

**Centro de Investigación Científica y de Educación  
Superior de Ensenada, Baja California**



**Programa de Posgrado en Ciencias  
en Ciencias de la Computación**

---

**Diseño de un sistema de seguimiento utilizando filtros de  
Kalman y filtros de correlación adaptativos**

Tesis

para cubrir parcialmente los requisitos necesarios para obtener el grado de  
Maestro en Ciencias

Presenta:

**Sergio Esteban Ontiveros Gallardo**

Ensenada, Baja California, México

2015

Tesis defendida por

# **Sergio Esteban Ontiveros Gallardo**

y aprobada por el siguiente Comité

---

Dr. Vitaly Kober  
Director del Comité

Dr. Josué Álvarez Borrego  
Dr. Hugo Homero Hidalgo Silva



---

Dra. Ana Isabel Martínez García  
Coordinador del Programa de Posgrado en Ciencias de la Computación

---

Dra. Rufina Hernández Martínez  
Director de Estudios de Posgrado

Resumen de la tesis que presenta Sergio Esteban Ontiveros Gallardo como requisito parcial para la obtención del grado de Maestro en Ciencias en Ciencias de la Computación.

## **Diseño de un sistema de seguimiento utilizando filtros de Kalman y filtros de correlación adaptativos**

Resumen aprobado por:

---

Dr. Vitaly Kober

Director de tesis

El seguimiento de objetos se utiliza comúnmente para aplicaciones como video vigilancia, monitoreo de tráfico en tiempo real y navegación de vehículos. En este trabajo se propone un sistema de seguimiento con la implementación de un filtraje de correlación adaptativo y una robusta predicción de estado debida al filtro de Kalman. El seguimiento se realiza por medio de múltiples detecciones del objeto en regiones reducidas de interés. Generamos un banco de filtros a partir de múltiples puntos de vista de un objetivo utilizando funciones discriminantes sintéticas (SDF) y un valor de capacidad de discriminación deseado. El enfoque adaptativo se utiliza para mejorar la capacidad de discriminación de los filtros sintetizados adaptándolos a múltiples tipos de fondos. Finalmente, presentamos los resultados obtenidos al evaluar el rendimiento del sistema propuesto en términos de eficiencia y precisión del seguimiento de objetos de detección con la ayuda de la simulación por computadora.

Palabras Clave: **Seguimiento, filtros de Kalman, filtros de correlación adaptativos.**

Abstract of the thesis presented by Sergio Esteban Ontiveros Gallardo as a partial requirement to obtain the Master of Science degree in Master in Computer Science in Computer Science.

### **Design of a tracking system using Kalman and adaptive correlation filters**

Abstract approved by:

---

Dr. Vitaly Kober

Thesis director

Object tracking is commonly used for applications such as video surveillance, motion based recognition, and vehicle navigation. In this work, a tracking system using adaptive correlation filters and robust Kalman prediction of target locations is proposed. Tracking is performed by means of multiple object detections in reduced frame areas. A bank of filters is designed from multiple views of a target using synthetic discriminant functions (SDF) and a desired discrimination capability value. An adaptive approach is used to improve the discrimination capability of the synthesized filters adapting them to multiple types of backgrounds. With the help of computer simulation, the performance of the proposed system in terms of detection efficiency and accuracy of object tracking is presented, evaluated and discussed.

Keywords: **Tracking, Kalman filter, adaptive correlation filters.**

## Dedicatoria

*A mi hermosa novia que estuvo a mi lado  
durante todo el proceso.*

## **Agradecimientos**

Mi más sincero agradecimiento:

A mi familia por su amor y su apoyo en todo. A Rosa por su amor, paciencia y apoyo incondicional.

A mi asesor, el Dr. Vitaly Kober, por la oportunidad y paciencia brindados.

A los miembros de mi comité de tesis, el Dr. Hugo Hidalgo y el Dr. Josué Álvarez, por sus valiosos comentarios.

A todos los compañeros con los que me tocó convivir en CICESE, en especial a la generación 2013. A mis compañeras América y Adriana por su amistad y apoyo. A Julia Díaz por ayudarme a resolver mis dudas siempre que fue requerido.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por brindarme el apoyo económico para realizar mis estudios de maestría.

Al Centro de Investigación Científica y de Educación Superior de Ensenada.

# Tabla de contenido

	Página
<b>Resumen en español</b>	<b>ii</b>
<b>Resumen en inglés</b>	<b>iii</b>
<b>Dedicatoria</b>	<b>iv</b>
<b>Agradecimientos</b>	<b>v</b>
<b>Lista de figuras</b>	<b>viii</b>
<b>Lista de tablas</b>	<b>x</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Definición del problema . . . . .	2
1.2. Objetivo general . . . . .	3
1.2.1. Objetivos específicos . . . . .	3
1.3. Importancia de la Investigación . . . . .	3
1.4. Limitaciones y suposiciones fundamentales . . . . .	3
1.5. Antecedentes . . . . .	4
1.5.1. Detección . . . . .	4
1.5.2. Seguimiento . . . . .	5
1.6. Organización de la tesis . . . . .	6
<b>2. Marco teórico</b>	<b>7</b>
2.1. Definición de imagen . . . . .	7
2.2. Sistemas lineales . . . . .	7
2.3. Transformada de Fourier (TF) . . . . .	8
2.3.1. Propiedades de la transformada de Fourier . . . . .	10
2.4. Transformada Discreta de Fourier (TDF) . . . . .	12
2.5. Correlación discreta . . . . .	14
2.6. Modelos de ruido . . . . .	15
2.6.1. Ruido aditivo . . . . .	15
2.6.2. Ruido multiplicativo . . . . .	16
2.6.3. Ruido impulsivo . . . . .	16
2.7. Distorciones geométricas . . . . .	16
2.7.1. Translación . . . . .	17
2.7.2. Rotación . . . . .	17
2.7.3. Escalamiento . . . . .	17
2.8. Filtro de Kalman . . . . .	18
2.8.1. El origen computacional del filtro . . . . .	19
2.8.2. El filtro de Kalman discreto . . . . .	20
2.9. Modelo cinemático . . . . .	21
2.10. Métricas de desempeño . . . . .	22
2.10.1. Razón señal a ruido (SNR) . . . . .	22
2.10.2. Error cuadrático medio (MSE) . . . . .	22
2.10.3. Métrica Energía de Pico de Correlación (PCE) . . . . .	23
2.10.4. Capacidad de discriminación (DC) . . . . .	23
2.10.5. Errores de localización (LE) . . . . .	24

## Tabla de contenido (continuación)

2.10.6. Tasa de reconocimiento . . . . .	24
2.10.7. Errores de clasificación . . . . .	24
<b>3. Filtros lineales clásicos</b>	<b>26</b>
3.1. Filtro digital . . . . .	26
3.2. Filtro de acoplamiento (MF) . . . . .	27
3.3. Filtro solo de fase (POF) . . . . .	28
3.4. Filtro inverso (IF) . . . . .	28
3.5. Filtros de correlación compuestos . . . . .	29
3.5.1. Funciones discriminantes sintéticas (SDF) . . . . .	30
3.5.2. Filtro de Mínimo Promedio de Energía de Correlación (MACE)	32
3.5.3. Funciones discriminantes sintéticas adaptativas (ASDF) . . . . .	33
3.6. Algoritmo básico para el reconocimiento de patrones por medio de filtros de correlación . . . . .	34
<b>4. Sistema de seguimiento propuesto</b>	<b>36</b>
4.1. Preprocesamiento . . . . .	36
4.2. Detección . . . . .	37
4.3. Predicción . . . . .	37
4.4. Resultados . . . . .	42
4.4.1. Elementos de prueba . . . . .	43
4.4.2. Parámetros utilizados . . . . .	43
4.4.3. Resultados experimentales . . . . .	45
4.4.3.1. Evaluación del banco de filtros . . . . .	46
4.4.3.2. Comparación del modelo de Kalman con el modelo cinemáti- co . . . . .	48
4.4.3.3. Errores de localización cuando el objeto se mantiene oculto	50
4.4.3.4. Seguidor de la orientación . . . . .	51
4.4.3.5. Tiempo de ejecución . . . . .	51
4.4.3.6. Discusión de los resultados . . . . .	52
<b>5. Conclusiones</b>	<b>55</b>
5.1. Conclusión . . . . .	55
5.2. Trabajo a futuro . . . . .	56
<b>Lista de referencias bibliográficas</b>	<b>57</b>



## Lista de figuras

Figura	Página	
1.	Podemos ver la imagen digital como un arreglo bidimensional de $m$ renglones y $n$ columnas donde cada punto en la figura representa un pixel (izquierda). La imagen muestra el orden lexicográfico que toman los pixeles de cada imagen de entrenamiento para formar la matriz $R$ (derecha). . . . .	31
2.	Diagrama de flujo del algoritmo iterativa para el diseño de ASDF . . . . .	34
3.	Ejemplo de implementación de filtros de correlación. Imágenes en el dominio espacial: (a) Imagen filtro y (b) Imagen problema. Respuesta calculadas para filtros clásicos. (c) MF, (d) POF y (e) IF. . . . .	35
4.	Ejemplo de respuesta de un filtro SDF en comparación con la salida de un filtro ASDF. Imágenes en el dominio espacial: (a) Imagen filtro $t_i$ , (b) Fondo utilizado para entrenamiento, (c) Imagen problema. Respuesta calculadas de los filtros SDF (d,f,h,j,l,n) y ASDF (e,g,i,k,m,ñ). (d) SDF objeto normal , (e) ASDF objeto normal, (f) SDF objeto escala 0.6, (g) ASDF objeto escala 1.4, (h) SDF objeto escala 0.6, (i) ASDF objeto escala 0.6, (j) SDF objeto rotación 90°, (k) ASDF objeto rotación 90°, (l) SDF objeto rotación 180° escala 1.4, (m) ASDF objeto rotación 180° escala 1.4, (n) SDF objeto rotación 270° escala 0.6 y (ñ) ASDF objeto rotación 270° escala 0.6. . . . .	38
5.	Diagrama de flujo de las operaciones básicas del sistema de seguimiento propuesto. . . . .	39
6.	Diagrama de flujo del sistema de seguimiento propuesto. . . . .	40
7.	Objetos de prueba. (a)Muñeco amarillo, (b)zapato negro, (c)pato sonriente, (d)pez rojo y blanco, (e)mono de nieve, (f)zapato, (g)gato amarillo, (h)oso de navidad blanco, (i)muñeco amarillo 3 y (j)vaselina. . . . .	43
8.	Fondos seleccionados de internet para pruebas. (a) Juguetes, (b) zapatos, (c) juguetes 2, (d) océano, (e) adornos de navidad, (f) zapatos 2, (g) juguetes 2, (h) juguetes 3, (i) juguetes 4 y (j) productos. . . . .	44
9.	Trayectorias creadas para pruebas. (a-e) Trayectorias (1-5), (f) objeto oculto primera posición, (g) objeto oculto segunda posición, (h) objeto oculto tercera posición . . . . .	44
10.	Datos resumidos: taza de reconocimiento y errores de clasificación. . . . .	48
11.	Comparación de modelos de seguimiento: Errores de localización máximo y promedio. . . . .	49
12.	Comparación de errores de localización del modelo con el objeto parcialmente oculto. . . . .	50
13.	Comparación gráfica de errores de estimación para objeto parcialmente oculto en trayectorias representativas. Posición real (amarillo), posición estimada con objeto presente (rojo) y posición estimada con objeto parcialmente oculto (verde). . . . .	51

14. Tiempo de ejecución del sistema. Donde Sn: Escena completa y Sncr: Región de interés. . . . . 52

## Lista de tablas

Tabla		Página
1.	Matriz de confusión para casos de clasificación. . . . .	24
2.	Argumentos de entrada de la función ConfigureKalmanFilter . . . . .	40
3.	Taza de reconocimiento y errores de clasificación del banco de filtros con secuencias de vídeo sin ruido. . . . .	46
4.	Taza de reconocimiento y errores de clasificación del banco de filtros con secuencias de vídeo en presencia de ruido aditivo con desviación estandar igual a 10. . . . .	46
5.	Taza de reconocimiento y errores de clasificación del banco de filtros con secuencias de vídeo en presencia de ruido aditivo con desviación estandar igual a 20. . . . .	47
6.	Taza de reconocimiento y errores de clasificación del banco de filtros con secuencias de vídeo en presencia de ruido aditivo con desviación estandar igual a 30. . . . .	47
7.	Taza de reconocimiento y errores de clasificación del banco de filtros con secuencias de vídeo en presencia de ruido aditivo con desviación estandar igual a 40. . . . .	47
8.	Errores de localización máximo y promedio para cada objeto con modelo de predicción cinemático. . . . .	49
9.	Errores de localización máximo y promedio para cada objeto con modelo de predicción de Kalman. . . . .	49
10.	Errores de localización máximo y promedio cuando el objeto se ocultó de forma parcial. . . . .	50
11.	Error máximo y promedio (En grados) para la estimación de la orientación. . . . .	51

# Capítulo 1. Introducción

---

El ser humano posee un complejo sistema de visión que le permite reconocer, clasificar e interpretar los objetos y el medio en el cual se desenvuelve. Aunque en la mayoría de los casos el reconocimiento del medio se hace de forma instantánea, el sistema visual requiere de todo un proceso para llevar a cabo su fin. Es así que la interpretación de imágenes artificiales mediante procesos computacionales como una simulación del sistema visual humano se ha convertido en un reto importante dentro de áreas como la robótica, visión e inteligencia artificial. Por otro lado, el desarrollo y la proliferación de equipo de cómputo con mejores capacidades de procesamiento de imagen y video han desencadenado una creciente demanda de técnicas de visión artificial, que aún con la contribución de estas tecnologías presenta ciertas limitantes cuando se tiene información incompleta o insuficiente para realizar un reconocimiento adecuado.

Este trabajo de investigación consiste en realizar la detección y el seguimiento de objetos presentes en escenas complejas utilizando secuencias de video producidas sintéticamente en el laboratorio, garantizando una variedad de escenarios y situaciones que permitan poner a prueba el método propuesto. Una secuencia de video está formada por un conjunto de cuadros o fotogramas proyectados a cierta frecuencia.

El seguimiento consta en una estimación confiable de la trayectoria de uno o varios objetivos cuando éstos se mueven a través del medio. El seguimiento de objetivos se estudia para el desarrollo de diversas aplicaciones como el monitoreo de tráfico, identificación e interceptación de objetos, navegación de robots autónomos, la interacción humano-computadora, entre otras. Recientemente se han propuesto varios métodos para mejorar el seguimiento de objetos basados en filtros de correlación (Ali *et al.*, 2014; Contreras *et al.*, 2013; Díaz-Ramírez *et al.*, 2014; Rodríguez *et al.*, 2010; Santiago-Ramírez *et al.*, 2012; Bolme *et al.*, 2010), sin embargo, la mayoría de estas propuestas resultan en una alta complejidad computacional cuando se implementan de forma digital.

Es posible identificar dos elementos que vuelven al seguimiento en un problema difícil de abordar: el objetivo y el ambiente (fondo) donde se desplaza; por un lado, el objetivo varía su posición de cuadro en cuadro, este movimiento del objeto se puede clasificar en rotación y translación, el más sencillo de estos problemas se presenta cuando el objeto se

traslada de forma perpendicular al eje de la cámara, debido a que hacerlo paralelamente implicaría un aumento o reducción de la perspectiva del objeto dependiendo si se aleja o se acerca; Por otro lado, los factores del ambiente incluyen variaciones en la iluminación y otros elementos que aparecen en la escena, la luz presente puede cambiar la manera de visualizar el objeto, las sombras también generan una gran problemática en este sentido al afectar la intensidad de discriminación del objeto en ciertas regiones, mientras que la presencia de otros objetos en la escena que presenten características parecidas al objetivo pueden ser la causa de falsas alarmas en la detección.

### **1.1. Definición del problema**

El seguimiento puede ser definido como el problema de la estimación de la trayectoria de un objeto en el plano de la imagen a través de una secuencia de vídeo. la estructura común del seguimiento consiste en efectuar la detección del objetivo y generar la estimación de su estado futuro (el estado de un objetivo en la escena se puede especificar como un vector de atributos tales como la posición, velocidad, aceleración etc.). Adicionalmente, un sistema de seguimiento puede proporcionar información como la orientación, área, o forma del objeto.

El seguimiento de objetos puede volverse complejo debido a diferentes factores que intervienen durante el proceso de captura de la escena, entre los cuales podemos destacar:

- La pérdida de información causada por la proyección del mundo 3D en una imagen 2D.
- El ruido en imágenes.
- El ocultamiento del objeto de interés.
- Requisitos de procesamiento en tiempo real.
- Entre otros.

## **1.2. Objetivo general**

Desarrollar un sistema basado en filtraje de Kalman y filtros de correlación adaptativos para la implementación de una metodología de reconocimiento y seguimiento de objetivos a través de escenas del mundo real.

### **1.2.1. Objetivos específicos**

- Obtener el conocimiento teórico necesario de las diferentes técnicas desarrolladas para el seguimiento de objetos.
- Desarrollo de banco de filtros capaz de hacer la discriminación de objetivos en cuadros de video sujetos a distorsiones geométricas.
- Elaboración de un algoritmo que acople el banco de filtros desarrollado a la implementación de filtros de Kalman para realizar el seguimiento de un objeto basado en predicciones de su estado futuro.
- Análisis estadístico del sistema propuesto mediante simulación por computadora, análisis teórico y pruebas en ambientes reales.
- Comparar los resultados obtenidos con los resultados de los métodos existentes.

## **1.3. Importancia de la Investigación**

El trabajo de investigación propuesto resulta sùmamamente importante como base para el desarrollo de nuevas tecnologías en diferentes áreas del conocimiento. El seguimiento de objetos es un tema de investigación que promueve el desarrollo tecnológico en áreas como medicina, la robótica, ciencias naturales, geografía, etc. Es así que mientras se requiera mayor precisión y eficiencia en el desempeño de estos sistemas, mayor será la necesidad de implementar nuevos algoritmos capaces de realizar el seguimiento de objetos.

## **1.4. Limitaciones y suposiciones fundamentales**

- El fondo de las escenas analizadas se mantiene fijo durante todo el proceso.

- El trabajo se centra en objetos rígidos sin articulaciones.
- El sistema propuesto se centra en el seguimiento de un solo objetivo.
- Los objetos de interés presentan solamente rotaciones dentro del plano.
- En todos los casos estudiados se asume que se tiene información de la apariencia de los objetos antes de iniciar el proceso de seguimiento.

## 1.5. Antecedentes

### 1.5.1. Detección

Uno de los métodos empleados para realizar la detección de los objetivos en los cuadros de la secuencia de video utiliza filtros de correlación. Un filtro de correlación puede diseñarse como un sistema lineal donde las coordenadas del valor máximo en la salida del sistema son estimaciones de las coordenadas del objetivo en la escena observada. Una característica importante de estos filtros es que son capaces de reconocer objetos aún cuando la imagen de referencia presenta ruido aditivo o cuando el objetivo sufre distorsiones geométricas (en escala y rotación por ejemplo) es posible aún diseñar un filtro capaz de hacer el reconocimiento. Algunos de estos métodos utilizan técnicas que se traducen en la substracción de pequeñas secciones de la escena actual para realizar el análisis y reducir el coste computacional del algoritmo (Contreras *et al.*, 2013; Díaz-Ramírez *et al.*, 2014; Miramontes-Jaramillo y Kober, 2013), otra alternativa para mejorar el coste es mediante la incorporación de una unidad de procesamiento gráfico (GPU). La principal ventaja de utilizar las GPU para el procesamiento de señales es que estos dispositivos pueden procesar grandes bloques de datos en paralelo. Por lo tanto, son muy adecuados para aplicaciones en tiempo real.

Existen también algunos métodos que utilizan el reconocimiento de los objetos a través de una caracterización del mismo por su histograma (Leichter *et al.*, 2010; Miramontes-Jaramillo y Kober, 2013; Zhou *et al.*, 2009). Otros métodos utilizan la detección de los objetos mediante la extracción de características esenciales de este mismo (Kalal *et al.*, 2011; Li *et al.*, 2012), utilizando un aprendizaje constante de las características que pueden obtenerse del objetivo a través de su desplazamiento en los cuadros de la imagen.

En contraste, otros métodos utilizan información a priori de los objetos a identificar, de esta manera, si utilizamos la información de un objeto sujeto a variaciones en tamaño y rotación es posible realizar un reconocimiento de patrones invariante a distorsiones mediante la implementación de filtros compuestos. Estos filtros se sintetizan mediante el entrenamiento de varias plantillas que representan diferentes vistas del objetivo, entonces un filtro compuesto puede ser utilizado para reconocer diferentes versiones del objetivo que se presentan de forma natural al trasladarse en la escena (Contreras *et al.*, 2013; Díaz-Ramírez *et al.*, 2014; Santiago-Ramírez *et al.*, 2012).

### 1.5.2. Seguimiento

Un método clásico que da solución al seguimiento de objetivos está dado por el filtro de Kalman (Li *et al.*, 2010; Liu *et al.*, 2013; Rodríguez *et al.*, 2010; Asaari *et al.*, 2014). El filtro de Kalman ofrece buenos resultados cuando el estado y las variables de la observación son aleatorias y gaussianas; sin embargo la labor de detección se lleva a cabo antes de la estimación de las posiciones. El filtro de Kalman requiere un estado inicial para cada objeto, por lo que la estimación del estado inicial se debe obtener primero con la detección de los objetivos. Otra propuesta de solución al seguimiento está dada por los filtros de partícula (Breitenstein *et al.*, 2009). Los filtro de partículas (FP) consisten en un modelo dinámico para la predicción y un modelo de observación para evaluar la probabilidad de un estado predicho. Este enfoque emplea la cadena de Markov (es decir, que sólo se basa en la información del curso y de la etapa de la última vez) para aproximar la solución para la estimación del estado de los objetivos.

Otro método exitoso en el seguimiento consiste en la implementación de algoritmos de desplazamiento promedio (Leichter *et al.*, 2010; Zhou *et al.*, 2009). Este algoritmo utiliza una forma básica (elipse) para representar al objetivo y utiliza el histograma como principal descriptor del objetivo. El algoritmo de desplazamiento promedio es computacionalmente eficiente para el seguimiento de la posición y orientación de un objetivo mediante la explotación de las transformaciones de perspectiva, un inconveniente principal del algoritmo de desplazamiento promedio es su alta sensibilidad a los cambios de iluminación y oclusiones.

Otro enfoque del problema descompone la tarea de seguimiento en seguimiento,



aprendizaje y detección (Kalal *et al.*, 2011). El seguidor rastrea al objeto de cuadro a cuadro, el detector localiza todas las apariencias que se han observado hasta el momento y corrige al rastreador si es necesario y en la etapa de aprendizaje se estiman los errores del detector y los actualiza para evitar estos errores en el estudio futuro. Además implementan métodos de aprendizaje semi-entrenado para dar soporte y afinar las predicciones fallidas y las detecciones falsas. La principal problemática con este enfoque es que el seguimiento puede irse a la deriva si el objetivo gira totalmente fuera del plano, además como se entrena solamente al detector y no al seguidor, el seguidor comete siempre los mismos errores.

## **1.6. Organización de la tesis**

Durante este trabajo de investigación proponemos la combinación de un filtrado de correlación adaptativo con una predicción de estado mediante la implementación del filtro de Kalman para seguir objetos con alta precisión. Con esta idea en mente el resto del documento se encuentra organizado de la siguiente manera.

En el capítulo II se describen algunas conceptos básicos de la teoría del procesamiento de imágenes así como los fundamentos matemáticos empleados, desde el concepto y tipos de ruido hasta las métricas consideradas para la evaluación del algoritmo propuesto. El capítulo III trata sobre la teoría de filtrado donde se da una breve descripción de los filtros de correlación clásicos como una introducción al procedimiento de correlación implementado. El capítulo IV tiene como propósito la descripción de la metodología propuesta y utilizada durante el presente trabajo de investigación así como una evaluación del algoritmo implementado en términos de su capacidad de discriminación, errores de localización y porcentaje de falsos positivos y falsos negativos. Finalmente, en el capítulo V se exponen las conclusiones finales y el trabajo a futuro.

## Capítulo 2. Marco teórico

---

### 2.1. Definición de imagen

Una imagen puede ser definida como una función bidimensional  $f(x, y)$  donde  $x$  y  $y$  son coordenadas espaciales y la amplitud de  $f$  para cada par de coordenadas  $(x, y)$  es conocida como la intensidad de la imagen en ese punto. Cuando los valores de  $x$ ,  $y$  y  $f$  son finitos y discretos, podemos llamar a la imagen como imagen digital. Sabemos entonces que una imagen está compuesta por un número finito de elementos, los cuales tienen una posición particular y un valor. Estos elementos son popularmente nombrados como pixeles. (Thompson *et al.*, 1981)

### 2.2. Sistemas lineales

Un sistema, es cualquier ente físico o proceso capaz de recibir una señal de entrada, o excitación  $f(x)$ , y transformarla en otra señal que denominaremos de salida o respuesta  $g(x)$ . Decimos entonces que un sistema  $T$  se puede representar por un proceso de transformación, de la siguiente manera:

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

Como puede apreciarse, la definición previa es de carácter muy general y esto se debe principalmente a que la teoría puede aplicarse al estudio de una gran cantidad de problemas reales de muy diversa naturaleza física. En el procesamiento de imágenes, una gran variedad de técnicas para modificar, mejorar o representar información visual se expresa en términos de sistemas lineales, para labores como el mejoramiento de contraste, eliminación de ruido, realce de contornos y hasta en labores de más alto nivel sobre los objetos que componen una imagen, como el reconocimiento de patrones.

En un proceso de correlación se involucran dos señales o imágenes: la primera (entrada), la escena y una plantilla almacenada o filtro (creado con información de la imagen de referencia); y la segunda (la salida), representada por el plano de correlación calculado. De este modo, un sistema basado en correlación presenta las ventajas de los sistemas

lineales, de forma más específica, a una subclase denominada sistemas lineales invariantes a desplazamientos (LSI). Estos sistemas pueden ser caracterizados completamente por la salida del sistema para una entrada en particular, la cual se denomina entrada puntual en el origen y a la salida resultante como respuesta al impulso.

Entonces, se dice que el sistema  $T$  es lineal si y solo si respeta el criterio de superposición, es decir, 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

$$T[af_1(x) + bf_2(x)] = aT[f_1(x)] + bT[f_2(x)], \quad (2)$$

para toda  $x$ . Así mismo, se dice que el sistema  $T$  es invariante a los desplazamientos si para cada  $f(x)$  se cumple:

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

para cualquier valor de  $x'$ . De lo anterior debe entenderse que un desplazamiento espacial de la señal de entrada produce la salida esperada, excepto por un desplazamiento de la misma magnitud.

En el procesamiento de imágenes, un factor importante de los sistemas lineales invariantes a desplazamientos es su relación con la transformada de Fourier ( $FT$ ). La transformada de Fourier  $H(u)$  de la respuesta al impulso  $h(x)$  de un sistema lineal se le conoce como respuesta en frecuencia o función de transferencia del sistema. Los filtros lineales en el procesamiento de imágenes se caracterizan en términos de su respuesta en frecuencia.

A continuación presentamos algunos conceptos que nos serán de utilidad en este trabajo con relación a la  $FT$ .

### 2.3. Transformada de Fourier (TF)

Diferentes campos incluyendo medicina, óptica, física e ingeniería electrónica usan la TF como una herramienta común. En general, la TF de una señal nos permite ver su

información en el dominio de la frecuencia en lugar de la información en el dominio del tiempo.

Dada una señal  $f(x)$ , la transformada de Fourier denotada por  $F(v)$  está definida como:

$$F(v) = \int_{-\infty}^{\infty} f(x)e^{-jvx} dx, \quad -\infty < v < \infty, \quad (4)$$

donde  $j = \sqrt{-1}$  y  $v$  es la variable continua de frecuencia.

Se dice que una señal  $f(x)$  tiene TF en el sentido ordinario, si la integral existe. La integral existe si  $f(x)$  tiene un número finito de discontinuidades, máximas y mínimas, dentro de un intervalo finito de tiempo, y si  $f(x)$  es completamente integrable, esto significa:

$$\int_{-\infty}^{\infty} |f(x)| dx < \infty. \quad (5)$$

De la misma manera, teniendo la TF  $F(v)$  de una señal, es posible obtener la señal original  $f(x)$  aplicando la Transformada Inversa de Fourier (IFT) dada:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(v)e^{jvx} dv. \quad (6)$$

Dado que en el presente trabajo realizaremos este tipo de transformaciones a imágenes, las cuales se pueden representar como secuencias (señales) discretas de espacio con dos componentes enteros, el análisis discreto será muy importante y lo explicaremos un poco más adelante. Las secuencias que utilizaremos pueden ser representadas entonces como  $g(x, y)$ , las cuales están definidas para todos los valores enteros de  $x$  y  $y$ . La TF en dos dimensiones estaría dada por:

$$F(g) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)e^{-j2\pi(ux+vy)} dx dy. \quad (7)$$

La transformada definida es por sí misma una función de valores complejos de dos

variables independientes  $u$  y  $v$ , a las cuales nos referimos generalmente como frecuencias. De forma similar, la Transformada Inversa de Fourier de una función  $G(u, v)$  ser representada por:

$$F^{-1}(G) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G(u, v) e^{j2\pi(ux+vy)} dudv. \quad (8)$$

### 2.3.1. Propiedades de la transformada de Fourier

Trabajar con una señal en el dominio de la frecuencia aporta una gran cantidad de información de la señal que al ser trabajada en el dominio del tiempo sería imposible descifrar. Sin embargo, para aprovechar de mejor manera todo lo que la señal puede ofrecernos, es necesario conocer ciertas propiedades de la TF. Tales como:

1.-Teorema de linealidad.  $F(\alpha g + \beta h) = \alpha F(g) + \beta F(h)$ ; esta propiedad nos dice que la transformada de la suma ponderada de dos o más funciones es igual a la suma ponderada de las transformadas de forma individual.

2.-Teorema de similaridad. Si  $F[g(x, y)] = G(u, v)$ , entonces:

$$F[g(ax, by)] = \frac{1}{|ab|} G\left(\frac{u}{a}, \frac{v}{b}\right). \quad (9)$$

Esta propiedad nos dice que cualquier escalamiento de las coordenadas en el dominio espacial  $(x, y)$  se traducirá en una contracción de las coordenadas en el dominio de la frecuencia  $(u, v)$  más un cambio en la amplitud total del espectro.

3.-Teorema de traslación. Si  $F[g(x, y)] = G(u, v)$ , entonces:

$$F[g(x - a, y - b)] = G(u, v) e^{-j2\pi(ua+vb)}. \quad (10)$$

Esta propiedad nos dice que una traslación en el dominio del espacio introduce una traslación lineal en el dominio de la frecuencia.

4.-Teorema de Parseval. Si  $F[g(x, y)] = G(u, v)$ , entonces:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g(x, y)|^2 dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |G(u, v)|^2 du dv. \quad (11)$$

En esta propiedad, la integral de la parte izquierda puede ser interpretada como la energía contenida en la forma de onda  $g(x, y)$ . Lo cual lleva a la idea de que la cantidad  $|G(u, v)|^2$  puede ser interpretada como la densidad de energía en el dominio de la frecuencia.

5.-Teorema de convolución. Si  $F[g(x, y)] = G(u, v)$  y  $F[h(x, y)] = H(u, v)$ , entonces:

$$F\left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\xi, \eta) h(x - \xi, y - \eta) d\xi d\eta\right] = G(u, v) H(u, v). \quad (12)$$

Esta propiedad nos dice que la convolución de dos funciones en el dominio espacial es equivalente a la multiplicación de sus transformadas individuales.

6.-Teorema de autocorrelación. Si  $F[g(x, y)] = G(u, v)$ , entonces:

$$F\left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\xi, \eta) g^*(\xi - x, \eta - y) d\xi d\eta\right] = |G(u, v)|^2. \quad (13)$$

De forma similar

$$F[|g(x, y)|^2] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G(\xi, \eta) G^*(\xi - u, \eta - v) d\xi d\eta. \quad (14)$$

Este teorema es considerado un caso especial del teorema de correlación en el cual se correlacionan  $g(x, y)$  con  $g^*(-x, -y)$ , donde \* es el complejo conjugado.

La TF está envuelta por una gran cantidad de propiedades, que para casos muy específicos son útiles, sin embargo, para el propósito de esta tesis se tomaron solo estas seis propiedades que nos parecieron más relevantes.

## 2.4. Transformada Discreta de Fourier (TDF)

El análisis de señales en tiempo discreto es usualmente utilizado en el procesamiento digital de señales. Para realizar un análisis en frecuencia de una señal en tiempo discreto  $x(n)$ , necesitamos convertir la secuencia en el dominio del tiempo a su representación equivalente en el dominio de la frecuencia. La transformada de Fourier para la señal ya mencionada será representada entonces como  $X(u)$ . Sin embargo,  $X(u)$  es una función de frecuencia continua y por lo tanto no es una representación conveniente para la señal  $x(n)$ . Computacionalmente es imposible llevar a cabo la representación de una señal continua. De forma digital, solo se harán aproximaciones a una señal continua, por eso mismo representaremos la secuencia  $x(n)$  con muestras de su espectro  $X(u)$ , de esta manera su representación en la frecuencia nos llevaría a la Transformada Discreta de Fourier (TDF)(Proakis y Manolakis, 1996), la cual es una poderosa herramienta computacional para la realización de análisis frecuenciales de señales en tiempo discreto.

Dada una señal en tiempo discreto  $x[n]$ , la TDF estará dada por:

$$X[u] = \sum_{n=-\infty}^{\infty} x[n]e^{-j\Omega n}. \quad (15)$$

La TDF es en general una función valuada de complejos de la variable real  $\Omega$ . Notar que la ecuación anterior es la equivalente en tiempo discreto a la TF, donde la integral es reemplazada por una sumatoria.

Se dice que una señal  $x[n]$  en tiempo discreto tiene TDF en el sentido ordinario, si la sumatoria converge, es decir, si la sumatoria es finita para todos los valores reales de  $\Omega$ , esto significa:

$$\sum_{n=-\infty}^{\infty} |x[n]| < \infty. \quad (16)$$

Tomando en cuenta nuevamente que se trabajará con secuencias discretas bidimensionales, podemos considerar una secuencia  $x_0(n_1, n_2)$ , la cual es periódica con periodo de  $N_1 \times N_2$ . Supongamos que formamos una señal con extensión finita  $x(n_1, n_2)$  que con-

tenga un periodo de  $x_0(n_1, n_2)$  y todos los demás valores en cero. Es decir:

$$x(n_1, n_2) = x_0(n_1, n_2)R_{N_1 \times N_2}(n_1, n_2), \quad (17)$$

donde

$$R_{N_1 \times N_2}(n_1, n_2) = \begin{cases} 1, & 0 \leq n_1 \leq N_1 - 1, 0 \leq n_2 \leq N_2 - 1. \\ 0, & \text{cualquier otro caso.} \end{cases} \quad (18)$$

Esto muestra claramente que podemos determinar  $x_0(n_1, n_2)$  de  $x(n_1, n_2)$  por:

$$x_0(n_1, n_2) = \sum_{r_1=-\infty}^{\infty} \sum_{r_2=-\infty}^{\infty} x(n_1 - r_1 N_1, n_2 - r_2 N_2). \quad (19)$$

Ahora, consideremos  $X_0(u, v)$  la Serie Discreta de Fourier de  $x_0(n_1, n_2)$ . Supongamos que formamos una señal con extensión finita  $X(u, v)$  que contenga un periodo de  $X_0(u, v)$  y todos los demás valores en cero. Es decir:

$$X(u, v) = X_0(u, v)R_{N_1 \times N_2}(u, v). \quad (20)$$

Notamos aquí también que el proceso es invertible ya que podemos obtener  $X_0(u, v)$  de  $X(u, v)$  por:

$$X_0(u, v) = \sum_{r_1=-\infty}^{\infty} \sum_{r_2=-\infty}^{\infty} X(u - r_1 N_1, v - r_2 N_2). \quad (21)$$

Finalmente tenemos nuestro par de TDF dadas por:

$$X(u, v) = \begin{cases} \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} x(n_1, n_2) e^{-\rho_1 n_1} e^{-\rho_2 n_2}, & 0 \leq u \leq N_1 - 1, 0 \leq v \leq N_2 - 1. \\ 0, & \text{cualquier otro caso} \end{cases} \quad (22)$$



$$x(n_1, n_2) = \begin{cases} \frac{1}{N_1 N_2} \sum_{u=0}^{N_1-1} \sum_{v=0}^{N_2-1} X(u, v) e^{-j \frac{2\pi}{N_1} u n_1} e^{-j \frac{2\pi}{N_2} v n_2}, & 0 \leq n_1 \leq N_1 - 1, 0 \leq n_2 \leq N_2 - 1. \\ 0, & \text{cualquier otro caso} \end{cases} \quad (23)$$

donde,  $\rho_1 = j(\frac{2\pi}{N_1})un_1$  y  $\rho_2 = j(\frac{2\pi}{N_2})vn_2$ . De las transformadas anteriores tenemos que un punto en  $N_1 \times N_2$  de la secuencia  $x(n_1, n_2)$  está representado en el dominio de la frecuencia por un punto en  $N_1 \times N_2$  de la secuencia  $X(u, v)$ . La secuencia  $X(u, v)$  es conocida como la Transformada Discreta de Fourier de  $x(n_1, n_2)$ , mientras que  $x(n_1, n_2)$  es conocida como la Transformada Discreta de Fourier Inversa (TDF Inversa) de  $X(u, v)$ .

## 2.5. Correlación discreta

Cuando se intenta simular el sentido de la vista de los humanos mediante algún tipo de algoritmo, es necesaria una comparación de ciertos patrones que contenga la imagen que quiera reconocerse con patrones provenientes de la imagen visualizada, este tipo de reconocimiento de patrones se hace a través de una función matemática conocida como correlación, la cual nos permite determinar qué tan parecido es un objeto a otro de forma cuantificada.

Para una señal en dos dimensiones, podemos definir a la correlación discreta como:

$$f(x, y) \otimes g(x, y) = \frac{1}{MN} \sum_{m=0}^M \sum_{n=0}^N f(m, n) g(x + m, y + n). \quad (24)$$

Lo cual podemos escribir como:

$$F[f(x, y) \otimes g(x, y)] = F(u, v) G^*(u, v). \quad (25)$$

Que nos indica que la TF de la correlación de dos señales en tiempo es igual al producto en la frecuencia de la TF de la primer señal por el complejo conjugado de la TF de la segunda señal. Así mismo tenemos:

$$F[f(x, y)g^*(x, y)] = F(u, v) \otimes G(u, v). \quad (26)$$

Aquí se obtiene que la TF del producto en el espacio de una señal por el complejo conjugado de otra señal es equivalente a la correlación de la TF de estas dos señales.

Finalmente concluimos que la correlación la podemos calcular como:

$$f(x, y) \otimes g(x, y) = F^{-1}[F(u, v)G^*(u, v)]. \quad (27)$$

Actualmente, algunos sistemas de correlación existentes en la actualidad permiten en ciertos casos reconocer objetos aún cuando este ha sido afectado en escala, posición y rotación.

## 2.6. Modelos de ruido

Las imágenes que se obtienen para mostrar o registrar información 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 se vuelve de suma importancia principalmente para las labores de detección y localización de objetos en imágenes.

Se denomina ruido a todas aquellas degradaciones que alteran los niveles de intensidad de la imagen capturada. El ruido se produce debido a imperfecciones en la fabricación de los sensores, a variaciones de voltajes con los que operan los sensores, entre otras. La visibilidad del ruido en las imágenes depende de factores como el tipo de ruido, intensidad y el fondo de la imagen.

A continuación presentamos algunas de las distorsiones más comunes consideradas para el diseño de filtros.

### 2.6.1. Ruido aditivo

El ruido aditivo con media cero se utiliza para modelar ruido debido a las imperfecciones de sensores y a los errores de cuantización. Una señal contaminada por ruido aditivo

se representa como:

$$S(x) = t(x) + n_s(x), \quad (28)$$

donde  $t(x)$  es la señal original sin perturbar y  $s(x)$  es la señal observada contaminada por ruido  $n_s(x)$

### 2.6.2. Ruido multiplicativo

El ruido multiplicativo se utiliza para modelar escenas bajo condiciones de iluminación no uniforme a lo largo de la imagen. El modelo formal es:

$$S(x) = t(x) * n_m(x), \quad (29)$$

donde  $n_m(x)$  denota el ruido multiplicativo. Este modelo frecuentemente se trata como el modelo aditivo aplicando la operación logaritmo a la escena de entrada.

### 2.6.3. Ruido impulsivo

Este ruido se caracteriza porque degrada 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. El modelo está dado por:

$$S(x) = (1 - e(x))t(x) + e(x)n_i(x). \quad (30)$$

Un caso particular del ruido impulsivo es el ruido sal y pimienta, donde  $n_i(x)$  toma valores extremos.

## 2.7. 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. Las distorsiones geométricas se clasifican principalmente en dos tipos:

### 2.7.1. Translación

Las distorsiones por translación se deben a cambios en la posición del objeto de interés en el plano de la imagen con respecto a la posición del objeto en la imagen filtro. La operación de translación es representada como una suma de vectores como se muestra:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a_x \\ a_y \end{pmatrix}. \quad (31)$$

### 2.7.2. Rotación

La distorsión por rotación se debe a cambios de la perspectiva del objeto en relación a la posición del dispositivo de captura. La rotación se clasifica en dos tipos; Rotaciones en el plano, que se refiere a las rotaciones que hace el objeto de forma perpendicular al eje de la cámara; Rotaciones fuera del plano, que se refiere a las rotaciones que realiza el objeto sobre el eje de la cámara. La operación de rotación puede expresarse de forma matricial:

$$T_R = \begin{pmatrix} \text{Cos}\theta & \text{Sen}\theta \\ -\text{Sen}\theta & \text{Cos}\theta \end{pmatrix}. \quad (32)$$

### 2.7.3. Escalamiento

La distorsión por escalamiento se presenta cuando un objeto cambia su distancia entre él y el dispositivo de captura, de manera que cambian las dimensiones aparentes en la escena capturada. La operación de escalamiento está dada por la siguiente matriz:

$$T_E = \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix}, \quad (33)$$

donde  $\alpha$  representa el factor de escala.

## 2.8. Filtro de Kalman

Los filtros de Kalman, desarrollados por Kalman (1960), consisten esencialmente en un conjunto de ecuaciones matemáticas que implementan un estimador óptimo de predicción-corrección que minimiza la media del error cuadrático. Este filtro aborda el problema de tratar de estimar el estado  $x \in \mathbb{R}^n$  de un proceso controlado en tiempo discreto que se define mediante el siguiente sistema de ecuaciones de estado (Bishop *et al.*, 2001):

$$x_k = Ax_{k-1} + Bu_k + w_{k-1}, \quad (34)$$

donde la matriz  $A$  de tamaño  $n \times n$  relaciona el estado del paso  $k - 1$  con el estado  $k$ , la matriz  $B$  de tamaño  $n \times l$  relaciona la entrada de control opcional  $u \in \mathbb{R}^l$  con el estado  $x$  y una medición  $z \in \mathbb{R}^m$ :

$$z_k = Hx_k + v_k, \quad (35)$$

donde la matriz  $H$  de tamaño  $m \times n$  contiene la información de la medición  $z_k$ . Las variables aleatorias  $w_k$  y  $v_k$  representan el ruido del proceso y el ruido de la medición respectivamente, los cuales se suponen como ruido blanco, independientes uno del otro y con distribución de probabilidad normal:

$$p(w) \sim N(0, Q), \quad (36)$$

$$p(v) \sim N(0, R). \quad (37)$$

En términos de seguimiento, nuestro principal interés al utilizar el filtro de Kalman se debe a que este método está capacitado para brindar información respecto a la orientación del objetivo, además como ya se mencionó anteriormente el proceso de retroalimentación del filtro mediante un proceso iterativo da como resultado una reducción del error

cuadrático medio entre la medición real y la estimada, así mismo, la implementación de este filtro colabora en la reducción de errores por oclusión de los objetos en las imágenes.

### 2.8.1. El origen computacional del filtro

Si decimos que  $\hat{x}_k^- \in \mathfrak{R}^n$  es nuestro estado estimado *a priori*, el paso  $k$  dada la información obtenida en el paso  $k - 1$ , y siendo  $\hat{x}_k \in \mathfrak{R}^n$  nuestra estimación de estado *a posteriori* el paso  $k$  dada la medición  $z_k$ , es posible definir los errores de la estimación *a priori* y *a posteriori* como:

$$e_k^- = x_k - \hat{x}_k^-, \quad (38)$$

y

$$e_k = x_k - \hat{x}_k. \quad (39)$$

De aquí, la covarianza del error estimado *a priori* estaría dado por

$$P_k^- = E[e_k^- e_k^{-T}], \quad (40)$$

y la covarianza del error estimado *a posteriori* por

$$P_k = E[e_k e_k^T]. \quad (41)$$

Trabajar con Kalman tiene como objetivo encontrar una ecuación que calcule una estimación del estado *a posteriori*  $\hat{x}_k$  como una combinación lineal de una estimación *a priori*  $\hat{x}_k^-$  y una diferencia ponderada entre la medición actual  $z_k$  y una predicción de la medición  $H\hat{x}_k^-$  como se muestra a continuación.

$$\hat{x}_k = \hat{x}_k^- + K(z_k - H\hat{x}_k^-). \quad (42)$$

De donde, la diferencia  $(z_k - H\hat{x}_k^-)$  es conocida como la *innovación* de la medición, o el *residual*. Aquí, un residual igual a cero implicaría que la predicción y la medición obtenida empatan perfectamente. La matriz  $K$  de la ecuación 42 conocida como el factor de *ganancia* de Kalman es elegido de tal manera que minimice la covarianza del error estimado *a posteriori* de la ecuación 41.

### 2.8.2. El filtro de Kalman discreto

El filtro de Kalman estima un proceso utilizando una forma de control por retroalimentación: El filtro estima el estado del proceso en cierto momento y luego obtiene la retroalimentación en forma de mediciones (ruidosas). Las ecuaciones para el filtro de Kalman se dividen en dos grupos: Las ecuaciones de actualización del tiempo (“time update”) o de *predicción* y las ecuaciones de actualización de la medición (“measurement update”) o de *corrección*. El primer tipo de ecuaciones son responsables de la proyección del estado al momento  $k$ , tomando como referencia el estado en el momento  $k - 1$ ; el segundo tipo de la retroalimentación, es decir, incorporan nueva información dentro de la estimación anterior con lo cual se llega a una estimación mejorada del estado.

Las ecuaciones de Kalman para la predicción están dadas por:

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_k, \quad (43)$$

$$P_k^- = AP_{k-1}A^T + Q. \quad (44)$$

Notese aquí que este par de ecuaciones relaciona la estimación del estado y la covarianza del paso  $k - 1$  con el paso  $k$ .

Las ecuaciones para la corrección son:

$$K_k = P^- H^T (HP_k^- H^T + R)^{-1} \quad (45)$$

$$\hat{x} = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-) \quad (46)$$

$$P_k = (I - K_k H)P_k^- \quad (47)$$

En esta etapa de corrección, lo primero es calcular la ganancia  $K_k$  mediante la ecuación 45. Lo siguiente es obtener una medición del proceso para obtener  $z_k$ , y con esto generar una estimación de estado *a posteriori*, como puede verse en 46. Como último paso, obtenemos una estimación del error de la covarianza *a posteriori* via la ecuación 47.

## 2.9. Modelo cinemático

La implementación del modelo cinemático nos brinda una alternativa para la solución a la estimación de estado para un objeto en un ambiente de seguimiento, (Li y Jilkov, 2003). Esta técnica realiza la predicción del estado del objeto en la escena subsecuente  $t+1$  con base en las estimaciones de estado realizadas para el tiempo  $t-1$  y  $t$ . El comportamiento del movimiento del objeto se caracteriza por las siguientes ecuaciones:

$$\alpha_{k+1} = \alpha_k + \frac{\text{sen}(\omega_k \Delta_t)}{\omega_k} \dot{\alpha}_k - \frac{1 - \text{cos}(\omega_k \Delta_t)}{\omega_k} \dot{\beta}_k + a_{\alpha,k} \frac{\Delta_t^2}{2},$$

$$\beta_{k+1} = \beta_k + \frac{1 - \text{cos}(\omega_k \Delta_t)}{\omega_k} \dot{\alpha}_k + \frac{\text{sen}(\omega_k \Delta_t)}{\omega_k} \dot{\beta}_k + a_{\beta,k} \frac{\Delta_t^2}{2},$$

$$\dot{\alpha}_{k+1} = \text{cos}(\omega_k \Delta_t) \dot{\alpha}_k - \text{sen}(\omega_k \Delta_t) \dot{\beta}_k + a_{\alpha,k} \Delta_t,$$

$$\dot{\beta}_{k+1} = \text{sen}(\omega_k \Delta_t) \dot{\alpha}_k + \text{cos}(\omega_k \Delta_t) \dot{\beta}_k + a_{\beta,k} \Delta_t,$$

$$\omega_{k+1} = \omega_k + a_{\omega,k}. \quad (48)$$



Donde las variables  $\alpha_k$  y  $\beta_k$  representan la posición del objeto en coordenadas cartesianas,  $\dot{\alpha}_k$  y  $\dot{\beta}_k$  son componentes de velocidad en las direcciones  $\alpha$  y  $\beta$ , respectivamente, y  $\omega_k$  es la velocidad angular del objeto. Aquí, las variables aleatorias  $a_{\alpha,k}$  y  $a_{\beta,k}$  representan los componentes de aceleración en las direcciones  $\alpha$  y  $\beta$ , respectivamente, y  $a_{\omega,k}$  es la aceleración angular.

Esta técnica ha demostrado un buen desempeño en la estimación de estado de objetos al trasladarse en escenas aún en la presencia de ruido en la medición (Contreras *et al.*, 2013; Díaz-Ramírez *et al.*, 2014; Santiago-Ramírez *et al.*, 2012; Miramontes-Jaramillo y Kober, 2013).

## 2.10. Métricas de desempeño

Las métricas de desempeño calculan la eficiencia con la cual está trabajando el sistema utilizado en el reconocimiento de patrones. Existe una gran cantidad de métricas para medir la eficiencia de los filtros de correlación, entre los cuales se encuentra: la relación señal a ruido (SNR), Error cuadrático medio (MSE), la relación energía del pico de correlación (PCE), la relación de la capacidad de discriminación (DC), y los errores de localización (LE), solo por mencionar las más utilizadas.

### 2.10.1. Razón señal a ruido (SNR)

Esta métrica caracteriza la variación del pico de correlación deseado a la salida del filtro con respecto al ruido en la entrada. Se define como:

$$SNR = \frac{|E\{y(x_s)\}|^2}{Var\{y(x_s)\}}; \quad (49)$$

donde  $y(x)$  denota la salida del filtro y 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.

### 2.10.2. Error cuadrático medio (MSE)

Una alternativa en el diseño de filtros de correlación consiste en especificar la salida deseada y diseñar un filtro que minimice el error cuadrático promedio en una secuencia

de imágenes. La salida de correlación ideal es un pico de muy alta intensidad en la localización del objeto sobre un fondo plano. La diferencia entre la salida ideal  $y^i(x)$  y la salida real del filtro está dada por:

$$MSE = E\left\{\frac{1}{\Gamma} \int_{-\infty}^{\infty} |y^i(x) - y(x)|^2 dx\right\}, \quad (50)$$

donde  $\Gamma$  es la duración espacial de la salida.

### 2.10.3. Métrica Energía de Pico de Correlación (PCE)

Esta métrica representa la relación que existe entre la energía en el origen medida en el plano de correlación y la energía total en el plano de correlación. Se define entonces como:

$$PCE = \frac{|E\{c(0, 0)\}|^2}{E\{|c(x, y)|^2\}}. \quad (51)$$

De donde la parte del numerador representa el valor esperado de la intensidad del pico de correlación y el denominador representa el valor esperado de la energía promedio en el plano de correlación.

### 2.10.4. Capacidad de discriminación (DC)

La métrica DC representa la capacidad del filtro de discriminar entre un objeto deseado y estructuras similares en la escena de entrada, esta definida como:

$$DC = 1 - \left|\frac{y_{max}^b}{y_{max}^t}\right|^2, \quad (52)$$

donde  $y_{max}^b$  y  $y_{max}^t$  denotan los valores de mayor intensidad en las regiones del fondo y del objeto, respectivamente. Valores de la DC cercanos a uno indican una buena capacidad del filtro de discriminar objetos no deseados de los objetos de la clase verdadera. Si el reconocimiento falla, es decir, ocurre un error, el valor de la DC puede tomar valores negativos.

### 2.10.5. Errores de localización (LE)

Se refiere a la diferencia entre la localización estimada del objeto y su localización real en la escena. Esta métrica está definida como (Kober y Campos, 1996):

$$LE = [(x - \hat{x})^2 + (y - \hat{y})^2]^{1/2}, \quad (53)$$

donde  $(x, y)$  y  $(\hat{x}, \hat{y})$  son las coordenadas real y estimadas, respectivamente.

### 2.10.6. Tasa de reconocimiento

En cada imagen de prueba se espera la presencia del objeto, la tasa de reconocimiento muestra el porcentaje de las imágenes de prueba estudiadas que cumplen esta condición y se calcula como:

$$Tasa\ de\ reconocimiento = \frac{Detecciones\ exitosas}{Detecciones\ esperadas} \times 100. \quad (54)$$

### 2.10.7. Errores de clasificación

En el reconocimiento de patrones es común encontrar casos en los cuales el método de detección utilizado es incapaz de determinar la presencia o ausencia del objeto de interés en la escena de forma adecuada.

Definiendo nuestras hipótesis como:

$H_0$ : El objeto está presente.

$H_1$ : El objeto no está presente.

Los errores de clasificación se pueden representar en la matriz de confusión de la tabla 1.

**Tabla 1: Matriz de confusión para casos de clasificación.**

	$H_0$ es cierta	$H_1$ es cierta
Se determinó $H_0$	No hay error (verdadero positivo)	Error de tipo II (falso negativo)
Se determinó $H_1$	Error de tipo I (falso positivo)	No hay error (verdadero negativo)

Donde el error de tipo I o falso positivo (Fp), es el error que se comete cuando no se determina la hipótesis  $H_0$  siendo esta verdadera. Por otro lado, el error de tipo II o falso negativo (Fn), se comete cuando se determina la hipótesis  $H_0$  siendo esta falsa.

## Capítulo 3. Filtros lineales clásicos

---

Este capítulo tiene por objetivo proporcionar ideas básicas para el uso de la correlación como una herramienta para el reconocimiento de patrones. La correlación involucra dos señales o imágenes, una de referencia (objetivo) y una de prueba (escena). La imagen de referencia es correlacionada con una imagen de prueba para detectar y localizar algún patrón en específico de la imagen de referencia. Esto es, la imagen de referencia se coloca en la parte superior izquierda de la escena analizada y se lleva a cabo la multiplicación de píxeles entre las dos matrices, todos los valores se suman para producir un valor de salida de correlación. El proceso se repite desplazando la imagen objetivo a la derecha y hacia abajo, produciendo de este modo una matriz de dos dimensiones como salida, llamada *plano de correlación*. Por lo tanto, la correlación se puede considerar como un sistema con una entrada (la escena), una plantilla o filtro (derivado de la imagen de referencia), y una salida (el plano de correlación)(Kumar *et al.*, 2005).

En un escenario de detección pueden presentarse diferentes factores que afecten la discriminación de los filtros de correlación, objetos con características parecidas al objeto de referencia pueden obtener valores grandes de correlación, lo que nos llevaría a falsas detecciones o clasificaciones. Otro problema puede surgir cuando los objetos presentan distorsiones geométricas como desplazamientos laterales, cambios de escala, rotación, etc. .

### 3.1. Filtro digital

La correlación se puede implementar tanto para el procesamiento óptico como digital. Los filtros digitales tienen su origen teórico en los filtros análogos utilizados en electrónica. La utilización de la computadora como un medio de simulación permitió el desarrollo de una gran cantidad de procesos electrónicos de forma digital, entre ellos la realización e implementación de filtros digitales sin la necesidad de armar ni conectar un solo componente.

Existe una gran cantidad de variantes en cuanto a la forma de desarrollo de un filtro en base al dominio al que se está trabajando, sin embargo, la base matemática es la misma. Un filtro digital es básicamente un algoritmo que permite el paso de ciertos datos

y reduce, elimina o transforma el valor de otros datos al ser utilizado en una señal, que para el propósito de esta tesis, es una imagen.

Desde el desarrollo del filtro de acoplamiento clásico propuesto por Lugt (1964), se han propuesto varios filtros de correlación para el reconocimiento de patrones. Estos filtros han sido ampliamente utilizados debido a sus características invariantes a desplazamiento laterales gracias a las propiedades de la transformada de Fourier y a la facilidad en la derivación de estos filtros para maximizar algún criterio de calidad en específico. En la literatura se pueden encontrar una gran cantidad de filtros de correlación, sin embargo, los discutidos en este trabajo de investigación se seleccionaron por brindar buenos resultados en experimentos similares a los presentados.

### 3.2. Filtro de acoplamiento (MF)

El filtro de acoplamiento (MF), propuesto por Lugt (1964), este filtro se caracteriza por maximizar el criterio SNR. Para conseguir esto, se requiere maximizar el numerador, que indica la intensidad esperada del pico de correlación en presencia del objeto, y minimizar el denominador, el cual corresponde a la densidad espectral del ruido. Podemos derivar el filtro a partir de la ecuación 49 y aplicar la desigualdad de Cauchy-Schwarz como se muestra a continuación:

$$\begin{aligned}
 SNR_{H(u,v)} &= \alpha \frac{|\int \int_{-\infty}^{\infty} T(u,v)H(u,v)dudv|^2}{\int \int_{-\infty}^{\infty} P_n(u,v)|H(u,v)|^2dudv} \\
 &= \alpha \frac{|\int \int_{-\infty}^{\infty} [\frac{T(u,v)}{\sqrt{P_n(u,v)}}][\sqrt{P_n(u,v)}H(u,v)]dudv|^2}{\int \int_{-\infty}^{\infty} P_n(u,v)|H(u,v)|^2dudv} \\
 &\leq \alpha \frac{|\int \int_{-\infty}^{\infty} \frac{|T(u,v)|^2}{P_n(u,v)}dudv \int \int_{-\infty}^{\infty} P_n(u,v)|H(u,v)|^2dudv|^2}{\int \int_{-\infty}^{\infty} P_n(u,v)|H(u,v)|^2dudv} \\
 &= \alpha \int \int_{-\infty}^{\infty} \frac{|T(u,v)|^2}{P_n(u,v)}dudv, \tag{55}
 \end{aligned}$$

donde se consigue la igualdad si y solo si

$$H(u, v) = \beta \frac{T^*(u, v)}{P_n(u, v)} \equiv H_{MF}(u, v), \quad (56)$$

con  $\beta$  una constante compleja arbitraria y  $P_n(u, v)$  la función de densidad espectral del ruido. Este filtro consiste básicamente en el complejo conjugado de la TF del objeto multiplicado por un factor constante que es inversamente proporcional a la densidad espectral de energía del ruido de entrada.

### 3.3. Filtro solo de fase (POF)

El filtro solo de fase (POF), propuesto por Horner y Gianino (1984), se deriva a partir del hecho de que la fase de la TF contiene información más significativa que el módulo. Este filtro produce picos de correlación más altos y más agudos que el MF. Esto se puede apreciar de forma intuitiva si consideramos al filtro solo de fase como una versión de realce de altas frecuencias, como las componentes de altas frecuencias de la TF están relacionadas con los bordes y la forma de un objeto, es decir con sus elementos característicos, el filtro solo de fase llega a ser más sensible a las diferencias entre imágenes.

El filtro solo de fase está definido como:

$$H(u, v) = \frac{T^*(u, v)}{|T(u, v)|} = e^{j\phi(u, v)}, \quad (57)$$

donde  $\phi(u, v)$  es la fase de la TF del objetivo. Este filtro con módulo unitario es considerado óptimo en términos de su relación pico-energía de correlación de la sección 2.10.3, sin embargo, tiene como principales desventajas su pobre capacidad de discriminación para objetivos de bajo contraste incorporado en un fondo complicado y bajo nivel de SNR en imágenes ruidosas.

### 3.4. Filtro inverso (IF)

El filtro inverso (IF) tiene como principal ventaja la optimización de la relación PCE de la ecuación 51, que básicamente indica un pico de correlación muy angosto. Si la imagen de referencia y la imagen problema son iguales, entonces se espera de salida una función  $\delta$  como resultado de la correlación.

Este filtro puede derivarse a partir de la ecuación 51 como se muestra a continuación:

$$\begin{aligned}
PCE_{H(u,v)} &= \alpha \frac{|\int \int_{-\infty}^{\infty} T(u,v)H(u,v)dudv|^2}{\int \int_{-\infty}^{\infty} |T(u,v)|^2|H(u,v)|^2dudv} \\
&= \alpha \frac{|\int \int_{-\infty}^{\infty} [\frac{T(u,v)}{|T(u,v)|}] [|T(u,v)|H(u,v)]dudv|^2}{\int \int_{-\infty}^{\infty} |T(u,v)|^2|H(u,v)|^2dudv} \\
&\leq \alpha \frac{\int \int_{-\infty}^{\infty} |T(u,v)|^2|H(u,v)|^2dudv \int \int_{-\infty}^{\infty} dudv}{\int \int_{-\infty}^{\infty} |T(u,v)|^2|H(u,v)|^2dudv} \\
&= \alpha \int \int_{-\infty}^{\infty} dudv = 1
\end{aligned} \tag{58}$$

y la igualdad se consigue si y solo si

$$H(u,v) = \beta \frac{T^*(u,v)}{|T(u,v)|^2} \equiv H_{IF}(u,v), \tag{59}$$

donde  $\beta$  es una constante arbitraria que indica la magnitud del impulso, en este caso, la magnitud del pico de correlación.

### 3.5. Filtros de correlación compuestos

Aún cuando cada uno de los filtros anteriormente descritos optimiza algún criterio de calidad, los filtros de correlación están diseñados utilizando información de la apariencia y de la forma del objeto de interés. Consecuentemente, el desempeño decrece rápidamente cuando la apariencia del objeto en la escena de entrada no coincide con la apariencia conocida al momento del diseño del filtro. Dentro del reconocimiento de patrones, uno de los principales retos consiste en diseñar sistemas invariantes, esto es, sistemas de reconocimiento que sean capaces de detectar la presencia de un objeto de interés en la escena de entrada, sin importar que este se encuentre desplazado lateralmente, rotado, presente distintas dimensiones o incluso con distintos niveles de iluminación (Fimbres-Castro *et al.*, 2012; Díaz-Ramírez *et al.*, 2014; Solorza-Calderón y Álvarez Borrego, 2015;



Ventura *et al.*, 2015).

Los filtros descritos durante este capítulo poseen ímplicitamente invariancia a los desplazamientos laterales del objeto de interés sobre la escena de entrada debido a las propiedades de la TF. Pero, para crear filtros más robustos con otro tipo de invariancias es necesario recurrir a otros métodos. El objetivo es diseñar un filtro  $H(u, v)$  con el cual obtengamos valores similares de salida tanto con imágenes distorsionadas como con las que no lo están, manteniendo una alta tasa de reconocimiento y un rango de error bajo.

Para agregar invarianza a los cambios de apariencia del objeto se introdujeron los *filtros de correlación compuestos*. Esta técnica consiste en emplear imágenes representativas de las posibles variaciones que pudiera presentar un objeto, proporcionadas 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 dos técnicas comúnmente utilizadas para el diseño de filtros de correlación compuestos.

### 3.5.1. Funciones discriminantes sintéticas (SDF)

El método de funciones discriminantes sintéticas (SDF), propuesto por Casasent (1984), consiste básicamente en realizar la combinación lineal de un conjunto de filtros de acoplamiento creados para diferentes versiones del objeto de interés. La ponderación asignada para cada imagen en la combinación lineal se elige de tal manera que la salida de correlación tome un valor deseado en las coordenadas del objeto. Por ejemplo, en un problema de dos clases, es común asignar el valor de 1 para objetos de interés y 0 para patrones a rechazar.

El diseño de filtros SDF se detalla a continuación. Sea  $t_i(x, y)$  el conjunto de  $N$  imágenes de entrenamiento, cada uno con  $d$  píxeles. La salida del filtro estaría dada por:

$$h(x, y) = \sum_{i=1}^N a_i t_i(x, y), \quad (60)$$

donde  $a_i, 1 \leq i \leq N$  son los coeficientes que deben satisfacer:

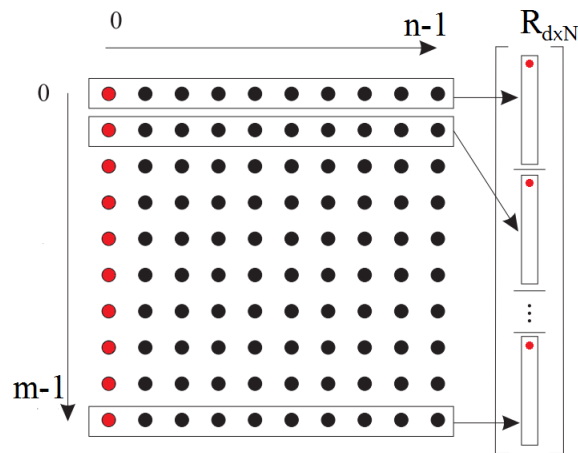


Figura 1: Podemos ver la imagen digital como un arreglo bidimensional de  $m$  renglones y  $n$  columnas donde cada punto en la figura representa un pixel (izquierda). La imagen muestra el orden lexicográfico que toman los pixeles de cada imagen de entrenamiento para formar la matriz  $R$  (derecha).

$$u_i = t_i(x, y) \otimes h(x, y), \quad (61)$$

donde  $u_i, 1 \leq i \leq N$  son los valores deseados para el origen en la salida de correlación para cada imagen de entrenamiento.

De forma matricial, podemos expresar las ecuaciones 60 y 61 como:

$$h = Ra, \quad (62)$$

$$u = R^T h, \quad (63)$$

donde  $R$  es una matriz de tamaño  $d \times N$ . Las columnas de  $R$  corresponden a las imágenes de entrenamiento en forma de vector (ver figura 1).

Sustituyendo la ecuación 62 en 63 obtenemos:

$$u = (R^T R)a. \quad (64)$$

Si  $(R^T R)$  es no singular, el filtro SDF puede representarse en su forma vectorial como:

$$h_{SDF} = R(R^T R)^{-1}u. \quad (65)$$

La desventaja principal en la implementación de SDF es la aparición de lóbulos laterales al área del objeto de interés, debido a que esta técnica tiene control exclusivamente sobre el pico de correlación y no en el resto del plano, lo que da como resultado una pobre capacidad de discriminación y una alta probabilidad de falsas alarmas.

### 3.5.2. Filtro de Mínimo Promedio de Energía de Correlación (MACE)

El filtro de Mínimo Promedio de Energía de Correlación (MACE) surge como una alternativa para la reducción de los lóbulos laterales presentes en el plano de correlación de salida respetando las condiciones de la ecuación 61. El objetivo es minimizar la energía promedio de la salida del filtro de manera que el fondo tienda a ser plano mientras que el pico de correlación sea lo más agudo posible. Al minimizar la energía en todo el plano, se suprimen los lóbulos laterales y por lo tanto se disminuye la probabilidad de falsas alarmas.

Este filtro se deriva a partir de la definición del promedio de la energía de correlación (ACE) que expresamos en su forma matricial de la siguiente manera (Kumar *et al.*, 2005):

$$ACE = h^+ Dh, \quad (66)$$

donde el superíndice  $+$  indica el transpuesto conjugado de un vector complejo y  $D = \frac{1}{d \cdot N} \sum_{i=1}^N T_i^* T_i$  representa una matriz diagonal de tamaño  $d \times d$ . De aquí y sujetos a la condición

$$T^+ h = d \cdot u, \quad (67)$$

se obtiene

$$h = D^{-1}T(T^+D^{-1}T)^{-1}U. \quad (68)$$

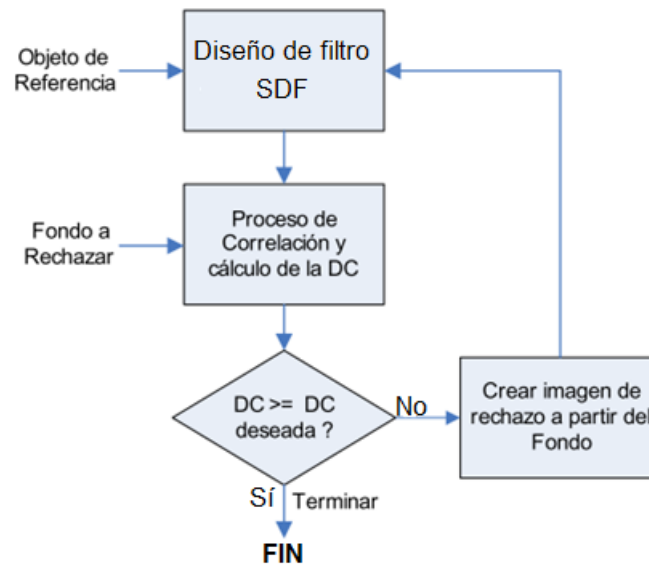
Aunque este filtro presenta ciertas ventajas sobre la técnica de SDF, aún presenta problemas debidos a la presencia de ruido en la imagen a reconocer, al mismo tiempo es sensible a las variaciones de los objetos entre-clases.

### 3.5.3. Funciones discriminantes sinteticas adaptativas (ASDF)

La técnica ASDF propuesta por González-Fraga *et al.* (2006), utiliza información sobre el objeto de interés y el fondo de la escena para diseñar un filtro de correlación adaptativo compuesto. El fondo se selecciona típicamente en base a la naturaleza y características del objeto de interés, para garantizar un pico de correlación afilado en presencia del objeto y valores cercanos a cero en el área de fondo. El filtro implementa un algoritmo iterativo para la reducción de los lóbulos laterales en el plano de salida, en el cual se sintetizan diferentes patrones del fondo como objetos de la clase falsa (patrones de rechazo). En la figura 2 se presenta un diagrama de flujo del método ASDF. El algoritmo ASDF sigue los siguientes pasos:

1. Sintetizar un filtro SDF utilizando diferentes versiones del objeto de interés.
2. Realizar el cálculo de la correlación entre el filtro SDF y el fondo típico seleccionado.
3. Calcular el valor DC del filtro.
4. Si el valor DC calculado es igual o mayor a un valor deseado, entonces el proceso iterativo termina; de otro modo el algoritmo continua con el paso 5.
5. Crear un nuevo patrón de rechazo centrado en la posición de mayor intensidad del plano de correlación calculado en el paso 2 y agregar dicho patrón como un objeto de la clase falsa.
6. Regresar al paso 1, utilizando el nuevo patrón de rechazo para la sintetización del nuevo filtro SDF.

El principal problema con la implementación de estos filtros es que su desempeño decae con un número creciente de imágenes de entrenamiento, por eso mismo se han pro-



**Figura 2: Diagrama de flujo del algoritmo iterativa para el diseño de ASDF**

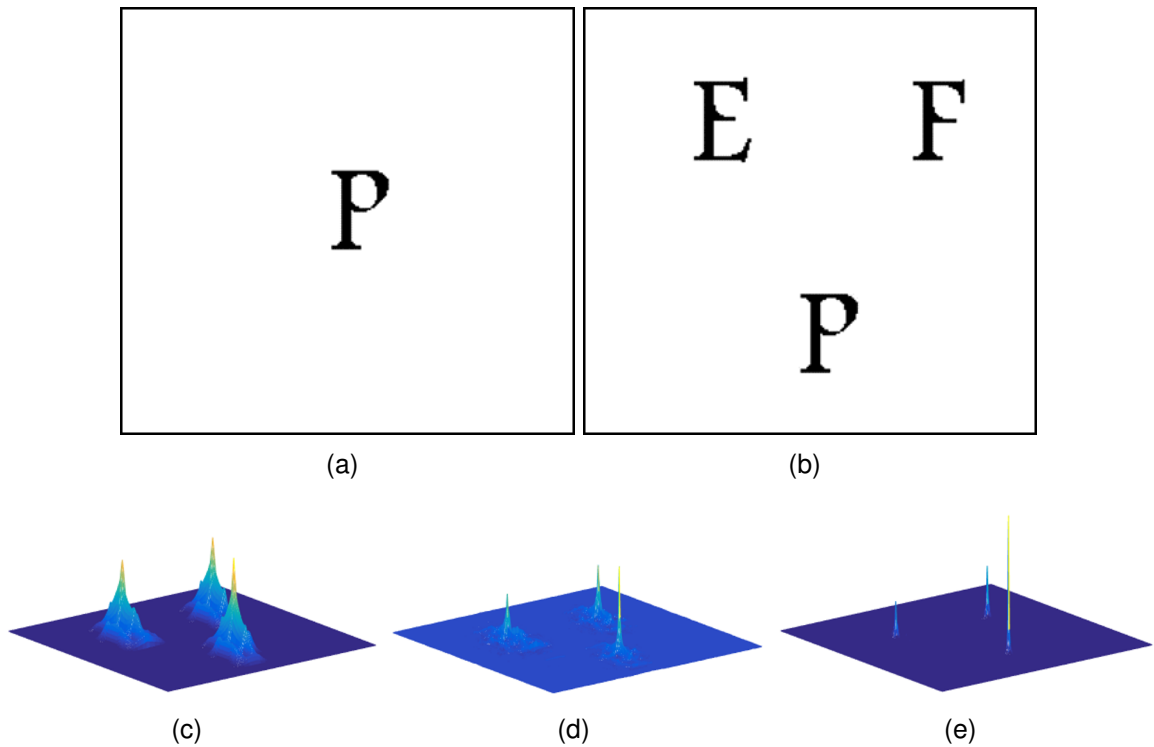
puesto recientemente trabajos que implementan un banco de filtros como una propuesta de solución al problema. Cada uno de estos filtros es diseñado con un subconjunto de imágenes de entrenamiento del objetivo. En este caso, la detección se lleva a cabo correlacionando la escena de entrada con cada uno de los filtros en el banco. Posteriormente, se realiza el cálculo de la DC para cada uno de estos filtros y se selecciona la salida con mejor desempeño.

### **3.6. Algoritmo básico para el reconocimiento de patrones por medio de filtros de correlación**

Este capítulo tiene como fin la descripción del método utilizado, así como los resultados calculados tras la implementación del mismo, sin embargo, antes de entrar en tema explicaremos brevemente el proceso básico para el reconocimiento de objetos por medio de los filtros de correlación descritos en el capítulo 3.

El proceso para determinar si el objeto de interés se encuentra presente dentro de una escena dada sigue los siguientes pasos:

1. Calcular la TF de la escena de entrada (imagen problema)  $s(x, y)$ , para obtener  $S(u, v) = F[s(x, y)]$ .
2. Calcular la función filtro  $H(u, v)$  conforme a los algoritmos descritos en las secciones



**Figura 3: Ejemplo de implementación de filtros de correlación. Imágenes en el dominio espacial: (a) Imagen filtro y (b) Imagen problema. Respuesta calculadas para filtros clásicos. (c) MF, (d) POF y (e) IF.**

3.2, 3.3 y 3.4.

3. Efectuar la operación punto a punto del producto  $S(u, v)$  con la función filtro  $H(u, v)$ .
4. Calcular el plano de correlación de salida aplicando la IFT al resultado del paso anterior, esto es  $c(x, y) = F^{-1}[S(u, v)H^*(u, v)]$ .
5. Localizar los picos de correlación con más intensidad en el plano de salida  $c(x, y)$ , si este valor es mayor a un umbral deseado, entonces se dice que el filtro tuvo éxito en el reconocimiento.

En la figura 3 presentamos un escenario de detección como el descrito hasta ahora. El objetivo del reconocimiento aquí es detectar al objeto de interés (letra P) dentro de una imagen problema propuesta, en la cual se añadieron otros elementos que comparten características parecidas al objeto de interés. En los incisos (c)-(e) se muestran las respuestas obtenidas para cada uno de los filtros probados.

## Capítulo 4. Sistema de seguimiento propuesto

---

En este capítulo realizaremos una descripción del sistema de seguimiento propuesto. El sistema se descompone en tres etapas: La etapa de preprocesamiento, donde creamos los filtros de correlación; etapa de detección, donde se realizan las correlaciones de los filtros creados con las escenas de la secuencia de video; y la etapa de predicción, en la cual implementamos el modelo de Kalman descrito en el capítulo 2.

De la misma manera presentamos los elementos utilizados para realizar pruebas a nuestro sistema sin ruido y en presencia de ruido aditivo a diferentes valores de desviación estandar  $\sigma$ . El sistema se evalúa en términos de su tasa de reconocimiento, errores de localización y porcentaje de falsos positivos y negativos.

### 4.1. Preprocesamiento

El filtrado adaptativo utilizado durante el desarrollo de este trabajo de tesis emplea un conjunto de filtros de correlación entrenados. Los filtros diseñados deben tener la capacidad de reconocer diferentes versiones del objeto cuando este se traslada a través del fondo.

El método comienza por sintetizar un banco de filtros de correlación. Primero, seleccionamos una imagen  $t_i$  que represente al objeto de interés. Esta imagen es escalada y rotada produciendo un conjunto de versiones del objeto modificadas geométricamente. Este conjunto de imágenes son utilizadas para crear filtros compuestos. Para el diseño de cada filtro compuesto utilizamos el método de funciones discriminantes sintéticas adaptativas descritas en la sección 3.5.3. Esta técnica nos capacita para generar un valor deseado en el área del objeto en el plano de correlación de salida con una reducción de lóbulos laterales en el área del fondo. Por lo tanto, para un buen funcionamiento e inicialización del sistema debemos tener información *a priori* del objeto a identificar y del fondo (o escena) por donde se moverá este.

En la figura 4 presentamos un ejemplo de la respuesta obtenida al implementar filtros creados por SDF y filtros creados por ASDF, en la figura 4 (a) se presenta el objeto, el cual ha sido afectado por diferentes degradaciones geométricas y colocado en un fondo típico de la figura 4 (b) dada la naturaleza del objeto. Se crearon 6 filtros SDF y 6 filtros ASDF a

partir de la información obtenida para cada versión creada del objeto y se presentan los planos de correlación correspondientes para cada filtro.

Es fácil ubicar la posición del objeto de interés en la imagen problema a la salida de los filtros ASDF debido a que el proceso iterativo de la figura 2 nos permite seleccionar un valor DC deseado (típicamente en el rango de  $[0.6 - 0.9]$ ) en el desempeño de cada filtro compuesto, de esta manera localizamos picos definidos a la salida, en contraste con las salidas de los filtros SDF, donde nos es imposible ubicar al objeto dada la cantidad de ruido generado por patrones con características parecidas al objeto en el fondo.

#### **4.2. Detección**

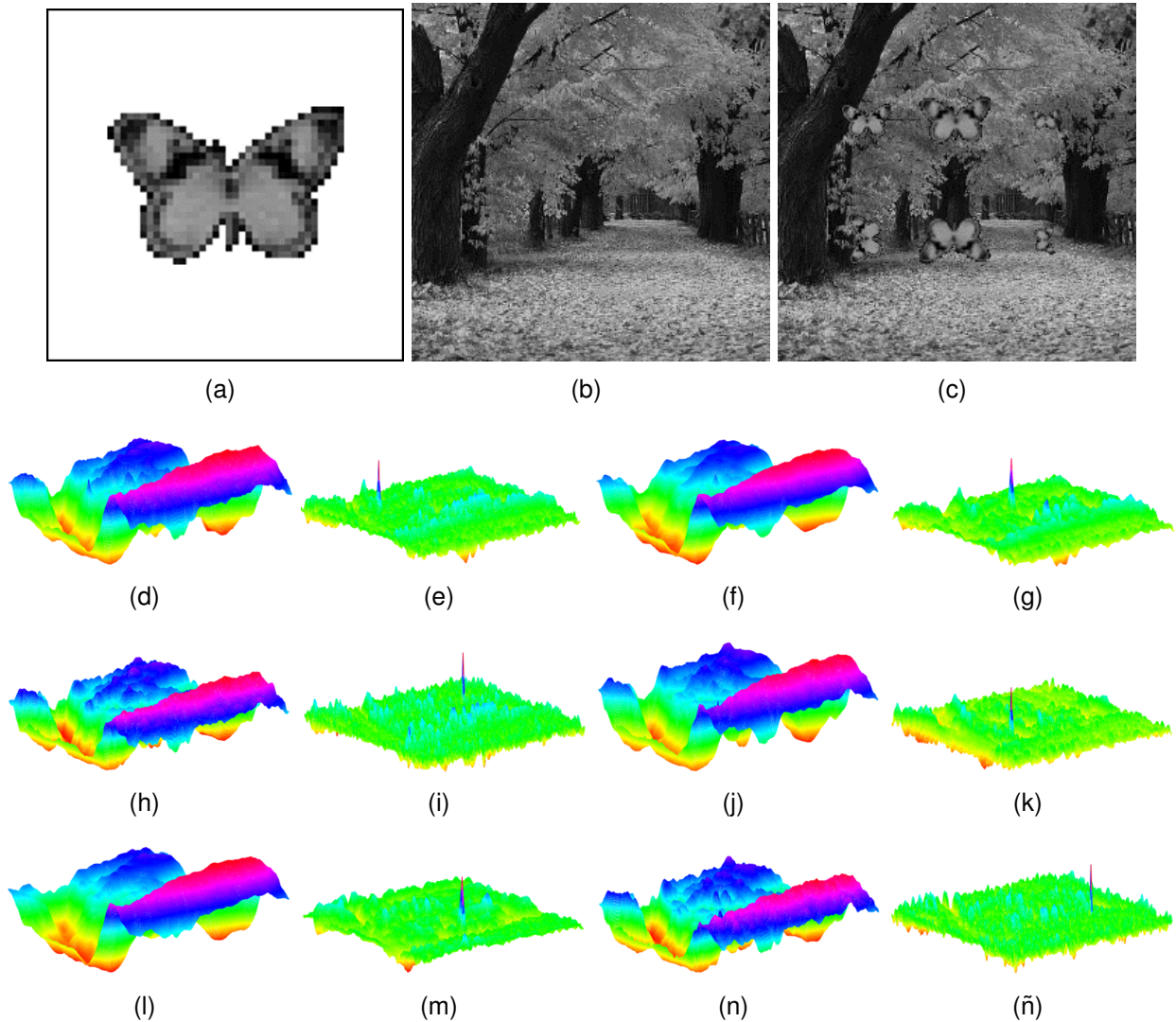
El banco de filtros es construido y entrenado antes de iniciar el proceso de seguimiento. La etapa de detección comienza cuando leemos la primer escena de la video secuencia. Para este tiempo, no sabemos si el objeto se encuentra presente en la escena o no, entonces, correlacionamos la escena con todos los filtros compuestos que conforman el banco, luego calculamos el valor DC de cada filtro y seleccionamos el que haya tenido mejor desempeño, si este valor es igual o mayor que un valor DC umbral especificado, entonces se dice que el objeto se encuentra en la escena. El valor DC umbral debe ser seleccionado de modo que garantice el reconocimiento del objeto con un número reducido de detecciones perdidas y falsas. Si el algoritmo determina la presencia del objeto, entonces su posición  $(x, y)$  en la escena y su orientación  $\theta$  es obtenida a partir del plano de correlación con mayor DC. En la figura 5 se presenta un diagrama básico de la metodología descrita.

Después de leer el primer fotograma de la secuencia que contiene al objeto, implementamos las estimaciones del modelo de Kalman para reducir el área de interés en el siguiente cuadro, el número de filtros correlacionados y por lo tanto el tiempo computacional, a partir de aquí el proceso de detección continúa en cada escena hasta leer la última escena de la secuencia.

#### **4.3. Predicción**

Después de que la primer detección es obtenida, es posible implementar la etapa de predicción con el fin de acelerar el proceso de seguimiento. En este trabajo, utilizamos





**Figura 4: Ejemplo de respuesta de un filtro SDF en comparación con la salida de un filtro ASDF. Imágenes en el dominio espacial: (a) Imagen filtro  $t_i$ , (b) Fondo utilizado para entrenamiento, (c) Imagen problema. Respuesta calculadas de los filtros SDF (d,f,h,j,l,n) y ASDF (e,g,i,k,m,ñ). (d) SDF objeto normal, (e) ASDF objeto normal, (f) SDF objeto escala 0.6, (g) ASDF objeto escala 1.4, (h) SDF objeto escala 0.6, (i) ASDF objeto escala 0.6, (j) SDF objeto rotación 90°, (k) ASDF objeto rotación 90°, (l) SDF objeto rotación 180° escala 1.4, (m) ASDF objeto rotación 180° escala 1.4, (n) SDF objeto rotación 270° escala 0.6 y (ñ) ASDF objeto rotación 270° escala 0.6.**

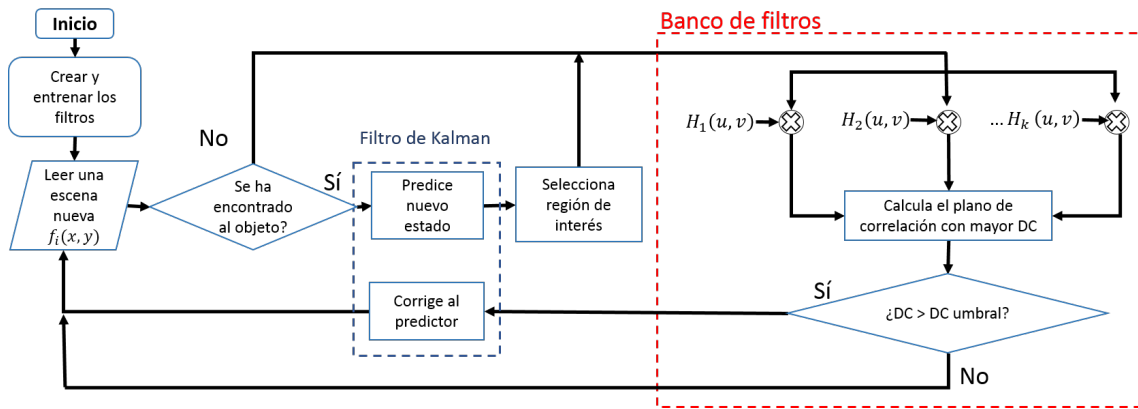


Figura 5: Diagrama de flujo de las operaciones básicas del sistema de seguimiento propuesto.

la implementación del filtro de Kalman incluida en el *toolbox* de sistemas de visión por computadora de matlab para llevar a cabo esta etapa (Mathworks, 1995-2015). La clase *Vision.KalmanFilter* incluida en el *toolbox* nos da la capacidad de crear un objeto de Kalman para generar las predicciones sin la necesidad de formar todas las ecuaciones de estado del modelo. Los parámetros del objeto pueden configurarse con la función *ConfigureKalmanFilter*, la cual está incluida en Matlab también. La función *ConfigureKalmanFilter* regresa un objeto del filtro de Kalman, pero es necesario que se ajusten ciertos parámetros primero, los cuales se describen en la tabla 2.

Entonces, necesitamos especificar cinco parámetros, entre los cuales es necesario calcular el primer vector de estado para predecir el comportamiento del objetivo en el siguiente fotograma. Como ya se mencionó anteriormente el algoritmo de Kalman implementa un estimador predictor-corrector. Estos requerimientos están incluidos en la clase de Kalman de matlab. El método de predicción utiliza los estados anteriores para estimar el estado actual. Por su parte, el método de corrección utiliza la medición actual para corregir el estado. Consecuentemente, cuando el objeto de Kalman ha sido inicializado, utilizamos el método de predicción para predecir el nuevo vector de estado  $[x_{t+1}, y_{t+1}, \theta_{t+1}]$ . Después, con la posición predicha  $[x_{t+1}, y_{t+1}]$  es posible elegir una región de interés en el siguiente fotograma en la zona donde el algoritmo ha estimado la nueva posición. del mismo modo utilizamos la orientación predicha  $\theta_{t+1}$  para seleccionar un conjunto de filtros del banco para reducir el número de correlaciones y por lo tanto el tiempo de procesamiento por cuadro.

En la figura 6, podemos ver el algoritmo de seguimiento propuesto. Donde *Hbf* repre-

Tabla 2: Argumentos de entrada de la función ConfigureKalmanFilter

Argumento	Descripción
<b>Modelo de movimiento</b>	El modelo de movimiento representa el comportamiento físico del movimiento del objeto. Es posible seleccionar entre <i>velocidad constante</i> o <i>aceleración constante</i> .
<b>Estado inicial</b>	Típicamente este valor del estado inicial corresponde con el primer estado obtenido tras la primer detección del objeto.
<b>Error inicial estimado</b>	Este parámetro representa el error incluido en la primer detección, es recomendable elegir valores altos dentro de este vector debido a que el primer estado puede ser ruidoso ya que se basa solamente en una detección.
<b>Ruido del modelo</b>	El ruido del modelo ayuda a especificar la tolerancia que tiene el objeto de Kalman a la desviación que existe entre los valores calculados por el modelo y las mediciones. Incrementar los valores de este vector puede ocasionar que el modelo de Kalman cambie su estado para ser más fiel a las detecciones.
<b>Ruido de la medición</b>	Los valores de este parámetro deben de seleccionarse de acuerdo a la eficiencia del detector. Es recomendable seleccionar valores grandes cuando el detector comete gran cantidad de errores de localización.

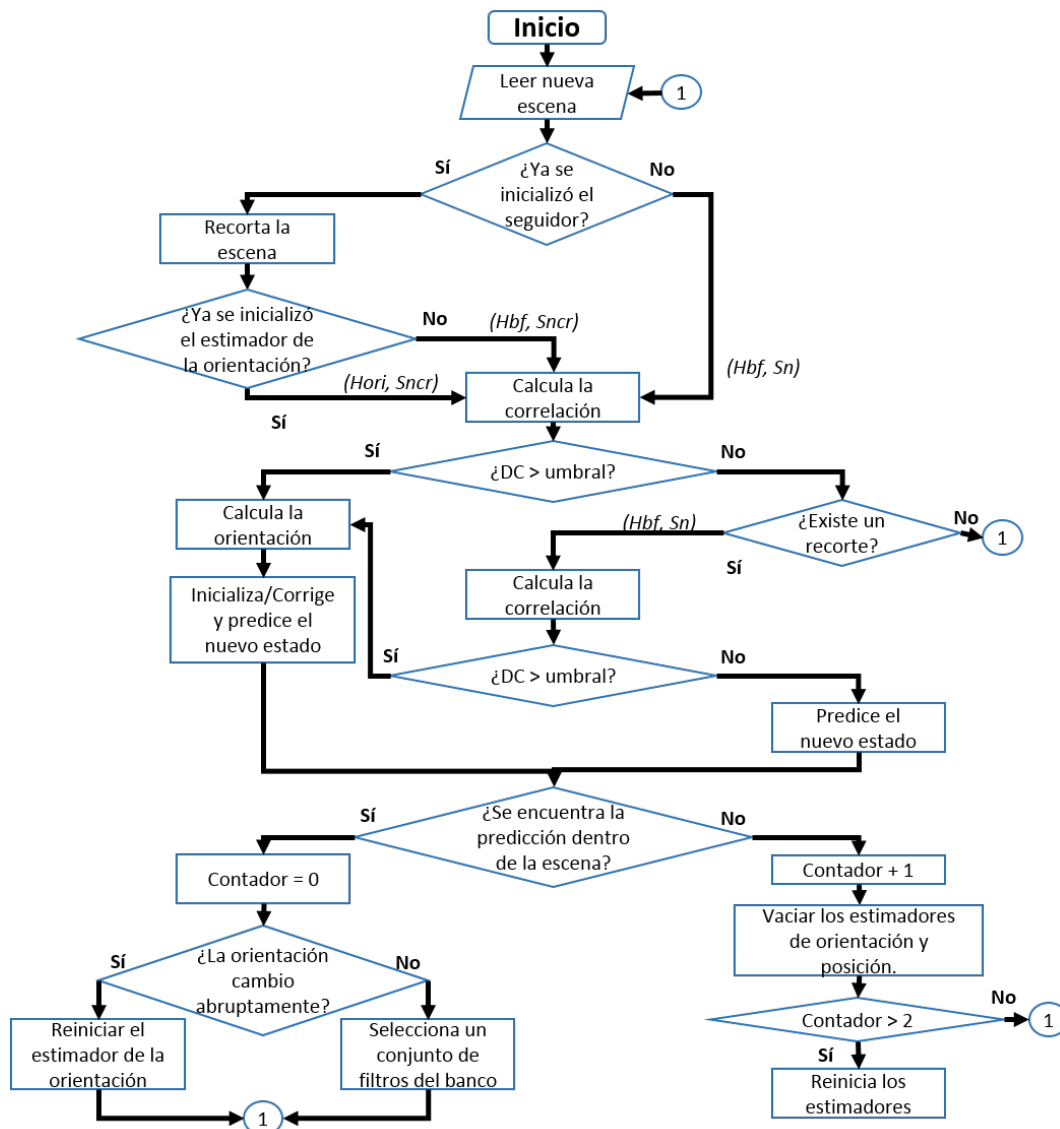


Figura 6: Diagrama de flujo del sistema de seguimiento propuesto.

senta a todos los filtros sintetizados,  $H_{ori}$  es el conjunto de filtros seleccionados dependiendo de la orientación predicha,  $S_n$  es el fotograma completo, y  $S_{ncr}$  es un fragmento de la escena  $S_n$ . Después de la primera estimación del objeto de Kalman, alguna de las siguientes situaciones puede ocurrir:

- El objeto se encuentra dentro de la región de interés en la siguiente escena. Este caso presenta el mejor de los escenarios. Cuando después de calcular la correlación de los filtros seleccionados ( $H_{ori}$ ) con el fragmento de la escena, obtenemos un valor DC mayor a cierto umbral, entonces implementamos el método de corrección para ajustar el modelo con la posición obtenida y posteriormente el método de predicción para continuar con el proceso. Aquí podemos tener dos casos: (i) La nueva posición predicha se encuentra dentro de los límites de la escena por lo cual se selecciona una nueva región de interés para el siguiente cuadro; (ii) La nueva posición predicha se encuentra fuera de los límites de la escena. En este último caso no es posible crear una nueva región de interés, por lo cual los filtros seleccionados se correlacionan con la escena completa para verificar que el objeto haya salido, en caso de lo contrario implementamos el método de corrección con la medición obtenida.
- El objeto se encuentra dentro de la región de interés en la siguiente escena pero con un cambio de orientación significativo. Debido a que el tiempo transcurrido entre cuadro y cuadro de una secuencia de video es muy corto, los objetos experimentan cambios de orientación muy pequeños, sin embargo, hay casos cuando este cambio de orientación se da entre valores positivos y valores negativos ( Por ejemplo, cuando el objeto cambia su orientación de  $2^\circ$  en el tiempo  $t$  a  $-2^\circ$  en el tiempo  $t + 1$ , entonces, el objeto de Kalman recibe una lectura de un cambio en la orientación de  $356^\circ$  cuando realmente el objeto giró solamente  $4^\circ$ , consecuentemente, el método de corrección introduce una actualización incorrecta por lo que el estimador puede perder la trayectoria del objeto en el siguiente cuadro). Lo mismo sucede en sentido contrario cuando el objeto cambia su orientación de valores grandes a valores pequeños. Para evitar este problema, proponemos reiniciar el estimador de la orientación cada vez que la orientación cruce el umbral de los positivos y los negativos, y comenzar la estimación a partir de la siguiente lectura. Debido a este caso, es recomendable crear dos objetos de la clase Kalman, uno para estimar la posición y

otro para estimar la orientación.

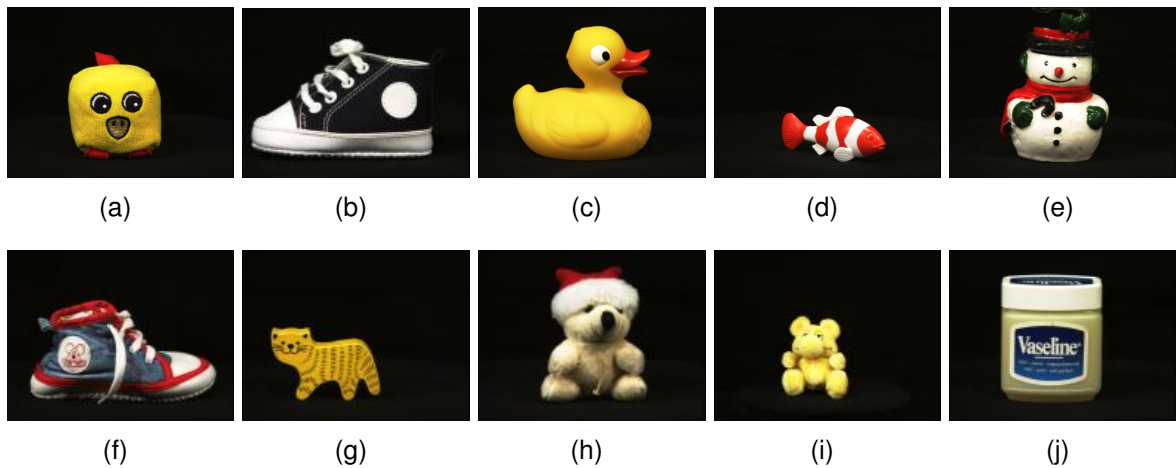
- El objeto no se encuentra dentro de la región de interés. Este problemático caso puede presentarse cuando el objeto se encuentra oculto por otros objetos o cuando este cambia su dirección o velocidad abruptamente. Si se calcula un valor DC menor a cierto umbral después de correlacionar los filtros seleccionados con la región de interés, entonces se prosigue a correlacionar los filtros del banco con la escena entera. De aquí podemos tener: (i) el objeto es encontrado en la escena, entonces, el método de corrección actualiza al modelo y el proceso continúa normalmente con una nueva predicción; (ii) el objeto no se encuentra en ningún lugar de la escena, por lo cual se asume que el objeto continúa oculto, así que no hay medición de corrección, en este caso utilizaríamos solamente el método de predicción para continuar analizando el siguiente cuadro.
- Existen varias predicciones continuas fuera de los límites de la escena. Si el modelo creado acumula más de dos estimaciones donde indique que el objeto se encuentra fuera de la región de la escena debido a que no ha sido posible ubicar al objeto mediante la correlación con el banco de filtros, entonces decimos que el objeto ha desaparecido de la escena y los estimadores de estado son reiniciados.

El proceso de detección-predicción continúa de la manera descrita hasta leer el último cuadro de la secuencia de video.

#### **4.4. Resultados**

En esta sección hablaremos sobre los resultados obtenidos tras las simulaciones realizadas. El desempeño del sistema propuesto es evaluado tomando en cuenta diferentes trayectorias creadas, la presencia de ruido aditivo en la escena y simulaciones en las cuales se retiró al objeto de la escena a diferentes alturas de cada trayectoria simulando que se mantuvo oculto.

A continuación se resumen las condiciones sobre las cuales se realizaron los experimentos.



**Figura 7: Objetos de prueba.** (a)Muñeco amarillo, (b)zapato negro, (c)pato sonriente, (d)pez rojo y blanco, (e)mono de nieve, (f)zapato, (g)gato amarillo, (h)oso de navidad blanco, (i)muñeco amarillo 3 y (j)vaselina.

#### 4.4.1. Elementos de prueba

Los experimentos se llevan a cabo utilizando 10 objetos diferentes seleccionados de la librería de objetos de Amsterdam (Geusebroek *et al.*, 2005). Estos objetos se puede apreciar en la figura 7. Por cada objeto, necesitamos un fondo, por lo cual seleccionamos 10 diferentes de internet (figura 8), los cuales debían presentar un contexto natural para la localización de cada objeto.

Todos los cuadros de las secuencias de video tienen un tamaño estandar de  $256 \times 256$  pixeles, mientras que cada objeto fue escalado de tal manera que pudiera contenerse dentro de un espacio de  $54 \times 54$  píxeles.

Creamos también cinco secuencias sintéticas de 30 cuadros con diferentes trayectorias como se muestra en la figura 9 (a)-(e). De la misma manera creamos secuencias tomando como base las anteriores, con la diferencia de que el objeto se desaparece en 3 fotogramas seguidos a diferentes alturas de tres trayectorias representativas como se muestra en la figura 9 (f)-(h), esto con el fin de simular que el objeto se mantiene oculto a diferentes alturas del proceso y poder obtener resultados más precisos.

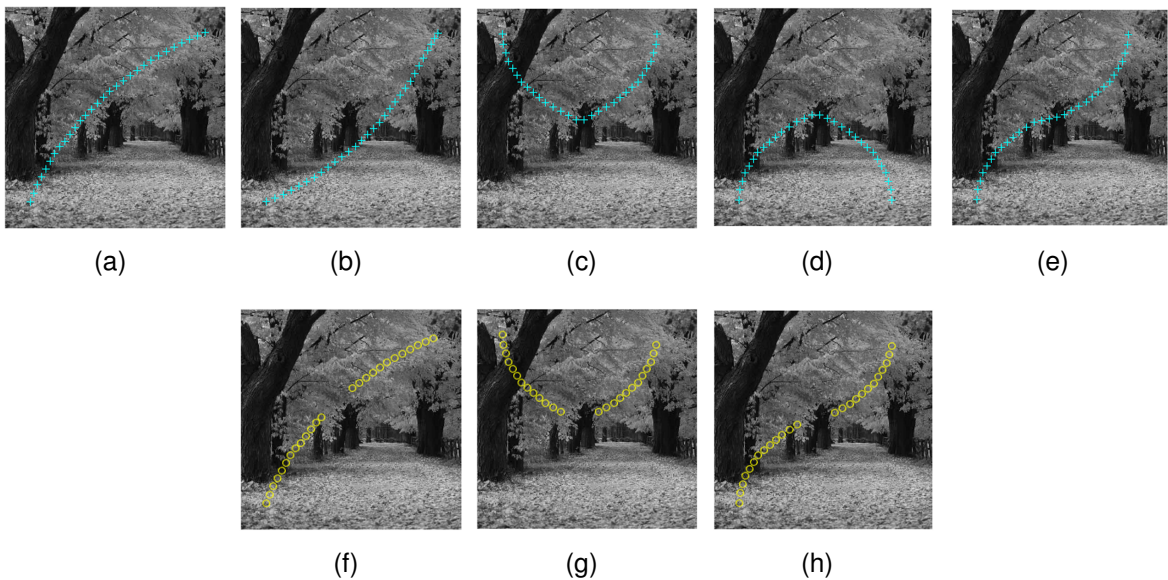
#### 4.4.2. Parámetros utilizados

Los objetos de la figura 7 son afectados por diferentes degradaciones como rotaciones dentro del plano en el intervalo  $[1^\circ - 360^\circ]$  y pequeñas variaciones de escalas alternadas





**Figura 8: Fondos seleccionados de internet para pruebas. (a) Juguetes, (b) zapatos, (c) juguetes 2, (d) océano, (e) adornos de navidad, (f) zapatos 2, (g) juguetes 2, (h) juguetes 3, (i) juguetes 4 y (j) productos.**



**Figura 9: Trayectorias creadas para pruebas. (a-e) Trayectorias (1-5), (f) objeto oculto primera posición, (g) objeto oculto segunda posición, (h) objeto oculto tercera posición**

en el intervalo  $[0.9 - 1.1]$ .

Los parámetros utilizados para la construcción de los objetos de Kalman son los siguientes: *Aceleración constante* como modelo de movimiento con lo cual se generan los arreglos  $A$  y  $H$  de las ecuaciones 34 y 35 respectivamente, el estado inicial corresponde a los valores calculados después de la primer detección (creamos un objeto que rastrea la posición y otro que rastrea la orientación según lo discutido en el tercer caso de la sección 4.3), el vector de error inicial estimado se llena de la siguiente manera  $[1, 1, 1] \times 10$ , el ruido del modelo como  $[15, 5, 5]$  y el ruido de la medición igual a 10. Estos valores pequeños son seleccionados con el fin de matener al estimador más fiel a las mediciones que al modelo de movimiento seleccionado.

La región de interés seleccionada para cada fotograma tiene un tamaño de  $1.5t_i$  y cuando el objeto no es encontrado dentro del fragmento seleccionado se prosigue a buscarlo en la escena completa.

Un valor DC umbral de 0.6 fue seleccionado según Díaz-Ramírez *et al.* (2014) para garantizar el reconocimiento de objetos en ambientes ruidosos.

El banco de filtros está compuesto por filtros construidos mediante ASDF, los cuales se generaron a partir de versiones del objeto rotadas un grado a la vez, y en escalas dentro del intervalo  $[0.9 - 1.1]$ , es decir contamos con un total de 1080 filtros en  $Hbf$ , mientras que los filtros seleccionados a partir de la predicción de la orientación se reducen a 11 por escala, dándonos un total de 33 filtros en  $Hori$ .

Durante el desarrollo de este trabajo de tesis, todos los programas y resultados fueron calculados y obtenidos con una computadora portátil Vaio Ultrabook serie T, modelo SVT151A11L, con un procesador Intel(R) Core(TM) i5-3337U 1.80GHz, con 8GB de memoria RAM y sistema operativo Windows 8.1 a 64 bits.

Los resultados obtenidos se muestran a continuación.

#### **4.4.3. Resultados experimentales**

En esta parte realizamos la recopilación de los resultados obtenidos al poner a prueba nuestro sistema de seguimiento. Cada experimento se realizó un total de 30 veces para



**Tabla 3: Taza de reconocimiento y errores de clasificación del banco de filtros con secuencias de vídeo sin ruido.**

Sin ruido						
Obj	Fp #	Fp %	Fn #	Fn %	Reconocimiento #	Reconocimiento %
a	0	0 %	0	0 %	150	100 %
b	1	1 %	0	0 %	149	99 %
c	0	0 %	0	0 %	150	100 %
d	0	0 %	0	0 %	150	100 %
e	0	0 %	0	0 %	150	100 %
f	0	0 %	0	0 %	150	100 %
g	1	1 %	0	0 %	149	99 %
h	0	0 %	0	0 %	150	100 %
i	0	0 %	0	0 %	150	100 %
j	0	0 %	0	0 %	150	100 %

**Tabla 4: Taza de reconocimiento y errores de clasificación del banco de filtros con secuencias de vídeo en presencia de ruido aditivo con desviación estandar igual a 10.**

Ruido con $\sigma = 10$						
Obj	Fp #	Fp %	Fn #	Fn %	Reconocimiento #	Reconocimiento %
a	0	0 %	0	0 %	150	100 %
b	1	1 %	0	0 %	149	99 %
c	1	1 %	0	0 %	149	99 %
d	0	0 %	0	0 %	150	100 %
e	0	0 %	0	0 %	150	100 %
f	0	0 %	0	0 %	150	100 %
g	1	1 %	0	0 %	149	99 %
h	0	0 %	0	0 %	150	100 %
i	0	0 %	0	0 %	150	100 %
j	0	0 %	0	0 %	150	100 %

ofrecer un resultado con validez estadística.

#### 4.4.3.1. Evaluación del banco de filtros

En primer lugar presentamos los resultados de la evaluación al banco de filtros creado en la tabla 3, y en presencia de ruido con desviación estandar  $\sigma = 10, 20, 30$  y  $40$  las tablas 4,5,6 y 7, respectivamente. La tasa de reconocimiento del algoritmo corresponde al porcentaje de detecciones exitosas en relación con el número de escena analizadas (150 escenas por objeto para este caso). De aquí, nos interesan también los errores de clasificación, decimos que un falso positivo se presenta cuando el sistema dice reconocer un objeto cuando este no se encuentra presente, mientras que un falso negativo se obtiene cuando el sistema dice no reconocer al objeto cuando este se encuentra presente.

**Tabla 5: Taza de reconocimiento y errores de clasificación del banco de filtros con secuencias de vídeo en presencia de ruido aditivo con desviación estandar igual a 20.**

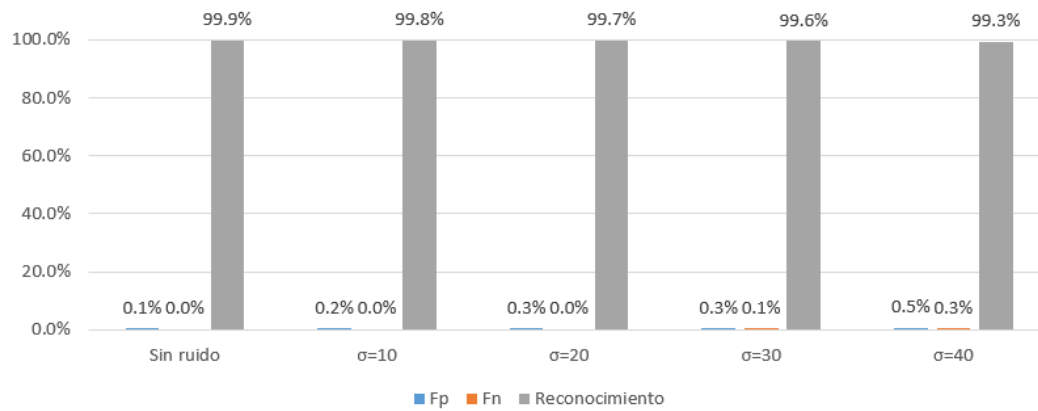
Ruido con $\sigma = 20$						
Obj	Fp #	Fp %	Fn #	Fn %	Reconocimiento #	Reconocimiento %
a	1	1 %	0	0 %	149	99 %
b	1	1 %	0	0 %	149	99 %
c	1	1 %	0	0 %	149	99 %
d	0	0 %	0	0 %	150	100 %
e	0	0 %	0	0 %	150	100 %
f	0	0 %	0	0 %	150	100 %
g	2	1 %	0	0 %	148	99 %
h	0	0 %	0	0 %	150	100 %
i	0	0 %	0	0 %	150	100 %
j	0	0 %	0	0 %	150	100 %

**Tabla 6: Taza de reconocimiento y errores de clasificación del banco de filtros con secuencias de vídeo en presencia de ruido aditivo con desviación estandar igual a 30.**

Ruido con $\sigma = 30$						
Obj	Fp #	Fp %	Fn #	Fn %	Reconocimiento #	Reconocimiento %
a	1	1 %	0	0 %	149	99 %
b	1	1 %	0	0 %	149	99 %
c	1	1 %	0	0 %	149	99 %
d	0	0 %	1	1 %	149	99 %
e	0	0 %	0	0 %	150	100 %
f	0	0 %	0	0 %	150	100 %
g	2	1 %	0	0 %	148	99 %
h	0	0 %	0	0 %	150	100 %
i	0	0 %	0	0 %	150	100 %
j	0	0 %	0	0 %	150	100 %

**Tabla 7: Taza de reconocimiento y errores de clasificación del banco de filtros con secuencias de vídeo en presencia de ruido aditivo con desviación estandar igual a 40.**

Ruido con $\sigma = 40$						
obj	Fp #	Fp %	Fn #	Fn %	Reconocimiento #	Reconocimiento %
a	1	1 %	0	0 %	149	99 %
b	1	1 %	0	0 %	149	99 %
c	1	1 %	0	0 %	149	99 %
d	0	0 %	4	3 %	146	97 %
e	0	0 %	0	0 %	150	100 %
f	0	0 %	0	0 %	150	100 %
g	4	3 %	0	0 %	146	97 %
h	0	0 %	0	0 %	150	100 %
i	0	0 %	0	0 %	150	100 %
j	0	0 %	0	0 %	150	100 %



**Figura 10: Datos resumidos: tasa de reconocimiento y errores de clasificación.**

Estos resultados se resumen en la gráfica de la figura 10. En ella podemos ver el promedio de los datos de cada tabla, es decir, son el resultado del promedio de los datos adquiridos para todos los objetos a diferentes intensidades de ruido.

#### 4.4.3.2. Comparación del modelo de Kalman con el modelo cinemático

Con la necesidad de ubicar nuestro sistema en contexto dentro del estado del arte, realizamos la comparación del método propuesto con el propuesto por Díaz-Ramírez *et al.* (2014), el cual implementa también una etapa de detección mediante filtros compuestos por ASDF, por lo tanto, es posible adecuar el modelo cinemático de predicción utilizado para que opere dentro con las mismas condiciones con las cuales evaluamos el nuestro.

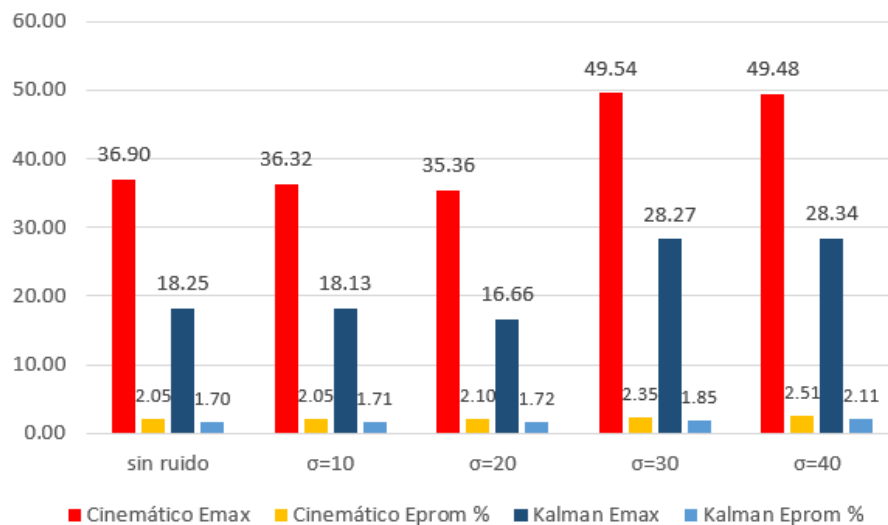
El modelo cinemático de predicción utilizado también en Santiago-Ramírez *et al.* (2012); Contreras *et al.* (2013); Miramontes-Jaramillo y Kober (2013) es comparado con el modelo de Kalman en términos de su error de localización máximo y error de localización promedio por trayectoria. Este modelo se caracteriza por utilizar las dos últimas lecturas de estado del objeto de interés para dar como salida una nueva predicción. Es por esto que este modelo presenta problemas cuando el objeto se encuentra parcialmente oculto, dado que al no tener una lectura no podría realizar la predicción. Los resultados obtenidos para cada objeto se muestran en las tablas 8 y 9, pero para su mayor apreciación, resumimos los datos en la gráfica de la figura 11.

**Tabla 8: Errores de localización máximo y promedio para cada objeto con modelo de predicción cinemático.**

C Obj	Sin ruido		de = 10		de=20		de=30		de=40	
	E <sub>max</sub>	E <sub>prom</sub>	E <sub>max</sub>	E <sub>prom</sub>	E <sub>max</sub>	E <sub>prom</sub>	E <sub>max</sub>	E <sub>prom</sub>	E <sub>max</sub>	E <sub>prom</sub>
a	7.00	1.75	7.62	1.81	6.08	1.86	7.07	1.84	7.07	1.81
b	109.29	2.85	107.23	2.79	112.16	2.83	109.17	2.85	108.17	2.82
c	90.05	2.34	87.82	2.36	90.92	2.48	91.42	2.64	90.92	2.57
d	8.00	1.83	6.00	1.86	6.40	1.99	129.00	4.09	129.50	3.44
e	6.40	1.74	6.08	1.82	6.71	1.90	8.00	1.88	8.60	1.78
f	6.00	1.84	6.00	1.71	6.70	1.78	6.40	1.80	7.07	1.85
g	123.17	3.05	122.79	3.03	104.75	2.94	125.26	3.02	123.56	5.38
h	6.71	1.87	7.21	1.71	7.28	1.83	6.32	1.85	7.21	1.91
i	6.32	1.68	6.32	1.62	6.32	1.78	6.08	1.77	6.32	1.76
j	6.08	1.57	6.08	1.74	6.32	1.66	6.71	1.72	6.40	1.75

**Tabla 9: Errores de localización máximo y promedio para cada objeto con modelo de predicción de Kalman.**

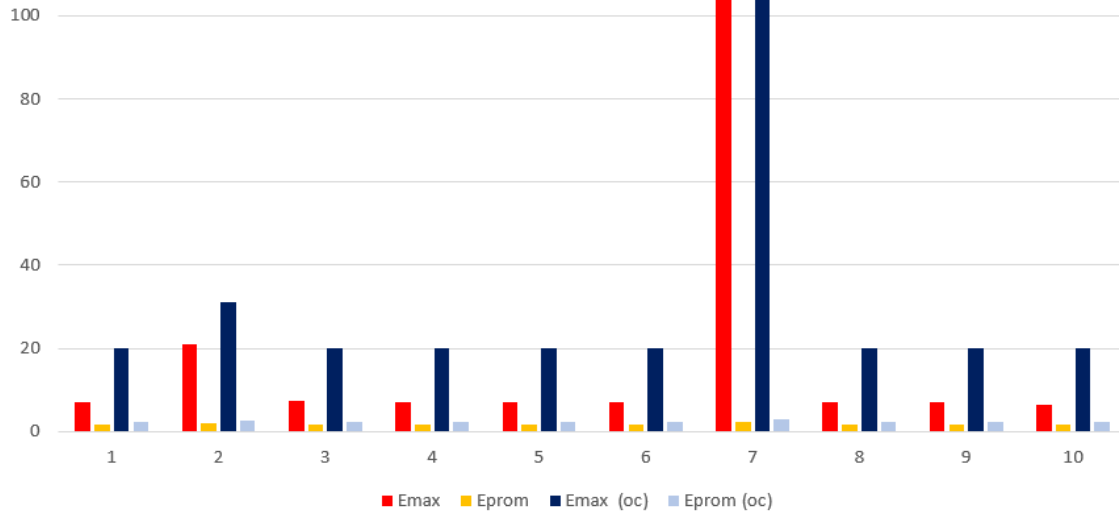
K Obj	Sin ruido		de = 10		de=20		de=30		de=40	
	E <sub>max</sub>	E <sub>prom</sub>	E <sub>max</sub>	E <sub>prom</sub>	E <sub>max</sub>	E <sub>prom</sub>	E <sub>max</sub>	E <sub>prom</sub>	E <sub>max</sub>	E <sub>prom</sub>
a	7.07	1.59	7.07	1.57	7.07	1.58	7.07	1.59	7.07	1.57
b	21.02	1.85	21.02	1.95	21.02	1.85	21.02	1.90	21.02	1.90
c	7.28	1.66	7.28	1.66	7.28	1.66	8.24	1.67	8.24	1.66
d	7.07	1.61	7.07	1.62	7.07	1.61	106.30	3.01	108.25	2.72
e	7.07	1.70	7.07	1.71	7.07	1.71	7.07	1.71	7.07	1.72
f	7.07	1.59	7.07	1.59	7.07	1.57	7.07	1.57	7.07	1.57
g	105.42	2.39	104.24	2.38	89.54	2.52	105.42	2.39	104.24	5.32
h	7.07	1.59	7.07	1.59	7.07	1.59	7.07	1.59	7.07	1.59
i	7.07	1.53	7.07	1.53	7.07	1.53	7.07	1.53	7.07	1.53
j	6.32	1.53	6.32	1.53	6.32	1.53	6.32	1.53	6.32	1.53



**Figura 11: Comparación de modelos de seguimiento: Errores de localización máximo y promedio.**

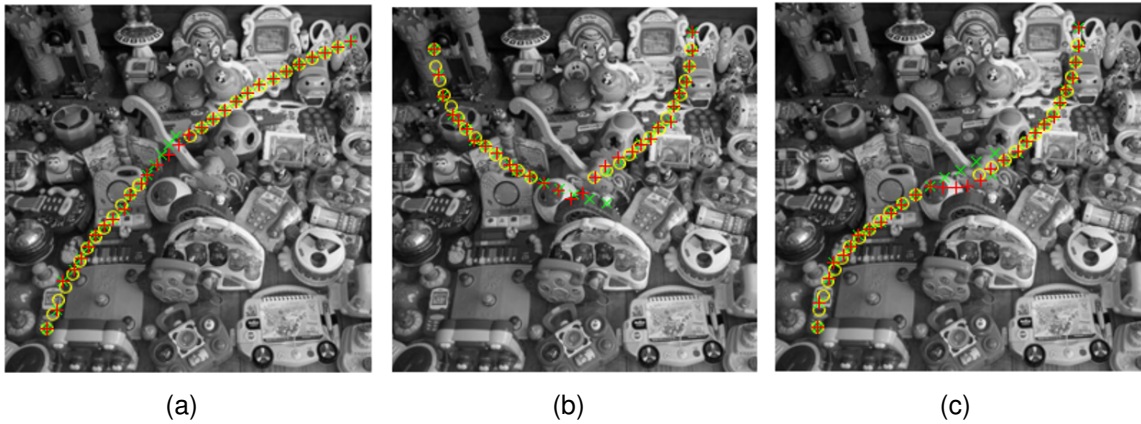
**Tabla 10: Errores de localización máximo y promedio cuando el objeto se ocultó de forma parcial.**

obj	1	2	3	4	5	6	7	8	9	10
E <sub>max</sub>	20.12	31.2	20.12	20.12	20.12	20.12	105.42	20.12	20.12	20.12
E <sub>prom</sub>	2.16	2.44	2.18	2.18	2.2	2.16	3.01	2.16	2.14	2.13

**Figura 12: Comparación de errores de localización del modelo con el objeto parcialmente oculto.**

#### 4.4.3.3. Errores de localización cuando el objeto se mantiene oculto

Cuando un objeto se desplaza normalmente por el medio, es común que en ocasiones se mantenga oculto o parcialmente oculto por otros objetos presentes en la escena, en estos casos el modelo de Kalman implementado nos brinda la capacidad de seguir estimando el estado del objeto sin tener una retroalimentación como medición. En el presente trabajo decidimos poner a prueba al sistema de seguimiento simulando que el objeto se mantiene oculto en tres posiciones diferentes de las trayectorias, como puede verse en la figura 9. Los resultados calculados para el error máximo y promedio de la localización en las trayectorias de cada objeto se enlistan en la tabla 10, mientras que en las figuras 12 y 13 podemos apreciar la comparación de los resultados cuando el objeto está presente siempre con respecto a cuando el objeto se retira de la escena, como puede observarse el modelo sufre una desviación mínima mientras el objeto se encuentra oculto, la cual se corrige rápidamente cuando el objeto reaparece en escena.



**Figura 13: Comparación gráfica de errores de estimación para objeto parcialmente oculto en trayectorias representativas. Posición real (amarillo), posición estimada con objeto presente (rojo) y posición estimada con objeto parcialmente oculto (verde).**

**Tabla 11: Error máximo y promedio (En grados) para la estimación de la orientación.**

obj	1	2	3	4	5	6	7	8	9	10
E <sub>max</sub>	167	166	56	147	118	56	57	55	85	179
E <sub>prom</sub>	14.4	12.2	10.8	11.3	11	10.1	10.4	10.8	10.9	11.9

#### 4.4.3.4. Seguidor de la orientación

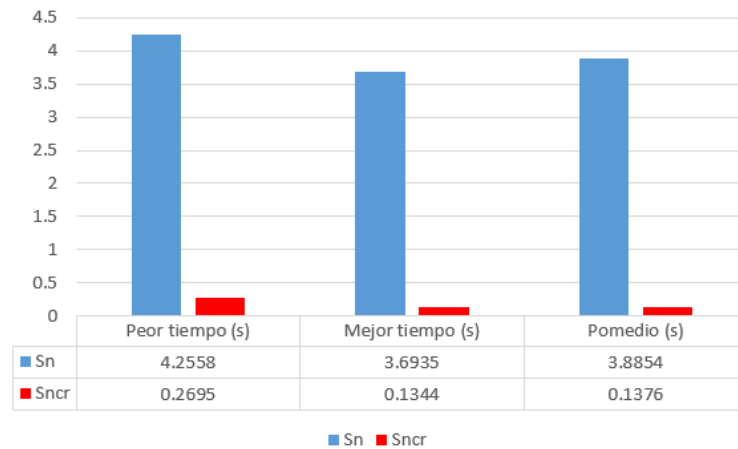
En la sección 4.3 se recomienda crear dos objetos de Kalman independientes, como una propuesta a la solución de cambios de umbral en los grados de la orientación de los objetos. Un objeto trabaja con la estimación de la posición y un objeto trabaja con la estimación de la orientación, de esta manera es posible reiniciar los objetos por separado cuando sea necesario.

En la tabla 11 presentamos los resultados (en grados) para la estimación de la orientación de cada objeto cuando este varía su rotación en el intervalo  $[1^\circ, 360^\circ]$ . El error promedio para la orientación de los objetos es de  $11^\circ$ .

#### 4.4.3.5. Tiempo de ejecución

Siempre que se emplea un algoritmo para resolver una tarea en específico, además de la precisión con la cual este solucione el problema, el tiempo que tarde en llegar a su objetivo es un factor importante a considerar.

En este caso, el tiempo de ejecución hace referencia al tiempo que el sistema tarda en procesar un cuadro de la secuencia de vídeo. Según sea el caso, el sistema puede



**Figura 14: Tiempo de ejecución del sistema. Donde Sn: Escena completa y Sncr: Región de interés.**

realizar el procesamiento de la escena completa, *Sn*, o de un fragmento de interés, *Sncr*, cuando exista una predicción del modelo de Kalman.

Dadas las posibilidades y limitantes que el dispositivo de cómputo empleado pudo ofrecer, los resultados obtenidos después de realizar el procesamiento de 30 escenas completas y 30 escenas fragmentadas se presentan en el gráfico de la figura 14, donde los valores enlistados están expresados en segundos.

#### 4.4.3.6. Discusión de los resultados

El sistema propuesto se evalúa de cinco maneras diferentes, esto con el fin de obtener un análisis más completo del alcance que una metodología como la propuesta puede alcanzar. En primero lugar, presentamos la evaluación al banco de filtros, como el proceso de predicción del sistema de seguimiento depende en gran cantidad de un sistema de reconocimiento fiable aún en presencia de ruido gaussiano en la escena, presentamos los resultados para 10 objetos presentes en escenas comunes a las cuales hemos añadido ruido. Podemos observar que el banco de filtros mantiene una tasa de reconocimiento promedio por arriba del 97% aún en presencia de ruido con  $\sigma = 40$ . En cuanto a los errores de clasificación, el banco de filtros obtiene un valor menor al 4% promedio para todos los niveles de ruido (ver figura 10). Entonces, es posible decir que para los 10 casos presentados el banco de filtros presenta resultados favorables en el reconocimiento sobre escenas con objetos con características similares utilizando un umbral  $DC = 0.6$ , lo cual nos brinda una base de detección sólida para resolver el problema de seguimiento.

En segundo lugar, probamos el modelo propuesto contra el modelo cinématico vigente entre técnicas empleadas en el estado del arte. Ambos modelos de predicción son comparados utilizando la misma metodología de detección por ASDF, las pruebas se realizaron para todos los objetos en secuencias de vídeo contaminadas por ruido, con el fin de evaluar qué tanto afecta el ruido incluido en la medición al siguiendo que da el modelo de predicción. Los resultados presentados en la gráfica de la figura 11 nos permiten observar cómo la presencia de ruido en las mediciones ocasiona que los estimadores aumenten en poca proporción el promedio de sus errores, siendo más notable el aumento de error de localización máximo para el modelo cinématico. En cuanto a los errores promedios de la estimación, el modelo de Kalman obtuvo mejores resultados en comparación al modelo cinématico para todos los casos, manteniéndose por debajo de un error con valor promedio de 3 pixeles por estimación en el caso con mayor cantidad de ruido en la escena.

En tercer lugar, decidimos ocultar al objeto durante tres posiciones consecutivas en tres segmentos de diferentes trayectorias, con el fin de evaluar qué tanto se afecta al desempeño del modelo cuando el objeto se mantiene oculto en diferentes momentos de su movimiento por la escena. En la figura 12 se presenta la comparación de los cálculos realizados después de ocultar al objeto con respecto a los resultados presentados cuando el objeto se mantiene presente. En todos los casos analizados el sistema consiguió realizar estimaciones del nuevo estado para el objeto, manteniéndose fiel al comportamiento de la trayectoria real. El cambio más notorio se presentó en el error máximo obtenido, sin embargo, la variación en los valores promedio es mínima en todos los casos (menor a un pixel del valor promedio), lo que demuestra que el sistema se mantiene siguiendo al objeto aún cuando este se mantiene oculto y logra ajustarse rápidamente cuando el objeto entra en escena nuevamente.

En cuarto lugar evaluamos las estimaciones de orientación para objetos que fueron variando su orientación con diferentes aceleraciones y cambios de dirección. Obtuvimos un error promedio menor a  $11^\circ$ , lo cual nos hace pensar que existen filtros con diferente orientación que realizan una buena discriminación del objeto en el fondo, por lo tanto, es posible depender de un conjunto de filtros menor en el banco creado.

Finalmente, se evalúa el sistema en base al tiempo que tarda en procesar una escena



nueva. En la figura 14 podemos notar que para el peor y el mejor de los casos, la ganancia en tiempo de ejecución es notablemente superior, sin embargo, estamos conscientes de que este valor puede variar al ajustar el tamaño del fragmento de interés y/o el número de filtros en *Hori*.

## Capítulo 5. Conclusiones

---

### 5.1. Conclusión

El seguimiento de objetos a través de una secuencia de vídeo se vuelve una labor muy compleja considerando los diversos factores que interfieren, los cuales van desde el ruido introducido en el proceso de formación de la imagen y debido a imperfecciones de los sensores, hasta la gran variedad de distorsiones que puede sufrir el objeto de interés durante su movimiento por la escena, considerando que si quisieramos incluir más versiones del objeto al banco de filtros, este llegaría a volverse más complejo y por lo tanto el proceso más lento. Aunado a esto, el sistema debe garantizar el reconocimiento del objeto aún cuando existan elementos con características similares en la escena, lo cual implica criterios de clasificación. Para simplificar el proceso, suponemos que siempre contamos con información de la apariencia del objeto antes de iniciar el proceso de búsqueda, sin embargo, en diversas aplicaciones prácticas esta suposición no se cumple.

En este trabajo combinamos las ventajas adaptativas de los filtros de correlación para una detección precisa y una robusta predicción basada en filtrado de Kalman. Con el fin de aumentar la eficiencia del procesamiento, el algoritmo propuesto selecciona fragmentos de interés de cuadros futuros, los cuales se determinan basándonos en las coordenadas predichas por el modelo de Kalman. El enfoque adaptativo es utilizado para brindar una capacidad de discriminación deseada a los filtros del banco, adaptándolos a diferentes fondos.

Los filtros desarrollados junto con el algoritmo de estimación propuesto se utilizan para rastrear 10 objetos en secuencias de vídeo creadas sintéticamente. Finalmente, se presentaron y discutieron los resultados obtenidos con ayuda de la simulación por computadora, donde se demuestra un nivel de confianza alto en términos de eficiencia en la detección y precisión en el seguimiento del objeto. El sistema demuestra que es posible seguir y estimar el estado de un objeto con gran precisión aún en la presencia de ruido moderado en las imágenes e incluso a través de fondos complejos. En comparación con el modelo cinématico se logró obtener resultados favorables en todos los casos de prueba considerados.

## 5.2. Trabajo a futuro

Aún cuando la mayoría de los objetivos planteados para este trabajo fueron alcanzados, queda aún mucho trabajo por hacer para generar un sistema de seguimiento más robusto.

En primer lugar, podría continuarse por extender la técnica propuesta para el seguimiento de objetos en imágenes a color o multiespectrales (como imágenes aéreas).

Proponiendo un ambiente más cercano a la realidad, podemos tratar con el seguimiento de objetos cuando el objeto, el fondo y el dispositivo de captura presentan movimientos independientes.

La reducción del número de filtros generados en un principio para el banco, de tal manera que no se ponga en riesgo el buen desempeño obtenido hasta el momento.

La implementación del sistema en GPU, con el fin de reducir el tiempo de ejecución al realizar correlaciones del banco de filtros en paralelo.

Así mismo, podría trabajarse con la idea de crear una adaptación local, es decir, crear filtros exclusivamente con base en la información del estado futuro para el objeto, de esta manera el entrenamiento de los filtros se actualizaría cada escena.

## Lista de referencias bibliográficas

- Ali, A., Jalil, A., Ahmed, J., Iftikhar, M. A., y Hussain, M. (2014). Correlation, Kalman filter and adaptive fast mean shift based heuristic approach for robust visual tracking. *Signal, Image and Video Processing*, pp. 1–19.
- Álvarez-Borrego, J., Solorza, S., y Bueno-Ibarra, M. A. (2013). Invariant correlation to position and rotation using a binary mask applied to binary and gray images. *Optics Communications*, **294**: 105–117.
- Asaari, M. S. M., Rosdi, B. A., y Suandi, S. A. (2014). Adaptive kalman filter incorporated eigenhand (akfie) for real-time hand tracking system. *Multimedia Tools and Applications*, pp. 1–27.
- Bishop, C. H., Etherton, B. J., y Majumdar, S. J. (2001). Adaptive sampling with the ensemble transform kalman filter. part i: Theoretical aspects. *Monthly weather review*, **129**(3): 420–436.
- Bolme, D. S., Beveridge, J. R., Draper, B., Lui, Y. M., et al. (2010). Visual object tracking using adaptive correlation filters. En: *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE*, pp. 2544–2550.
- Breitenstein, M. D., Reichlin, F., Leibe, B., Gool, E. K.-m. L. V., y Leuven, K. U. (2009). Robust Tracking-by-Detection using a Detector Confidence Particle Filter. **12**: 1515–1522.
- Casasent, D. (1984). Unified synthetic discriminant function computational formulation. *Applied Optics*, **23**(10): 1620–1627.
- Contreras, V., Díaz-Ramírez, V. H., Kober, V., y Tapia-Armenta, J. J. (2013). Real-time object tracking with correlation filtering and state prediction. *SPIE Applications of Digital Image Processing*, **8856**: 19.
- Díaz-Ramírez, V. H., Picos, K., y Kober, V. (2014). Target tracking in nonuniform illumination conditions using locally adaptive correlation filters. *Optics Communications*, **323**: 32–43.
- Fimbres-Castro, C., Álvarez-Borrego, J., y Bueno-Ibarra, M. A. (2012). Invariant nonlinear correlation and spectral index for diatoms recognition. *Optical Engineering*, **51**(4): 047201–1.
- Geusebroek, J., Burghouts, G., y Smeulders, A. (2005). The amsterdam library of object images. *Int. J. Comp. Vis.*, **61**(3): 103–112.
- González-Fraga, J. Á., Kober, V., y Álvarez-Borrego, J. (2006). Adaptive synthetic discriminant function filters for pattern recognition. *Optical Engineering*, **45**(5): 057005–057005.
- Horner, J. L. y Gianino, P. D. (1984). Phase-only matched filtering. *Applied optics*, **23**(6): 812–816.
- Kalal, Z., Mikolajczyk, K., y Matas, J. (2011). Tracking-Learning-Detection. *IEEE transactions on pattern analysis and machine intelligence*, **6**(1): 1–14.

- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, **82**(1): 35–45.
- Kober, V. y Campos, J. (1996). Accuracy of location measurement of a noisy target in a nonoverlapping background. *JOSA A*, **13**(8): 1653–1666.
- Kumar, B. V., Mahalanobis, A., y Juday, R. D. (2005). *Correlation pattern recognition*, Vol. 25. Cambridge University Press.
- Leichter, I., Lindenbaum, M., y Rivlin, E. (2010). Mean Shift tracking with multiple reference color histograms. *Computer Vision and Image Understanding*, **114**(3): 400–408.
- Li, J., Wang, Y., y Wang, Y. (2012). Visual tracking and learning using speeded up robust features. *Pattern Recognition Letters*, **33**(16): 2094–2101.
- Li, X., Zhang, T., Shen, X., y Sun, J. (2010). Object tracking using an adaptive Kalman filter combined with mean shift. *Optical Engineering*, **49**(2): 020503.
- Li, X. R. y Jilkov, V. P. (2003). Survey of maneuvering target tracking. part i. dynamic models. *Aerospace and Electronic Systems, IEEE Transactions on*, **39**(4): 1333–1364.
- Liu, R., Li, X., Han, L., y Meng, J. (2013). Track infrared point targets based on projection coefficient templates and non-linear correlation combined with Kalman prediction. *Infrared Physics & Technology*, **57**: 68–75.
- Lugt, A. V. (1964). Signal detection by complex spatial filtering. *Information Theory, IEEE Transactions on*, **10**(2): 139–145.
- Mathworks (1995-2015). vision.kalmanfilter class. Recuperado de: <http://www.mathworks.com/help/vision/ref/vision.kalmanfilter-class.html>. octubre 6, 2015.
- Miramontes-Jaramillo, D. y Kober, V. (2013). A fast kernel tracking algorithm based on local gradient histograms. **8856**(3918): 885618.
- Proakis, J. G. y Manolakis, D. G. (1996). *The discrete Fourier transform: Its properties and applications*. Prentice Hall, Inc. pp. 394–447.
- Rodríguez, A., Panza, J., Kumar, B. V. K. V., y Mahalanobis, A. (2010). Automatic recognition of multiple targets with varying velocities using quadratic correlation filters and Kalman filters. *2010 IEEE Radar Conference*, pp. 446–451.
- Santiago-Ramírez, E., González-Fraga, J., y Lázaro-Martínez, S. (2012). Face recognition and tracking using unconstrained non-linear correlation filters. *Procedia Engineering*, **35**: 192–201.
- Solorza-Calderón, S. y Álvarez Borrego, J. (2015). Position and rotation-invariant pattern recognition system by binary rings masks. *Journal of Modern Optics*, **62**(10): 851–864.
- Thompson, M., Gonzalez, R. C., y Wintz, P. (1981). *Digital Image Processing*, Vol. 14. p. 256.
- Ventura, A. S., Borrego, J. Á., y Solorza, S. (2015). Adaptive nonlinear correlation with a binary mask invariant to rotation and scale. *Optics Communications*, **339**: 185–193.

Vijaya Kumar, B. y Pochapsky, E. (1986). Signal-to-noise ratio considerations in modified matched spatial filters. *JOSA A*, **3**(6): 777–786.

Zhou, H., Yuan, Y., y Shi, C. (2009). Object tracking using SIFT features and mean shift. *Computer Vision and Image Understanding*, **113**(3): 345–352.