de cada filtro utilizar dependiendo de las estadísticas de las imágenes en la aplicación. Se demuestra optimizar que la métrica clásica de error cuadrático medio para la estimación de la apariencia del objeto no necesariamente logra el mejor desempeño de reconocimiento.

Se muestra que la técnica de estimación utilizada es importante para el reconocimiento de objetos. Para imágenes espacialmente homogéneas el filtrado de Wiener es suficiente y los métodos propuestos tienen un buen desempeño para la mayor parte

---

del espacio de parámetros de los modelos considerados. Sin embargo, para imágenes no homogéneas es conveniente utilizar técnicas localmente adaptativas.

## Capítulo VI

---

# Diseño de filtros adaptativos para el reconocimiento de objetos mediante imágenes ruidosas de referencia

---

Los filtros de correlación utilizan la información de apariencia y forma del objeto de interés para llevar a cabo el reconocimiento. Por este motivo, su desempeño se degrada rápidamente si la apariencia del objeto en la escena de interés difiere de la apariencia utilizada en el proceso de diseño. La apariencia del objeto en la escena puede variar debido a distorsiones geométricas, tales como rotaciones, cambios de escala y cambios de perspectiva.

El propósito de este capítulo es proponer filtros adaptativos de correlación para el reconocimiento de objetos utilizando imágenes ruidosas de referencia. El diseño de filtros compuestos se basa en un algoritmo modificado de funciones discriminantes sintéticas. Se utiliza un algoritmo iterativo para diseñar filtros de correlación con una capacidad de discriminación deseada tomando en cuenta múltiples imágenes de referencia, información estadística *a priori* sobre los procesos de ruido y fondos disjuntos en las escenas.

## VI.1. Filtros compuestos de respuestas de impulso con adaptación a fondos típicos

En esta sección se utiliza información sobre el objeto de interés, el fondo en la escena y el ruido aditivo para diseñar un filtro de correlación adaptativo compuesto. El filtro se diseña para producir un pico de correlación afilado en la presencia del objeto de interés y valores cercanos a cero en el área del fondo. El fondo de la escena se puede describir en forma estocástica (como la realización de un proceso estocástico) o en forma determinística (como la imagen de un fondo típico). El modelo de señales que se utiliza es el modelo AN, en el que las imágenes de referencia usadas para diseñar el filtro se suponen contaminadas por ruido aditivo, mientras que el filtro se diseña para detectar un objeto en una escena con un fondo disjunto. Dado un conjunto de imágenes de referencia  $r^i(x)$ , el modelo de señales usado es el siguiente:

$$r^i(x) = t^i(x) + n_r^i(x), \quad (166)$$

$$s(x) = t^i(x - x_s) + b_s(x) \bar{w}^i(x - x_s) + n_s(x), \quad (167)$$

donde  $t^i(x)$ ,  $\bar{w}^i(x)$  y  $n_r^i(x)$  denotan la apariencia del objeto, la región de soporte correspondiente, y la instancia de ruido aditivo en la  $i$ -ésima imagen de referencia, respectivamente. Para este modelo de señales, la respuesta de frecuencia del filtro óptimo está dada por

$$H^i(\omega) = \frac{T^i(\omega) T_s^i(\omega) R^{i*}(\omega)}{(|T^i(\omega)|^2 + N_r^i(\omega)) (|T_s^i(\omega)|^2 + \tilde{N}_s^i(\omega))}, \quad (168)$$

donde  $T^i(\omega)$ ,  $T_s^i(\omega)$ ,  $R^i(\omega)$  son las transformadas de Fourier de  $t^i(x)$ ,  $t_s^i(x) = t^i(x) + \mu_s \bar{w}^i(x)$  y  $r^i(x)$ , respectivamente, y  $N_r^i(\omega)$  y  $N_s^i(\omega)$  son las densidades espectrales de

la imagen de referencia y de la escena de entrada, respectivamente.

Sea  $h^i(x)$  la transformada inversa de la respuesta de frecuencia estimada del filtro óptimo obtenido a partir de la  $i$ -ésima imagen de referencia. Se puede diseñar un filtro SDF como combinación lineal del conjunto  $\{h^i(x), i = 1, 2, \dots, k\}$  donde  $k$  es la cantidad de imágenes de entrenamiento disponibles. Los coeficientes  $\alpha_i$  de la combinación lineal se seleccionan para satisfacer valores de salida prefijados para cada imagen de entrenamiento. Sea  $\{p^i(x), i = 1, 2, \dots, m\}$  el conjunto de patrones a rechazar. Sea  $\mathbf{S} = \{t_s^1(x), \dots, t_s^k(x), p^1(x), \dots, p^m(x)\}$  el conjunto de  $k + m$  patrones de entrenamiento y sea  $\mathbf{H} = \{h^1(x), \dots, h^k(x), p^1(x), \dots, p^m(x)\}$  el conjunto de respuestas de impulso y de patrones a rechazar. El filtro deseado se compone de la siguiente manera:

$$h_{\text{SDF}}(x) = \sum_{i=1}^k \alpha_i h^i(x) + \sum_{i=k+1}^{k+m} \alpha_i p^{i-k}(x). \quad (169)$$

Usando notación matriz-vector se denota por  $\mathbf{R}$  una matriz con  $k + m$  columnas, donde las columnas de  $\mathbf{R}$  son los elementos de  $\mathbf{H}$  en forma de vector. Si el vector de coeficientes es  $\mathbf{a} = [\alpha_1, \dots, \alpha_{k+m}]^T$  (donde el superíndice  $T$  denota transpuesta), la ecuación (169) se puede reescribir como

$$\mathbf{h}_{\text{SDF}} = \mathbf{R}\mathbf{a}. \quad (170)$$

El vector de respuestas deseadas para cada patrón de entrenamiento se denota por  $\mathbf{u}$  y se fija como

$$\mathbf{u} = \left[ \underbrace{1, \dots, 1}_{k \text{ unos}}, \underbrace{0, \dots, 0}_{m \text{ ceros}} \right]^T. \quad (171)$$

Si la matriz  $\mathbf{Q}$  se compone con los elementos de  $\mathbf{S}$  como columnas, las restricciones de

la salida del filtro se pueden escribir como

$$\mathbf{u} = \mathbf{Q}^+ \mathbf{h}_{\text{SDF}}, \quad (172)$$

donde  $^+$  denota transpuesta conjugada. Usando las ecuaciones (170) y (172), la solución para  $\mathbf{h}_{\text{SDF}}$  es

$$\mathbf{h}_{\text{SDF}} = \mathbf{R} [\mathbf{Q}^+ \mathbf{R}]^{-1} \mathbf{u}. \quad (173)$$

El valor del pico de correlación al usar el filtro en la ecuación (173) se espera que sea cerca de la unidad para objetos de la clase verdadera y cerca de cero para objetos de la clase falsa. Sin embargo, dado que no se controla en forma explícita todo el plano correlación, es posible que ocurran lóbulos laterales muy grandes. Para reducir la probabilidad de falsas alarmas, se utiliza un algoritmo iterativo para reducir la intensidad de la salida del filtro para un fondo típico dado. En cada paso, el filtro se aplica al fondo típico y se sintetiza un patrón de entrenamiento a rechazar en la posición del lóbulo lateral de mayor intensidad. Un diagrama del algoritmo de adaptación se muestra en la figura 19. El algoritmo resume de la siguiente manera:

1. Componer un filtro SDF utilizando los patrones de entrenamiento que incluyen las posibles vistas del objeto conocidas *a priori* utilizando la ecuación (173).
  2. Realizar la correlación del filtro con un fondo típico y calcular el valor de la DC utilizando la ecuación (64).
  3. Si el valor de la DC es mayor o igual que un valor prefijado, el algoritmo termina. De lo contrario, se continuar al paso 4.
  4. Sintetizar un patrón para rechazar en la posición del lóbulo lateral de mayor valor y se agrega al conjunto de entrenamiento.
-

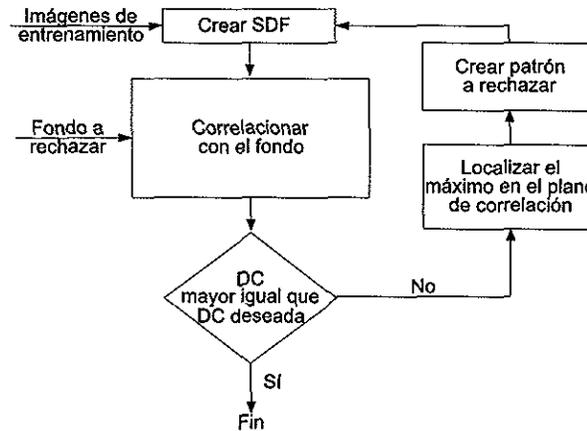


Figura 19: Diagrama del algoritmo para el diseño de filtros adaptativos.

5. Continuar al paso 1.

### VI.1.1. Resultados experimentales

Para analizar el desempeño del filtro adaptativo propuesto, se realizaron simulaciones por computadora para distintas rotaciones, escalamiento e intensidades de ruido, mientras varía la ubicación del objeto en la escena. Se analizan los filtros en términos de tolerancia a distorsiones y tolerancia a ruido. Los resultados de las simulaciones se presentan a continuación. En las simulaciones siguientes, el filtro adaptativo propuesto se denota por  $AGOF_{AN}$ . La DC deseada para el proceso de adaptación se fija en 0.9. Para los experimentos realizados el procedimiento de adaptación requiere de menos de 10 iteraciones para convergir a la DC deseada. El filtro diseñado con el algoritmo propuesto en la sección V.2 se denota por  $GOF_{AN}$ , esto es, el filtro que se determina que tendrá mejor desempeño entre el  $AN_1$  y el  $AN_2$ .

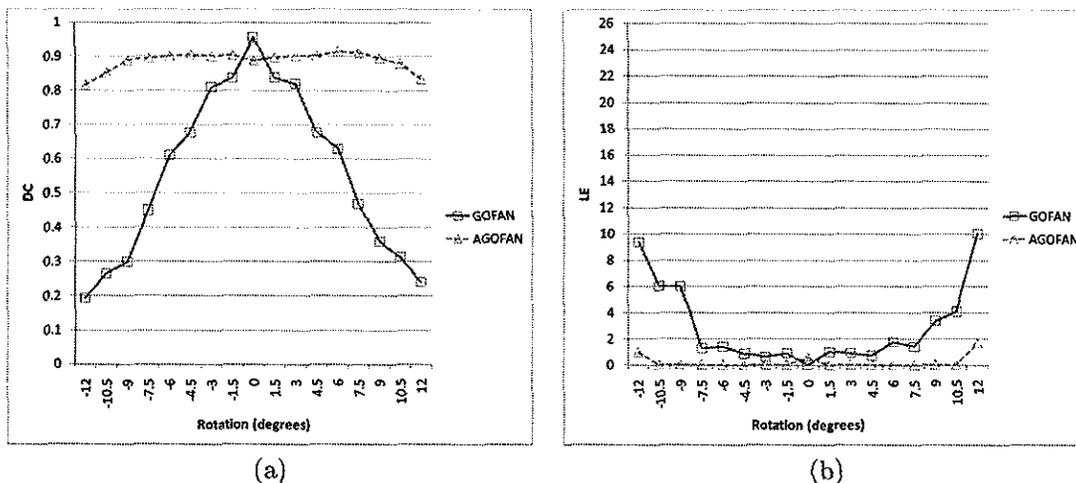


Figura 20: Desempeño de los filtros compuestos, en términos de (a) DC y (b) LE cuando se varía la rotación del objeto. La desviación estándar del ruido en las imágenes de entrenamiento es 10.

### Tolerancia a distorsiones geométricas

Primero se investiga la tolerancia que presentan los filtros a rotaciones del objeto. El objeto aparece sin cambios de escala en la escena y se fija el valor de la desviación estándar del ruido en las imágenes de entrenamiento en 10. La rotación se varía de  $-12^\circ$  a  $12^\circ$  en incrementos de  $1.5^\circ$ . Cabe destacar que el conjunto de rotaciones que se usan en la escena no pertenecen al conjunto de rotaciones que se usaron para el diseño del filtro. Los resultados de las simulaciones se muestran en la figura 20(a) para el criterio DC. Se puede observar que el filtro  $GOF_{AN}$  tiene mejor desempeño que el filtro  $AGOF_{AN}$  sólo en el caso en el que el objeto aparece sin rotación. El desempeño del  $GOF_{AN}$  decae rápidamente conforme la rotación del objeto aumenta. el desempeño del filtro  $AGOF_{AN}$  sólo baja para rotaciones fuera del intervalo de entrenamiento. Los resultados para el criterio LE se muestran en la figura 20(b). Estos son consistentes con los resultados para el criterio DC. Los errores de localización se vuelven grandes para rotaciones más allá de los  $7.5^\circ$ .

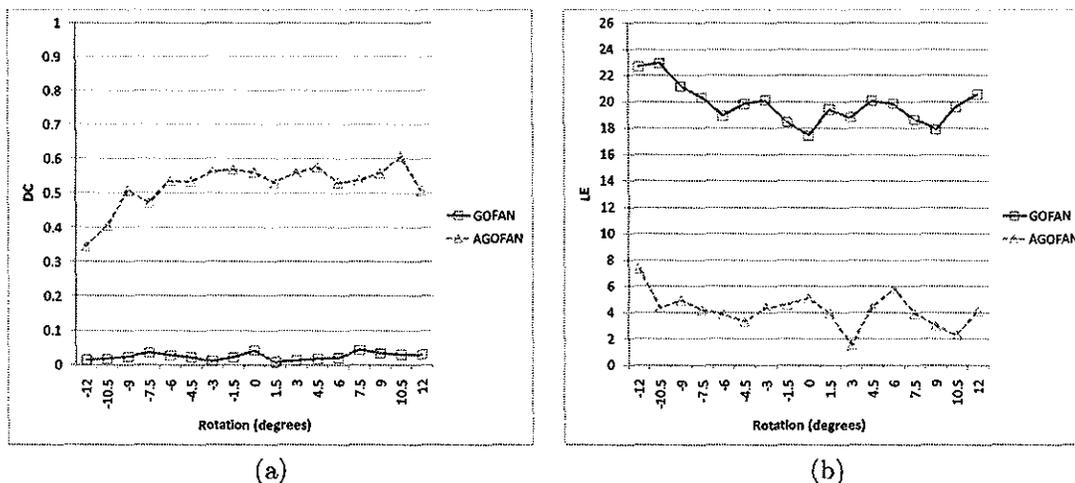


Figura 21: Desempeño de los filtros propuestos en términos de (a) DC y (b) LE cuando se varía la rotación del objeto. El objeto aparece escalado a 0.7 veces su tamaño original y la desviación estándar del ruido en las imágenes de referencia es 15.

Ahora se considera una situación en la que el reconocimiento es más difícil. Para los resultados que se muestran en la figura 21, el objeto se escala a 0.7 veces su tamaño original y la desviación estándar del ruido en las imágenes de entrenamiento se aumenta hasta 15. El  $GOF_{AN}$  falla completamente al detectar el objeto, mientras que el  $AGOF_{AN}$  todavía es capaz de detectarlo. Los valores de DC que se muestran en la figura 21(a) son más bajos que para el caso anterior debido al incremento de la intensidad del ruido y a que el valor de escala está fuera del intervalo de valores que se usó para el entrenamiento. Los valores de LE que se muestran en la figura 21(b) indican que la capacidad para localizar el objeto se ve afectada, pero que las estimaciones de localización siguen cercanas a la verdadera ubicación del objeto.

Para caracterizar el impacto de cambios de escala sobre el desempeño de los filtros, se realizaron simulaciones en las que se varía la escala desde 0.7 hasta 1.3 veces el tamaño original del objeto, en incrementos de 0.15. La desviación estándar del ruido en las imágenes de entrenamiento se fija en 10. En las figuras 22(a) y 22(b) se muestran los resultados de las simulaciones para los criterios DC y LE, respectivamente. Se puede

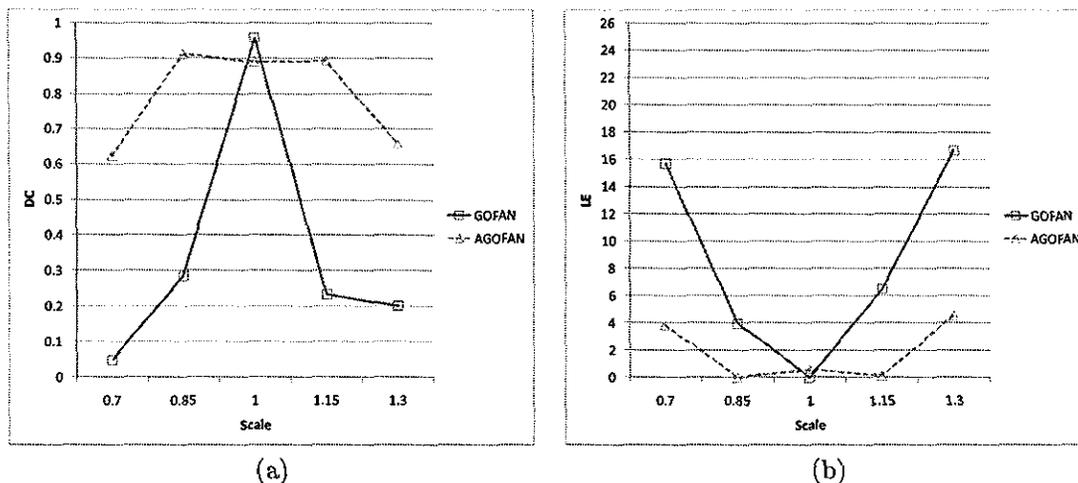


Figura 22: Desempeño de los filtros propuestos en términos de (a) DC y (b) LE cuando se varía la escala del objeto. La desviación estándar del ruido en las imágenes de entrenamiento es 10.

observar que el desempeño del filtro  $GOF_{AN}$  decae rápidamente conforme el escalamiento del objeto cambia. El filtro  $AGOF_{AN}$  es capaz de detectar el objeto aún en casos en los que la distorsión está fuera del intervalo de entrenamiento. Es interesante notar que los cambios de escala tienen un impacto más significativo sobre el desempeño final de los filtros que las rotaciones. Esto se refleja claramente en los resultados de errores de localización: los errores máximos de localización al variar la escala llegan hasta 4 píxeles, mientras que el error máximo al variar rotación sólo llega a 2 píxeles.

### Tolerancia a ruido en las imágenes de entrenamiento

Ahora se analiza el efecto que tiene el ruido en las imágenes de entrenamiento sobre el desempeño de los filtros. Primero se muestran experimentos en los que el objeto aparece sin distorsiones geométricas. La desviación estándar del ruido en las imágenes de entrenamiento se varía de 5 a 20 en incrementos de 5. Los resultados de las simulaciones se muestran en las figuras 23(a) y 23(b) en términos de DC y LE, respectivamente. Nótese que los valores de la DC son cercanos a 0.9 y los valores de LE

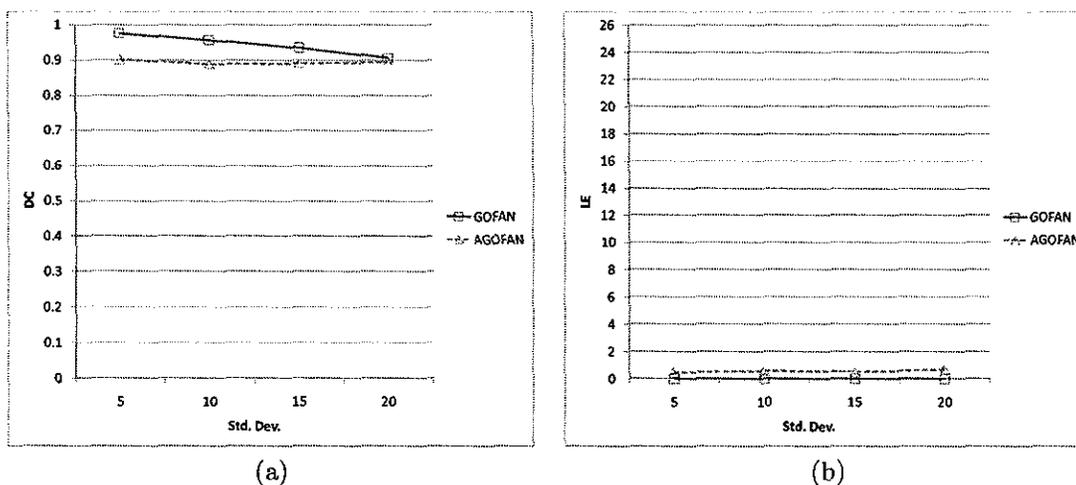


Figura 23: Desempeño de los filtros propuestos en términos de (a) DC y (b) LE cuando se varía la desviación estándar del ruido en las imágenes de entrenamiento.

son cero para ambos filtros y para todos los valores probados de la desviación estándar. Sin embargo, cuando el objeto aparece distorsionado el aumento del nivel de ruido en las imágenes de entrenamiento tiene un impacto significativo sobre el desempeño del  $GOF_{AN}$ . Esto se puede observar en la figura 24(a). Por el contrario, el filtro  $AGOF_{AN}$  no baja notoriamente su desempeño conforme aumenta el ruido. Esta tolerancia al ruido se debe al proceso de adaptación en el algoritmo de diseño. Los valores del criterio LE se muestran en la figura 24(b). Es claro que cuando la desviación estándar del ruido aumenta a más de 10, el  $GOF_{AN}$  falla al detectar y localizar al objeto, mientras que el filtro  $AGOF_{AN}$  sigue siendo capaz de localizar al objeto con precisión menor o igual a un pixel.

## VI.2. Aplicación a seguimiento de objetos (tracking)

El incremento en poder de cómputo de los dispositivos modernos ha hecho posible el seguimiento de objetos en tiempo real. Los sistemas de seguimiento se usan en apli-

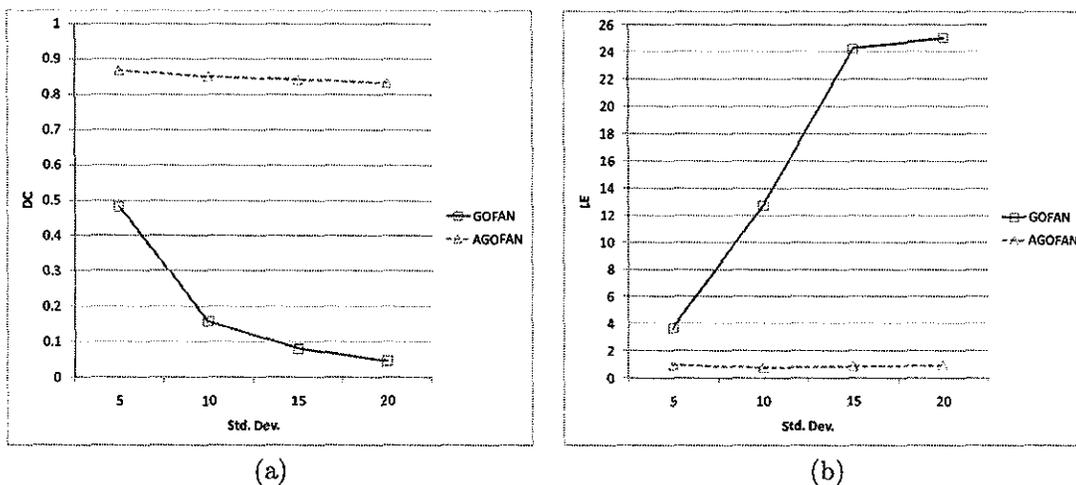


Figura 24: Desempeño de los filtros propuestos en términos de (a) DC y (b) LE cuando se varía la desviación estándar del ruido en las imágenes de entrenamiento. El objeto aparece en la escena escalado a 1.15 veces su tamaño original y rotado  $10.5^\circ$ .

caciones como vigilancia, navegación automática, entre otras (Yilmaz *et al.*, 2006). El seguir objetos en una secuencia de video requiere de procesar grandes cantidades de información. Para este propósito, se pueden aplicar dos enfoques: reducir la cantidad de información para procesar y realizar el procesamiento más rápido. Para reducir la cantidad de información se utiliza un enfoque en el que se extraen características de los patrones a seguir a partir de una escena observada (Duda *et al.*, 2000). Un extractor de características, idealmente obtiene un pequeño número de características. El encontrar la correspondencia de estas características entre una escena y la otra es lo que provee la información de desplazamiento. Más aún, cuando la cantidad de imágenes capturadas por segundo es alta, se puede realizar un preprocesamiento como eliminar el fondo de un cuadro al siguiente, de manera que la información resultante esté presente sólo en las regiones de la imagen donde tuvo lugar el movimiento. La principal desventaja de los sistemas basados en características es que pueden fallar en condiciones ruidosas, por ejemplo, cuando aparece una versión distorsionada del objeto en un fondo complejo y corrompido por ruido. Muchos algoritmos de extracción de características requieren de

segmentación de los objetos. Sin embargo, en las condiciones descritas es común que los algoritmos de segmentación fallen.

Por otro lado, el seguimiento de objetos se puede realizar sobre cada cuadro sin preprocesar ni extraer características utilizando métodos basados en apariencia (Yilmaz *et al.*, 2006; Davey *et al.*, 2008). Este enfoque requiere que o la apariencia del objeto no cambie mucho a lo largo de la secuencia de imágenes o que se conozcan *a priori* las posibles apariencias del objeto. Esto resulta en un alto costo computacional, por lo que se vuelve deseable el uso de esquemas de procesamiento rápido. Los correladores ópticos se han sido ampliamente utilizados para el reconocimiento de objetos (VanderLugt, 1964; Díaz-Ramírez *et al.*, 2006; Hebert y Henneberger, 1998). Realizan detección rápida al explotar el paralelismo inherente a los sistemas ópticos. En caso de que el seguimiento requiere conocer únicamente la ubicación del objeto en un momento dado, el problema se puede resolver mediante el enfoque de detecciones sucesivas (Bruno *et al.*, 2008). En este enfoque, múltiples filtros de correlación (uno por cada vista del objeto) se aplican en rápida sucesión generando una secuencia de salidas de correlación. El uso de correladores ópticos para detección rápida de objetos permite aplicaciones de seguimiento en tiempo real. El enfoque de seguimiento mediante detecciones sucesivas presenta ciertas ventajas. Por ejemplo, los filtros se diseñan optimizando una métrica de desempeño en forma analítica, como minimizar los errores de detección (Yaroslavsky, 1993), por lo que el desempeño de detección en cada cuadro es óptimo respecto al criterio optimizado. Los filtros de correlación también se pueden diseñar para minimizar los errores de localización (Kumar *et al.*, 1992; Kober y Campos, 1996). Además, como el objeto se está detectando en cada escena, no hay problema en caso de que el objeto desaparezca temporalmente de la escena (ya sea por que sale de la escena o porque se oculta por otro objeto) ya que será detectado correctamente al volver a aparecer. Los filtros

---

compuestos de correlación se propusieron para considerar múltiples vistas del objeto de interés en una sola operación de correlación (Casasent, 1984; Kerekes y Kumar, 2008; Kumar, 1992; Mahalanobis *et al.*, 1987, 1994). Entonces, se puede diseñar un filtro óptimo para cada vista del objeto y luego múltiples vistas se combinan en un sólo filtro compuesto (Ramos-Michel y Kober, 2008).

Si se dispone de información sobre el fondo donde se va a llevar a cabo la detección, la capacidad de discriminación de los filtros compuestos se puede mejorar utilizando un enfoque adaptativo (Díaz-Ramírez *et al.*, 2006; González-Fraga *et al.*, 2006). En este enfoque, el filtro compuesto se diseña para optimizar el desempeño para un conjunto de entrenamiento y un fondo típico, en lugar de diseñarse para tener un buen desempeño promedio sobre un ensamble de imágenes. Se sabe que la razón señal a ruido de los filtros compuestos se degrada gradualmente conforme crece el número de vistas utilizadas (Kerekes y Kumar, 2008; Kumar y Pochapsky, 1986). Este problema se resuelve utilizando un banco de filtros compuestos, donde cada filtro se diseña para un subconjunto de vistas conocidas del objeto (Bhuiyan *et al.*, 2007, 2006; Billert y Singher, 2002). El seguimiento se realiza mediante la correlación de la escena con cada filtro en el banco. Se calcula la DC en cada plano de salida y se escoge el plano con mayor valor de DC. La localización del objeto se toma como la localización del pico de correlación en el plano seleccionado. Cabe mencionar que el número de filtros en el banco se puede escoger para garantizar la precisión de la localización con la menor cantidad posible de correlaciones.

En esta sección se presenta un sistema opto digital de seguimiento utilizando un banco de filtros adaptativos. Los filtros se componen a partir de múltiples vistas del objeto utilizando un algoritmo iterativo. El algoritmo optimiza la capacidad de discriminación de cada filtro en el banco, y minimiza la cantidad de filtros necesarios garantizando una

---

precisión de localización prefijada. De esta manera, se optimiza la razón de desempeño contra complejidad computacional.

### VI.2.1. Formulación del problema

Se considera una secuencia discreta de imágenes  $\{s_i(x)\}$  donde  $s_i(x)$  denota la  $i$ -ésima imagen. Se denota por  $t(x; \theta)$  el objeto, donde el vector de parámetros  $\theta$  determina la apariencia (distorsión) del objeto, como rotación o escalamiento. Sean  $x_i$  y  $\theta_i$  la localización y distorsión del objeto en la  $i$ -ésima imagen. No se hace ninguna suposición sobre la relación entre  $x_i$  y  $x_{i+1}$ . La estimación de la posición del objeto se realiza usando un banco de filtros  $\{H_i(\omega), i = 1, 2, \dots, N_h\}$ , siendo  $N_h$  el número de filtros en el banco. Cada filtro  $H_i(\omega)$  se compone de las respuestas de frecuencia de filtros óptimos diseñados a partir de imágenes de entrenamiento. El diagrama del algoritmo se muestra en la figura 25.

Sea  $r_i(x)$  la  $i$ -ésima imagen de entrenamiento que muestra una vista del objeto y sea  $\bar{w}_i(x)$  la región inversa de soporte para esa vista. La imagen de entrada se representa con el modelo disjunto, es decir

$$s(x) = t_i(x - x_s) + b(x) \bar{w}_i(x - x_s) + n_s(x). \quad (174)$$

No se usa el sufijo para indicar el número de cuadro del que se está tratando, porque el filtro se diseña para detectar al objeto en un solo cuadro. De esta manera, el modelo de señales corresponde al modelo AN estudiado en el Capítulo V. Se modela la situación en la que las imágenes de referencia se capturan en un ambiente controlado pero con equipo de baja calidad y no se realiza preprocesamiento de las mismas. El parámetro  $\theta_i$  no se muestra porque un filtro se diseña para una vista del objeto de referencia y no hay la necesidad de distinguir entre diferentes vistas. Los procesos de ruido y variables

aleatorias se consideran estadísticamente independientes. El filtro óptimo para el modelo disjunto es el filtro GOF (Javidi y Wang, 1994a). Sin embargo, cuando la imagen de referencia contiene ruido, se usa una aproximación del filtro óptimo. En el capítulo V se muestran dos aproximaciones  $AN_1$  y  $AN_2$  y un algoritmo para seleccionar la mejor estimación dependiendo de las estadísticas del objeto y el fondo en la escena. Se denota por  $GOF_{AN}$  el filtro obtenido de esta manera.

### VI.2.2. Diseño de filtros compuestos para reconocimiento invariante a distorsiones geométricas

Se diseña un filtro óptimo, con respecto al criterio POE, para cada vista del objeto disponible para entrenamiento. Luego, se utiliza un algoritmo modificado de funciones discriminantes sintéticas para diseñar un filtro compuesto de respuestas de impulso de múltiples filtros óptimos. Sea  $h_i(x)$  la respuestas de impulso del filtro óptimo diseñado a partir de la  $i$ -ésima imagen de entrenamiento  $i \in \{1, 2, \dots, N_v\}$  donde  $N_v$  es el número de vistas de entrenamiento disponibles. Sean  $\{\alpha_1, \alpha_2, \dots, \alpha_{N_h}\}$   $N_h$  subconjuntos tomados del conjunto de vistas de entrenamiento. Cada conjunto se utiliza para sintetizar la respuesta de impulso de un filtro compuesto  $g_j(x)$ . Los filtros compuestos  $g_j(x)$  se diseñan como una combinación lineal de las respuestas de impulso del  $j$ -ésimo conjunto de entrenamiento más objetos a rechazar utilizando la ecuación (173).

Finalmente, se aplica el algoritmo de adaptación que se muestra en la sección VI.1. El fondo que se utiliza para el algoritmo de adaptación puede ser un fondo representativo de la aplicación en la que se va a utilizar el filtro o se puede describir en forma estocástica como realización de un proceso estocástico con estadísticas similares al fondo en el que se va a realizar el rastreo. En el fondo que se usa para adaptación puede contener objetos

falsos o estructuras similares al objeto.

En la figura 25 se muestra un diagrama del algoritmo propuesto para el diseño del banco de filtros. El algoritmo se detalla a continuación:

1. Para cada una de las  $N_v$  imágenes imagen de entrenamiento se diseña un filtro óptimo  $h_i(x)$ .
  2. Se busca el mínimo número de filtros requeridos en el banco utilizando el algoritmo de búsqueda binaria (Cormen *et al.*, 2009). El intervalo inicial de búsqueda es  $[1, N_v]$ .
    - a) Si hay solo un elemento en el intervalo actual de búsqueda, el procedimiento ha terminado. De lo contrario, continuar al paso siguiente.
  3. Sea  $k$  la mediana del intervalo de búsqueda actual.
  4. Se divide el conjunto de entrenamiento en  $k$  subconjuntos.
  5. Para cada subconjunto se sintetiza un filtro compuesto utilizando los filtros óptimos para cada vista del conjunto de la siguiente forma:
    - a) Se realiza la correlación entre el filtro actual y el fondo de entrenamiento y se calcula el valor de la DC utilizando la ecuación (64).
    - b) Si el valor de la DC es mayor o igual al valor deseado, el procedimiento de adaptación termino para este subconjunto. De lo contrario, continuar al paso siguiente.
    - c) Se crea un patrón a partir del fondo en las coordenadas de mayor intensidad en el plano de correlación. Luego, se agrega ese patrón al conjunto de patrones a rechazar.
-

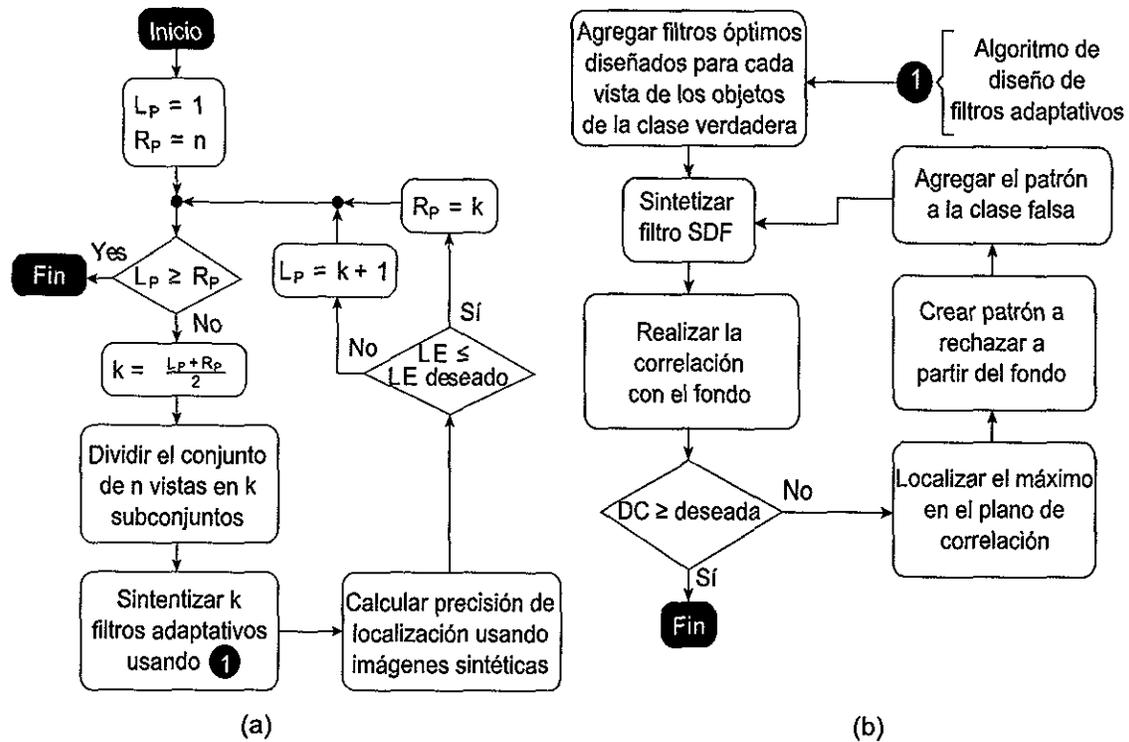


Figura 25: Diagrama del algoritmo de diseño del banco de filtros.

d) Se diseña un nuevo filtro SDF utilizando la ecuación (173) y se continúa al paso 7.

6. Se calcula la precisión de localización utilizando imágenes sintéticas.

7. Si la precisión de localización es mayor o igual a un valor prefijado, se continúa la búsqueda en la mitad izquierda del intervalo actual de búsqueda, de lo contrario se continúa la búsqueda en la mitad derecha del intervalo. Se continúa al paso 3.

Cabe mencionar que los patrones a rechazar no necesariamente coinciden para todos los filtros compuestos en el banco.

### VI.2.3. Sistema opto digital de rastreo

Los sistemas híbridos opto digitales se pueden utilizar para implementar procesadores basados en correlación (Díaz-Ramírez y Kober, 2007; Díaz-Ramírez *et al.*, 2006; VanderLugt, 1964; Weaver y Goodman, 1966). Se usan dos arquitecturas básicas de correladores ópticos: el correlador 4F (VanderLugt, 1964) y el correlador de transformada conjunta (JTC<sup>1</sup>) (Weaver y Goodman, 1966). La principal ventaja del JTC sobre el 4F es que el primero es menos sensible a problemas de alineación del sistema óptico. Un diagrama del sistema óptico propuesto se muestra en la figura 26. Como correlador óptico se puede utilizar un JTC compacto (Javidi y Horner, 1989). El banco de filtros considera las múltiples apariencias posibles del objeto. El *i*-ésimo cuadro en la secuencia de video se procesa con todos los filtros en el banco en sucesión rápida. La salida del sistema en un momento dado es el plano de correlación con mayor valor de DC. Para obtener este plano se realiza el siguiente procedimiento: en cada plano se localiza el pico de mayor intensidad; luego, utilizando la región de soporte del objeto se localiza el lóbulo lateral de mayor intensidad sobre el área del fondo. Estas intensidades se usan para calcular el valor de la DC para cada plano. Finalmente, el plano con máxima DC se selecciona como la salida del sistema. Los valores de la DC deberían ser cercanos al valor con el que se diseñaron los filtros compuestos. Si el valor de la DC en el plano seleccionado está por debajo de un umbral que se calcula en función de la DC deseada en el proceso de adaptación, el sistema determina que el objeto no está presente en la escena. La trayectoria del objeto como función del tiempo se forma a partir de las posiciones de los picos en la salida del sistema. Para algunas aplicaciones, como la vigilancia aérea, el banco de filtros se puede adaptar para los cambios en la apariencia del fondo debidos a los cambios de terreno. El algoritmo adaptativo se puede utilizar

---

<sup>1</sup>Joint Transform Correlator

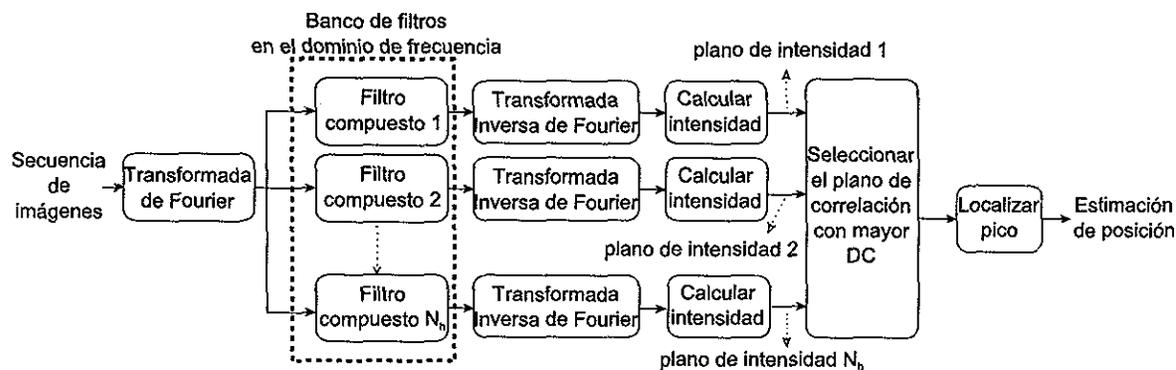


Figure 26: Diagrama del sistema opto digital de seguimiento.

con diferentes fondos, representativos de cada tipo de terreno. Los filtros compuestos resultantes se incluyen en el banco de filtros.

Una implementación más sencilla se puede realizar con un solo filtro compuesto. Mientras que esto presenta una ventaja en términos de tiempo de procesamiento, la precisión de localización se degrada si se usa una gran cantidad de vistas en la composición de un sólo filtro. Por el contrario, un banco con más filtros logra mejor precisión de localización a costa de mayor tiempo de procesamiento.

#### VI.2.4. Resultados experimentales

En esta sección se analiza el desempeño del sistema propuesto para el rastreo de objetos en un fondo no traslapado. El criterio de desempeño es la precisión de localización. Las imágenes que se utilizan son de  $256 \times 256$  píxeles. Se muestra un ejemplo de imagen de referencia en la figura 27(a); el objeto de interés es la mariposa en el centro de la imagen de referencia. El objeto es de tamaño  $28 \times 44$  píxeles, tiene intensidad media de 100 y desviación estándar 25. Un ejemplo de escena de entrada se muestra en la figura 27(b). Las secuencias de imágenes se generan utilizando el siguiente algoritmo:

1. Se generan parámetros iniciales para el objeto: posición  $x_s$ , rotación  $\tau$  y escala  $\sigma$ .
2. Se genera un vector de dirección en forma polar  $\nu = (\rho, \phi)$ .
3. Se genera una secuencia aleatoria  $\Delta_i\tau$  con un factor de correlación 0.95 que se utiliza para perturbar  $\tau$ .
4. Se generan secuencias aleatorias de media cero  $\Delta_i\rho$ ,  $\Delta_i\phi$  y  $\Delta_i\sigma$  que se usan como perturbaciones para  $\rho$ ,  $\phi$  y  $\sigma$ , respectivamente.
5. Para cada cuadro en la secuencia:
  - a) El objeto se rota  $\tau$  grados, se escala por un factor de  $\sigma$  y se ubica en la posición  $x_s$  sobre el fondo.
  - b) Se agrega ruido blanco a la escena.
  - c) Se actualiza la posición del objeto  $x_s \leftarrow x_s + \nu$ .
  - d) Se actualizan los parámetros:  $\tau \leftarrow \tau + \Delta_i\tau$ ,  $\rho \leftarrow \rho + \Delta_i\rho$ ,  $\phi \leftarrow \phi + \Delta_i\phi$  y  $\sigma \leftarrow \sigma + \Delta_i\sigma$ .

La escala se mantiene en el intervalo  $[0.8, 1.2]$ , mientras que el resto de los parámetros se permite que varíen libremente.

Las imágenes que se muestran en la figura 28 contienen versiones degradadas del objeto con factores de escalamiento de 0.9, 1.0 y 1.1, y rotadas por -9, 3, 3, y 9 grados. El fondo que se usa para el algoritmo adaptativo se muestra en la figura 29. Cabe resaltar que el fondo que se utiliza para entrenar los filtros no es el mismo que se utiliza en la generar la secuencia de imágenes. Un cuadro ejemplo y su salida de correlación correspondiente se muestran en las figuras 30(a) y 30(b), respectivamente.

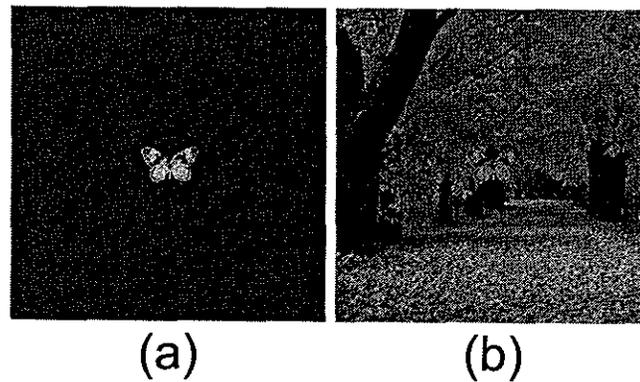


Figura 27: Imágenes usadas en los experimentos: (a) Imagen de referencia ejemplo, (b) escena ejemplo.

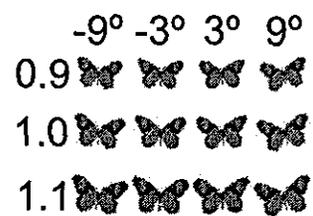


Figure 28: Imágenes del objeto de interés distorsionadas por rotación y escalamiento.

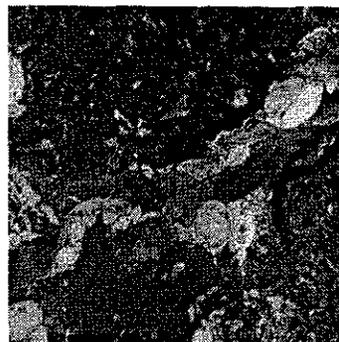


Figura 29: Fondo usado como fondo ejemplo en el algoritmo adaptativo de diseño de filtros.

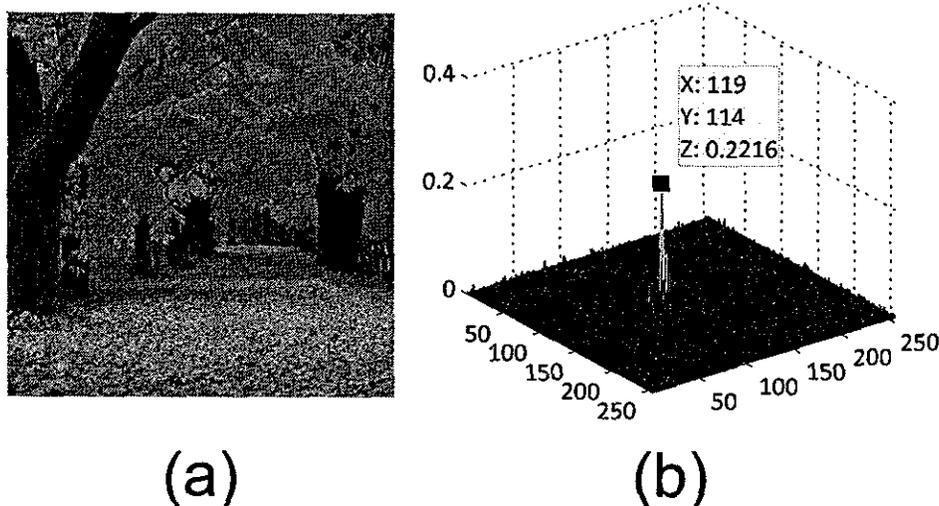


Figura 30: (a) Cuadro ejemplo y (b) su salida de correlación correspondiente.

Para los experimentos, se generaron 100 secuencias de imágenes con 100 cuadros cada una. Los filtros compuestos se diseñaron utilizando imágenes de referencia contaminadas con ruido aditivo con desviación estándar 10. Se prueban cuatro sistemas. El primer sistema, denotado como sistema A utiliza un sólo filtro compuesto con todas las vistas de entrenamiento. Los siguientes sistemas utilizan bancos de 4, 6 y 9 filtros, y se denotan por B4, B6 y B9, respectivamente. El desempeño de los filtros se muestra en la figura 31. La gráfica muestra los errores de localización, por cada cuadro, promediados sobre todas las secuencias generadas. Se puede observar que el sistema A es capaz de seguir la posición del objeto con un error promedio de 4.5 píxeles. Se puede apreciar también cómo los errores de localización disminuyen conforme se aumenta el número de filtros en el banco. Para las imágenes utilizadas, nueve filtros compuestos son suficientes para garantizar precisión de localización de un píxel o menos.

La figura 32 muestra una trayectoria ejemplo del objeto en una secuencia de 25 imágenes. El contorno del objeto se muestra para ilustrar la trayectoria a lo largo de la secuencia. El objeto inicia en el centro de la escena, avanza lentamente hasta que

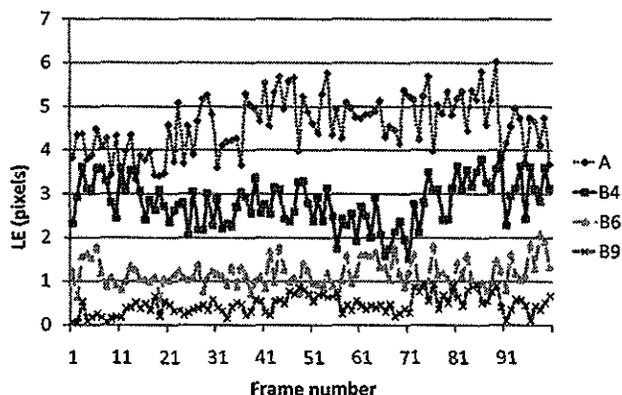


Figura 31: Errores de localización en cada cuadro en la secuencia de imágenes para el sistema que utiliza uno, cuatro, seis y nueve filtros compuestos.

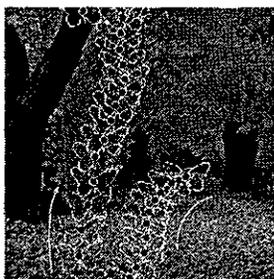


Figura 32: Secuencia ejemplo con 25 cuadros.

sale por la parte inferior de la escena, y posteriormente vuelve a entrar por la parte superior. Esta trayectoria ilustra dos casos típicos en los que las técnicas comunes de rastreo presentan dificultades. El primer caso es cuando el objeto desaparece de la escena y debe ser detectado nuevamente al volver a aparecer. El segundo caso es cuando la velocidad del objeto es mayor que la velocidad de captura de la cámara. En la secuencia se observa que hay ocasiones en las que no hay traslape entre una posición del objeto y la siguiente. Esto no satisface la suposición comúnmente utilizada de que el desplazamiento entre cuadros sucesivos es pequeño. Esto no presenta problema alguno para las técnicas de detección sucesiva ya que no se hacen ninguna suposición respecto al desplazamiento máximo entre cuadros consecutivos.

## VI.3. Resumen

En este capítulo se presenta el diseño de filtros compuestos adaptativos para la detección y localización de objetos en escenas no traslapadas. Los filtros se componen utilizando un algoritmo modificado de funciones discriminantes sintéticas que produce un filtro como combinación lineal de respuestas de impulso de filtros óptimos y de filtros de correspondencia. Los filtros óptimos se diseñan a partir de imágenes ruidosas de entrenamiento. La apariencia y forma del objeto se estima a partir de estas imágenes ruidosas. Utilizando un algoritmo iterativo, el filtro compuesto se adapta a un fondo típico representativo de las situaciones en las que se va a utilizar. Se muestra mediante simulaciones por computadora que los filtros obtenidos mediante el algoritmo propuesto logran buenos resultados en términos de capacidad de discriminación y errores de localización aún utilizando imágenes ruidosas de referencia. Se concluye que los cambios de escala tienen un mayor impacto en el desempeño de los filtros propuestos que los cambios de rotación.

Se presenta también la aplicación de los filtros adaptativos propuestos al problema de rastrear objetos en secuencias de video. Se propone la arquitectura de un sistema de rastreo de objetos en tiempo real que utiliza un banco de filtros y un correlación óptica. El banco de filtros se diseña a partir de imágenes ruidosas de entrenamiento optimizando dos criterios: se minimizan los errores de localización utilizando la menor cantidad de filtros compuestos en el banco. Las imágenes se utilizan tal como son capturadas sin necesidad de preprocesamiento manual. Se muestra mediante simulaciones por computadora que el algoritmo propuesto permite estimar la trayectoria de un objeto a lo largo de una secuencia de imágenes con alta precisión. Se mostraron cuatro variantes del sistema propuesto en las que se intercambia tiempo de procesamiento y

---

precisión de localización.

# Capítulo VII

---

## Conclusiones

---

El diseño de filtros de correlación comúnmente supone que se conoce la apariencia ideal del objeto de interés así como su forma. Sin embargo, en diversas aplicaciones prácticas, dicha suposición no se cumple. Las imágenes inevitablemente se encuentran corrompidas por ruido debido al proceso de formación de imagen, debido a imperfecciones de los sensores, entre otras causas. Aunado a esto, los objetos en el mundo real aparecen incrustados sobre fondos complejos. Para obtener la apariencia del objeto de interés, es indispensable realizar algún tipo de preprocesamiento manual por parte de un operador humano para identificar exactamente la apariencia y forma del objeto. La motivación de este trabajo fue por encontrar métodos que permitan el reconocimiento de objetos en forma automatizada, sin necesidad de intervención humana. El problema tiene diversas áreas de aplicación tales como la vigilancia y búsquedas utilizando imágenes de ejemplo. La hipótesis fundamental del presente trabajo es que el incorporar la presencia de ruido al modelo matemático utilizado para el reconocimiento obtiene mejores resultados que cuando se tratan los problemas de restauración, segmentación y reconocimiento por separado.

La aportación de este trabajo se centra en el desarrollo de filtros óptimos para el reconocimiento de objetos cuando únicamente se dispone de imágenes de entrenamiento ruidosas. Se propusieron modelos de escena nuevos en los que se considera la presencia de ruido en la imagen de referencia. Se propusieron estimaciones para obtener la información necesaria para el reconocimiento a partir de las imágenes de referencia

---

disponibles. Además, se estudió analíticamente el desempeño de las estimaciones propuestas y se propusieron técnicas para seleccionar la estimación adecuada dependiendo de los parámetros estadísticos de las imágenes en las que el reconocimiento se lleva a cabo. Para esto, se propuso la métrica de errores de correspondencia que funciona como una buena estimación del desempeño de los filtros en términos de las métricas de detección. Se mostró que esta métrica refleja mejor el desempeño de los filtros que la métrica de error cuadrático medio que se usa comúnmente para problemas de restauración.

Dada la sensibilidad de los filtros de correlación a cambios en la apariencia del objeto debidos a distorsiones geométricas, se propuso también un algoritmo para el diseño de filtros adaptativos compuestos que, además de considerar el ruido en la imagen de referencia, son tolerantes a distorsiones como rotaciones y cambios de escala. Se propuso un algoritmo que permite garantizar un desempeño en cuanto precisión de localización utilizando un banco de filtros con el mínimo número posible de filtros.

Los filtros desarrollados junto con los algoritmos de estimación propuestos se utilizaron en una aplicación para rastrear objetos en secuencias de video. Se propuso una arquitectura híbrida opto digital para realizar el rastreo de objetos en tiempo real.

Finalmente, se presentaron y discutieron simulaciones por computadora que demuestran la efectividad de los filtros propuestos junto con las estimaciones correspondientes en gran parte del espacio de parámetros de los modelos de escena estudiados. Se mostró que es posible detectar y localizar objetos con precisión aceptable aún en la presencia de niveles moderados de ruido en las imágenes de referencia e incluso cuando la escena de referencia contiene un fondo complejo.

---

## VII.1. Trabajo futuro

Aunque se cumplieron los objetivos planteados para este trabajo de investigación, quedan aún líneas de investigación que se pueden continuar. Una primer dirección que se puede seguir es extender las técnicas propuestas para el reconocimiento de objetos en imágenes a color o multiespectrales (como imágenes aéreas o satelitales). Otra área de oportunidad es la aplicación del modelo para utilizar imágenes de entrenamiento con fondos complejos (modelo NN) para rastrear objetos en secuencias de video. Siguiendo la línea de rastrear objetos en secuencias de imágenes, se puede investigar el rastrear objetos cuando el objeto, el fondo y el dispositivo de captura pueden presentar movimientos.

Otra línea de investigación que se puede seguir es el utilizar el modelo de escena que considera imágenes ruidosas de referencia para derivar filtros no lineales de correlación.

## VII.2. Publicaciones derivadas de este trabajo

Publicaciones en revistas indexadas:

- Aguilar-González, P.M. y Kober, V. 2011. Design of correlation filters for pattern recognition with disjoint reference image. *Optical Engineering*, **50**(11), 117201 p.
  - Aguilar-González, P.M. y Kober, V. 2011. Design of correlation filters for pattern recognition using a noisy reference. Enviado a *Optics Communications*. En prensa.
  - Aguilar-González, P.M., Kober, V. y Díaz-Ramírez, V.H. 2011. Adaptive composite correlation filters for object tracking using noisy training images. Enviado a *Optical Engineering*. En revisión.
-

## Publicaciones en revistas arbitradas:

- Aguilar-González, P.M., Kober, V. y Ovseyevich, I. 2010. Pattern recognition in nonoverlapping background with noisy target image. *Pattern Recognition and Image Analysis*, **20**(2), 163–168 p.

## Publicaciones en congresos:

- Aguilar-González, P.M. y Kober, V. 2008. Correlation Filters for Pattern Recognition Using a Noisy Reference. En J. Ruiz-Shulcloper y W.G. Kropatsch, eds. *Progress in Pattern Recognition, Image Analysis and Applications*. CIARP 2008. La Habana, Cuba. LNCS 5197. 38–45 p.
  - Aguilar-González, P.M. y Kober, V. 2008. Pattern recognition of an implicitly given target. En A.G. Tescher, ed. *Proceedings of SPIE, Applications of Digital Image Processing XXXI*. SPIE. 707303 p.
  - Aguilar-González, P.M. y Kober, V. 2009. Correlation filters for object detection in nonoverlapping background noise using a noisy reference image. En A.G. Tescher, ed. *Proceedings of SPIE, Applications of Digital Image Processing XXXII*. SPIE. 74431G p.
  - Aguilar-González, P.M. y Kober, V. 2009. Correlation Pattern Recognition in Nonoverlapping Scene Using a Noisy Reference. In E. Bayro-Corrochano y J.O. Eklundh, eds. *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. CIARP 2009. Guadalajara, México. LNCS 5856. 555–562 p.
  - Aguilar-Gonzalez, P.M., Kober, V., y Ovseyevich, I. 2010. Object detection in a cluttered background with a noisy image. 9th International Conference Computer Data Analysis and Modeling. Minsk, Bielorrusia.
-

- Aguilar-González, P.M. y Kober, V. 2010. Correlation filter design using a single cluttered training image for detecting a noisy target in a nonoverlapping scene. En A.G. Tescher, ed. *Proceedings of SPIE, Applications of Digital Image Processing XXXIII*. SPIE. 779806 p.
  - Aguilar-Gonzalez, P.M., Kober, V., y Karnaukhov V. 2011. Pattern Recognition with Noisy Target. 11th International Conference on Pattern Recognition and Information Processing. Minsk, Bielorrusia.
  - Aguilar-González, P.M. y Kober, V. 2011. Design of Correlation Filters for Pattern Recognition Using a Noisy Training Image. En J.F. Martínez-Trinidad et al., eds. *Pattern Recognition*. MCPRI 2011. Cancún, México. LNCS 6718. 194–201 p.
  - Aguilar-González, P.M. y Kober, V. 2011. Pattern recognition with composite correlation filters designed from noisy training images. En A.G. Tescher, ed. *Proceedings of SPIE, Applications of Digital Image Processing XXXIV*. SPIE. 81350B–81350B-8 p.
-

## Bibliografía

- Avi-Itzhak, H. y Diep, T. (1996). Arbitrarily tight upper and lower bounds on the bayesian probability of error. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **18**(1): 89–91.
- Bhuiyan, S. M. A., Alam, M. S., y Sims, S. R. F. (2006). Target detection, classification, and tracking using a maximum average correlation height and polynomial distance classification correlation filter combination. *Optical Engineering*, **45**(11): 116401.1–116401.13.
- Bhuiyan, S. M. A., Alam, M. S., y Alkanhal, M. (2007). New two-stage correlation-based approach for target detection and tracking in forward-looking infrared imagery using filters based on extended maximum average correlation height and polynomial distance classifier correlation. *Optical Engineering*, **46**(8): 086401.1–086401.14.
- Billert, O. y Singher, L. (2002). Adaptive multiple filtering. *Optical Engineering*, **41**(1): 55–68.
- Bovik, A. C. (2005). *Handbook of Image and Video Processing*. Academic Press, New York, USA, segunda edición. ISBN 0121197921. 1384 p.
- Bruno, M. G. S., Araújo, R. V., y Pavlov, A. G. (2008). Sequential monte carlo methods for joint detection and tracking of multiaspect targets in infrared radar images. *EURASIP Journal on Advances in Signal Processing*, **2008**: 1–14.
- Casasent, D. (1984). Unified synthetic discriminant function computational formulation. *Applied Optics*, **23**(10): 1620–1627.
- Cormen, T. H., Leiserson, C. E., Rivest, R. L., y Stein, C. (2009). *Introduction to Algorithms*. The MIT Press, Cambridge, USA, tercera edición. ISBN 0262033844. 1312 p.
- Cortes, C. y Vapnik, V. (1995). Support-vector networks. *Machine Learning*, **20**: 273–297.
-

- Davey, S. J., Rutten, M. G., y Cheung, B. (2008). A comparison of detection performance for several Track-before-Detect algorithms. *EURASIP Journal on Advances in Signal Processing*, **2008**: 1–11.
- Díaz-Ramírez, V. H. y Kober, V. (2007). Adaptive phase-input joint transform correlator. *Applied Optics*, **46**(26): 6543–6551.
- Díaz-Ramírez, V. H., Álvarez-Borrego, J., y Kober, V. (2006). Pattern recognition with an adaptive joint transform correlator. *Applied Optics*, **45**(23): 5929–5941.
- Duda, R. O., Hart, P. E., y Stork, D. G. (2000). *Pattern Classification*. Wiley-Interscience, New York, USA, segunda edición. ISBN 0471056693. 654 p.
- Gonzalez, R. C. y Woods, R. E. (1992). *Digital Image Processing*. Addison-Wesley Pub (Sd), Boston, USA, tercera edición. ISBN 0201508036. 528 p.
- González-Fraga, J., Kober, V., y Álvarez-Borrego, J. (2006). Adaptive synthetic discriminant function filters for pattern recognition. *Optical Engineering*, **45**: 057005.1–057005.10.
- Goudail, F. y Réfrégier, P. (1996). Optimal detection of a target with random gray levels on a spatially disjoint background noise. *Optics Letters*, **21**(7): 495–497.
- Guérault, F. y Réfrégier, P. (1998). Unified statistically independent region processor for deterministic and fluctuating target in non-overlapping background. *Optics Letters*, **23**: 412–414.
- Guérault, F., Signac, L., Goudail, F., y Réfrégier, P. (1997). Location of target with random gray levels in correlated background with optimal processors and preprocessings. *Optical Engineering*, **36**(10): 2660–2670.
- Hebert, T. J. y Henneberger, B. E. (1998). Optical filtering approach to regularized tracking of an object's position and orientation. *Applied Optics*, **37**(32): 7577–7586.
- Horner, J. y Gianino, P. (1984). Phase-only matched filtering. *Applied Optics*, **23**(6): 812–816.
- Jain, A. (1989). *Fundamentals of digital image processing*. Prentice Hall, New York, USA. ISBN 0133361659. 592 p.
-

- Javidi, B. (1994). *Real-Time Optical Information Processing*. Academic Press, New York, USA. ISBN 0123811805. 536 p.
- Javidi, B. y Horner, J. L. (1989). Single spatial light modulator joint transform correlator. *Applied Optics*, **28**(5): 1027–1032.
- Javidi, B. y Wang, J. (1992). Limitation of the classic definition of the correlation signal-to-noise ratio in optical pattern recognition with disjoint signal and scene noise. *Applied Optics*, **31**(32): 6826–6829.
- Javidi, B. y Wang, J. (1994a). Design of filters to detect a noisy target in nonoverlapping background noise. *Journal of the Optical Society of America A*, **11**(10): 2604–2612.
- Javidi, B. y Wang, J. (1994b). Optimum filter for detection of a target in nonoverlapping scene noise. *Applied Optics*, **33**(20): 4454–4458.
- Javidi, B., Zhang, G., y Parchekani, F. (1996). Minimum-mean-square-error filters for detecting a noisy target in background noise. *Applied Optics*, **35**: 6964–6975.
- Kerekes, R. A. y Kumar, B. V. K. V. (2008). Selecting a composite correlation filter design: a survey and comparative study. *Optical Engineering*, **47**(6): 067202.1–067202.18.
- Kober, V. y Campos, J. (1996). Accuracy of location measurement of a noisy target in a nonoverlapping background. *Journal of the Optical Society of America A*, **13**(8): 1653–1666.
- Kumar, B. V. K. V. (1992). Tutorial survey of composite filter designs for optical correlators. *Applied Optics*, **31**(23): 4773–4801.
- Kumar, B. V. K. V. y Hassebrook, L. (1990). Performance measures for correlation filters. *Applied Optics*, **29**(20): 2997–3006.
- Kumar, B. V. K. V. y Pochapsky, E. (1986). Signal-to-noise ratio considerations in modified matched spatial filters. *Journal of the Optical Society of America A*, **3**(6): 777–786.
- Kumar, B. V. K. V., Dickey, F. M., y DeLaurentis, J. M. (1992). Correlation filters minimizing peak location errors. *Journal of the Optical Society of America A*, **9**(5): 678–682.
-

- Kumar, B. V. K. V., Mahalanobis, A., y Juday, R. (2005). *Correlation pattern recognition*. Cambridge Univ Press, New York, USA. ISBN 0521571030. 402 p.
- Mahalanobis, A., Kumar, B. V. K. V., y Casasent, D. (1987). Minimum average correlation energy filters. *Applied Optics*, **26**(17): 3633–3640.
- Mahalanobis, A., Kumar, B. V. K. V., Song, S., Sims, S. R. F., y Epperson, J. F. (1994). Unconstrained correlation filters. *Applied Optics*, **33**(17): 3751–3759.
- Mardia, K. y Hainsworth, T. (1988). A spatial thresholding method for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, **10**(6): 919–927.
- McDonough, R. N. y Whalen, A. D. (1995). *Detection of Signals in Noise*. Academic Press, New York, USA, segunda edición. ISBN 0127448527. 495 p.
- Milton, A., Barone, F., y Kruer, M. (1985). Influence of nonuniformity on infrared focal plane array performance. *Optical Engineering*, **24**(5): 855–862.
- Papageorgiou, C. y Poggio, T. (2000). A trainable system for object detection. *International Journal of Computer Vision*, **38**(1): 15–33.
- Papoulis, A. (1991). *Probability, Random Variables and Stochastic Processes*. Mcgraw-Hill College, New York, USA, tercera edición. ISBN 0070484775. 666 p.
- Pratt, W. K. (2007). *Digital Image Processing*. Wiley-Interscience, New York, USA, cuarta edición. ISBN 9780471767770. 812 p.
- Ramos-Michel, E. M. y Kober, V. (2008). Adaptive composite filters for pattern recognition in linearly degraded and noisy scenes. *Optical Engineering*, **47**(4): 047204.1–047204.7.
- Réfrégier, P. y Goudail, F. (2003). *Statistical Image Processing Techniques for Noisy Images: An Application-Oriented Approach*. Kluwer Academic Publishers, New York, USA. ISBN 030647865X. 266 p.
- Réfrégier, P., Javidi, B., y Zhang, G. (1993). Minimum mean square error filter for pattern recognition with spatially disjoint target and scene noise. *Optics Letters*, **18**(17): 1453–1455.
-

- Suetens, P., Fua, P., y Hanson, A. J. (1992). Computational strategies for object recognition. *ACM Computing Surveys*, **24**: 5–62.
- Towghi, N. y Javidi, B. (2001a). Image recognition in the presence of non-Gaussian noise with unknown statistics. *Journal of the Optical Society of America A*, **18**(11): 2744–2753.
- Towghi, N. y Javidi, B. (2001b). Optimum receivers for pattern recognition in the presence of gaussian noise with unknown statistics. *Journal of the Optical Society of America A*, **18**(8): 1844–1852.
- VanderLugt, A. (1964). Signal detection by complex spatial filtering. *IEEE Transactions on Information Theory*, **10**(2): 139–145.
- Weaver, C. S. y Goodman, J. W. (1966). A technique for optically convolving two functions. *Applied Optics*, **5**(7): 1248–1249.
- Webb, A. R. (2002). *Statistical Pattern Recognition*. Wiley, West Sussex, UK, segunda edición. ISBN 0470845147. 534 p.
- Yaroslavsky, L. (2003). *Digital Holography and Digital Image Processing: Principles, Methods, Algorithms*. Kluwer Academic Publishers, Boston, USA, primera edición. ISBN 9781402076343. 600 p.
- Yaroslavsky, L. P. (1993). The theory of optimal methods for localization of objects in pictures. En: E. Wolf, editor, *Progress in Optics*, páginas 145–201, Amsterdam. Elsevier.
- Yilmaz, A., Javed, O., y Shah, M. (2006). Object tracking: A survey. *ACM Computing Surveys*, **38**(4).
-

## Apéndice A

En este apéndice se presenta un método para generar ruido correlacionado. El modelo de función de correlación que se utiliza para los experimentos presentados en esta tesis es el del modelo exponencial (Jain, 1989). Para señales de una dimensión, la función de autocorrelación está dada por:

$$R(\tau) = \sigma^2 \rho^\tau, \quad (175)$$

donde  $\sigma^2$  es la varianza del campo aleatorio y  $\rho$  es el coeficiente de correlación. Si la media de las variables del campo es cero, la función de autocorrelación es la misma que la autocovarianza. Fijando el valor de  $\sigma = 1$  y utilizando el modelo de correlación exponencial, se cumplen las siguientes condiciones:

$$\begin{aligned} E\{X(i)X(i)\} &= 1, \\ E\{X(i-1)X(i)\} &= \rho, \\ E\{X(i-1)\xi\} &= 0, \\ E\{X(i)\xi\} &= 1, \end{aligned} \quad (176)$$

donde  $\xi$  es una variable aleatoria con distribución Gaussiana, media cero y varianza uno. La dependencia de la variable  $X(i)$  sobre su vecino y la variable aleatoria  $\xi$  se puede escribir como

$$X(i) = AX(i-1) + B\xi. \quad (177)$$

Multiplicando la ecuación (177) por  $X(i)$  y por  $X(i-1)$  y tomando el valor esperado, se obtiene el sistema de ecuaciones

$$\begin{aligned} E\{X(i)X(i)\} &= A\rho + B = 1, \\ E\{X(i)X(i-1)\} &= A = \rho, \end{aligned} \quad (178)$$

La solución del sistema para  $A$  y  $B$  es entonces

$$\begin{aligned} A &= \rho, \\ B &= 1 - \rho^2. \end{aligned} \quad (179)$$

Para generar campos aleatorios correlacionados de dos dimensiones, el modelo de función de autocorrelación es

$$R(\tau_h, \tau_v) = \sigma^2 \rho_h^{\tau_h} \rho_v^{\tau_v}, \quad (180)$$

donde  $\rho_v$  y  $\rho_h$  son los coeficientes de correlación vertical y horizontal, respectivamente, y  $\tau_v$  y  $\tau_h$  son los desplazamientos en las direcciones vertical y horizontal, respectivamente.

La relación entre una variable aleatoria y sus vecinas se describe por las condiciones

$$\begin{aligned}
 E\{X(i, j) X(i, j)\} &= 1, \\
 E\{X(i, j) X(i-1, j)\} &= \rho_v, \\
 E\{X(i, j) X(i, j-1)\} &= \rho_h, \\
 E\{X(i, j) X(i-1, j-1)\} &= \rho_v \rho_h, \\
 E\{X(i, j) \xi\} &= 1, \\
 E\{X(i-1, j) \xi\} &= 0, \\
 E\{X(i, j-1) \xi\} &= 0, \\
 E\{X(i-1, j-1) \xi\} &= 0.
 \end{aligned} \tag{181}$$

Planteando la dependencia de la variable  $X(i, j)$  sobre sus vecinos y sobre  $\xi$  como

$$X(i, j) = AX(i-1, j) + BX(i, j-1) + CX(i-1, j-1) + D\xi, \tag{182}$$

se plantea un sistema de ecuaciones cuya solución para las variables  $A$ ,  $B$ ,  $C$ , y  $D$  está dada por

$$\begin{aligned}
 A &= \rho_v, \\
 B &= \rho_h, \\
 C &= -\rho_v \rho_h, \\
 D &= \sqrt{(1 - \rho_v^2)(1 - \rho_h^2)}.
 \end{aligned} \tag{183}$$

## Apéndice B

En éste apéndice se deriva el umbral óptimo para separar dos distribuciones Gaussianas minimizando el número de errores de clasificación. Sean  $P(t)$  y  $P(b)$  las probabilidades *a priori* de que un pixel pertenezca a la región del objeto o a la región del fondo, respectivamente. Sean  $p(\nu|t)$  y  $p(\nu|b)$  las funciones de densidad de probabilidad condicional de que un pixel tenga una intensidad  $\nu$  dado que pertenece a la región del objeto o a la región del fondo, respectivamente. Finalmente, sean  $P(t|\nu)$  y  $P(b|\nu)$  las probabilidades *a posteriori* de que un pixel pertenezca a la al objeto o al fondo, respectivamente, dado que tiene una intensidad  $\nu$ .

La regla de decisión de mínimos errores de clasificación es la siguiente (Duda *et al.*, 2000): Decidir que un pixel pertenece al fondo si  $P(b|\nu) > P(t|\nu)$ . Usando el teorema de Bayes, estas probabilidades se pueden calcular como sigue:

$$P(t|\nu) = \frac{p(\nu|t) P(t)}{p(\nu)}, \quad (184)$$

$$P(b|\nu) = \frac{p(\nu|b) P(b)}{p(\nu)}. \quad (185)$$

Suponiendo que el objeto puede aparecer en cualquier lugar en la imagen de referencia con probabilidad uniforme, la probabilidad *a priori* de que un pixel pertenezca al objeto se calcula como la razón del tamaño del objeto al tamaño de la imagen ( $A_t$ ). Como las regiones del objeto y del fondo son complementarias, la probabilidad *a priori* de que un pixel pertenezca al fondo es  $1 - A_t$ . Las funciones de densidad de probabilidad condicionales dependen de las estadísticas de las regiones del objeto y fondo después de

filtrar, es decir

$$p(\nu|t) = \frac{1}{\sqrt{2\pi(\tilde{\sigma}_t^2 + \tilde{\zeta}_r^2)}} \exp\left(-\frac{(\nu - \mu_t)^2}{2(\tilde{\sigma}_t^2 + \tilde{\zeta}_r^2)}\right), \quad (186)$$

$$p(\nu|b) = \frac{1}{\sqrt{2\pi\tilde{\zeta}_r}} \exp\left(-\frac{\nu^2}{2\tilde{\zeta}_r^2}\right). \quad (187)$$

Sustituyendo las probabilidades *a priori* y las densidades de probabilidad condicionales en las ecuaciones (184) y (185), la regla de decisión se puede reescribir como

$$\frac{1}{\sqrt{2\pi\tilde{\zeta}_r}} \exp\left(-\frac{\nu^2}{2\tilde{\zeta}_r^2}\right) > \frac{1}{\sqrt{2\pi(\tilde{\sigma}_t^2 + \tilde{\zeta}_r^2)}} \exp\left(-\frac{(\nu - \mu_t)^2}{2(\tilde{\sigma}_t^2 + \tilde{\zeta}_r^2)}\right) \times \frac{A_t}{1 - A_t}. \quad (188)$$

Finalmente, aplicando la operación logaritmo a la desigualdad anterior y despejando términos, se llega al valor del umbral óptimo:

$$\tau = -\frac{\mu_t \tilde{\zeta}_r^2}{\tilde{\sigma}_t^2} + \frac{\hat{\sigma}_t \tilde{\zeta}_r}{\tilde{\sigma}_t^2} \sqrt{\mu_t^2 + 2\tilde{\sigma}_t^2 \log\left[\frac{(1 - A_t) \hat{\sigma}_t}{A_t \tilde{\zeta}_r}\right]}, \quad (189)$$

donde  $\hat{\sigma}_t^2 = \tilde{\sigma}_t^2 + \tilde{\zeta}_r^2$ .

## Apéndice C

En este apéndice se analiza el desempeño de las estimaciones para el modelo AN en términos de la métrica MSE. Para obtener una estimación del objeto compuesto es necesario realizar dos estimaciones: una estimación lineal para restaurar la apariencia del objeto y una estimación no lineal para determinar la forma del mismo. El orden en que estas dos estimaciones se realizan lleva a valores significativamente distintos de MSE. Sea  $G_i(\omega)$  la respuesta de frecuencia del filtro Wiener para cada estimación con  $i = 1, 2$ .  $G_i(\omega)$  se calcula como

$$G_i(\omega) = \frac{|T_i(\omega)|^2}{|T_i(\omega)|^2 + N_r(\omega)}, \quad (190)$$

donde  $T_1(\omega) = T(\omega)$  y  $T_2(\omega) = T_s(\omega)$ . El MSE de la imagen restaurada es el siguiente (Jain, 1989):

$$\text{MSE}_i = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_i^e(\omega) d\omega, \quad (191)$$

siendo  $S_i^e(\omega)$  la densidad espectral de potencia del ruido residual en la estimación  $i$ .  $S_i^e(\omega)$  se calcula de la siguiente manera:

$$S_i^e(\omega) = |1 - G_i(\omega)|^2 |T_i(\omega)|^2 + |G_i(\omega)|^2 N_r(\omega). \quad (192)$$

Sustituyendo  $G_i(\omega)$  en la ecuación (192), la densidad espectral del ruido se puede reescribir como

$$S_i^e(\omega) = \frac{|T_i(\omega)|^2 N_r(\omega)}{|T_i(\omega)|^2 + N_r(\omega)}. \quad (193)$$

Se utilizan las siguientes suposiciones:

- Las medias del objeto y del fondo de la escena son positivas.

- El ruido aditivo en la referencia tiene media cero.
- La densidad espectral de potencia del ruido en la referencia es  $N_r(\omega) = L\zeta_r^2$ , donde  $L$  es una constante de normalización.

Para que se cumpla que  $\text{MSE}_1 < \text{MSE}_2$  se necesita que se cumpla la siguiente condición:

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} S_1^e(\omega) d\omega < \frac{1}{2\pi} \int_{-\infty}^{\infty} S_2^e(\omega) d\omega. \quad (194)$$

Para poder llevar a cabo el análisis, se consideran los puntos del espacio de parámetros en los que  $S_2^e(\omega)$  domina a  $S_1^e(\omega)$  en todas las frecuencias. Comparando los términos en la ecuación (194) en cada frecuencia, se debe cumplir que

$$\frac{\zeta_r^2 |T_2(\omega)|^2}{|T_2(\omega)|^2 + \zeta_r^2} - \frac{\zeta_r^2 |T_1(\omega)|^2}{|T_1(\omega)|^2 + \zeta_r^2} > 0. \quad (195)$$

Descartando los factores multiplicativos en ambos lados de la ecuación (195) y algunas manipulaciones algebraicas, la condición se reduce a

$$|T_2(\omega)|^2 - |T_1(\omega)|^2 > 0. \quad (196)$$

Sustituyendo los valores de  $T_1(\omega)$  y  $T_2(\omega)$  en la ecuación (196) y desarrollando términos:

$$|T(\omega)|^2 + \mu_s^2 |\delta(\omega) - W(\omega)|^2 + 2\mu_s \text{Re}\{T^*(\omega) [\delta(\omega) - W(\omega)]\} - |T(\omega)|^2 > 0, \quad (197)$$

donde  $W(\omega)$  es la transformada de Fourier de la región de soporte del objeto, es decir,  $\mathcal{F}\{1 - \bar{w}(x)\}$ . Es fácil ver que esta desigualdad siempre se cumple para  $\omega = 0$ . Para las frecuencias distintas de cero, se obtiene

$$\mu_s |W(\omega)|^2 > 2\text{Re}\{T^*(\omega) W(\omega)\}. \quad (198)$$

Sea  $s_t$  el tamaño del objeto. Comparando ambos lados de esta ecuación en el dominio

espacial, el lado izquierdo es la autocorrelación de la región de soporte con valor máximo  $\mu_s s_t$ ; el lado derecho es una correlación cruzada con valor máximo  $2\mu_t s_t$ . Si  $\sigma_t = 0$ , la condición se reduce a

$$\mu_s > 2\mu_t. \quad (199)$$

En caso contrario, en el pico de la correlación cruzada se están sumando  $s_t$  variables aleatorias con desviación estándar  $\sigma_t$ . Para determinar el umbral se necesita suponer una distribución para los valores de intensidad en el área del objeto y fijar un grado de confianza con el cual asegurar que la desigualdad se cumple. A falta de información sobre la distribución de la señal del objeto, se supone una distribución normal. Finalmente, la condición que se obtiene es la siguiente:

$$\mu_s > 2 \left( \mu_t + k \frac{\sigma_t}{\sqrt{s_t}} \right),$$

donde  $k$  es una constante que depende del nivel de confianza deseado.