Seguimiento de múltiples objetos pequeños con manejo de oclusión

Por:

Mario Leopoldo Bustillo Eguiluz

Tesis sometida como requisito parcial
para obtener el grado de:

MAESTRÍA EN CIENCIAS EN LA ESPECIALIDAD DE CIENCIAS COMPUTACIONALES

en el

Instituto Nacional de Astrofísica, Óptica y Electrónica

Febrero, 2015
Tonantzintla, Puebla

Dirigida por:

Dr. Leopoldo Altamirano Robles

©INAOE 2015
Derechos reservados
El autor otorga al INAOE el permiso de
reproducir esta tesis en su totalidad o en partes

Tabla de Contenidos

I Introducción 1

1. Introducción 3
   - 1.1. Introducción General . . . 3
   - 1.2. Planteamiento del problema . . . 5
   - 1.3. Hipótesis . . . 8
   - 1.4. Objetivos . . . 9
     - 1.4.1. Objetivo General . . . 9
     - 1.4.2. Objetivos Específicos . . . 10
   - 1.5. Motivación . . . 10
   - 1.6. Contribuciones . . . 11
   - 1.7. Organización del documento . . . 12

II Marco Teórico 13

2. Marco Teórico 15
   - 2.1. Características y representación de objetos . . . 15
     - 2.1.1. Principales características de los objetos . . . 16
     - 2.1.2. Principales representaciones de objetos . . . 18
   - 2.2. Algoritmos de detección de objetos . . . 19
     - 2.2.1. Algoritmos de segmentación . . . 20
     - 2.2.2. Algoritmos de modelado del fondo . . . 23
     - 2.2.3. Algoritmos de aprendizaje supervisado . . . 28
   - 2.3. Conceptos preliminares de seguimiento de objetos . . . 29
     - 2.3.1. Estado del objetivo . . . 29
     - 2.3.2. Función de distribución de probabilidad . . . 29

2.3.3. Teorema de Bayes 30
2.3.4. Estimación Bayesiana 30
2.3.5. Proceso de Markov 31
2.3.6. *Random Finite Sets* (RFS) 31
2.3.7. Estadísticas de conjuntos finitos 33
2.4. Algoritmos de seguimiento 34
2.4.1. Filtrado para un solo objetivo 34
2.4.2. Filtrado para múltiples objetivos 40
2.5. Asociación de datos 44
2.5.1. Introducción 44
2.5.2. Vecino más cercano 45
2.5.3. Asociación de datos probabilista (PDA) 46
2.5.4. Asociación de datos probabilista conjunta (JPDA) 47
2.5.5. Seguimiento por múltiples hipótesis (MHT) 47
2.5.6. Algoritmo Hungaro 48

III Trabajo relacionado **51**

3. Trabajo relacionado **53**
3.1. Introducción 53
3.2. *Sequential Monte Carlo Probability Hypothesis Density* (SMC-PHD) . 55
3.2.1. SMC-PHD manejo de oclusión y asignación de etiquetas . 55
3.3. Filtro *Gaussian Mixture-Probability Hypotheis Density* (GM-PHD) . 57
3.3.1. GM-PHD manejo de oclusión y asignación de etiquetas . 58
3.4. Discusión 60

IV Diseño **63**

4. Diseño de GM-PHD adaptado **65**
4.1. Introducción 65
4.1.1. Arquitectura del método 67
4.1.2. Caracterización de la oclusión 68
4.1.3. Modelado del sistema 71
4.1.4. Adaptación de la probabilidad de detección 74

4.1.5. Adaptación de la distribución de nacimiento y supervivencia de objetos . . . 75
4.1.6. Asignación de etiquetas . . . 79

V Resultados y Experimentos 83

5. Resultados y Experimentos **85**
5.1. Detalles de implementación . . . 85
5.2. Descripción de base de datos . . . 86
5.3. Construcción del *Ground Truth* . . . 90
5.4. Métricas de evaluación . . . 90
5.4.1. Métricas para algoritmos de seguimiento . . . 91
5.4.2. *Multiple object tracking accuracy* (MOTA) . . . 91
5.4.3. *Optimal Subpattern Assigment* . . . 92
5.4.4. Métricas para la detección de objetos . . . 94
5.5. Experimentos y Resultados . . . 95
5.5.1. Detección de objetos . . . 95
5.5.2. Caracterización de la oclusión . . . 100
5.5.3. Seguimiento en videos de objetos pequeños . . . 107
5.5.4. Seguimiento en escenarios sintéticos . . . 117

VI Conclusiones y trabajo futuro 129

6. Conclusiones y trabajo futuro **131**
6.1. Conclusiones . . . 131
6.2. Trabajo futuro . . . 133
6.3. Revisión de objetivos . . . 135

A. Pruebas de significancia **137**

Lista de Figuras

1.1. Planteamiento del problema 7
1.2. Método general para seguimiento de múltiples objetos [**?**]. 9

2.1. Ejemplo de umbralización con Otsu 22
2.2. Ejemplo de la técnica K-means para imágenes. 22
2.3. Ejemplo de segmentación con Watershed. 24
2.4. Modelado del fondo con Mezcla de Gaussianas. 27
2.5. Adaboost. 29
2.6. Asociación de datos con un solo objeto 45
2.7. Asociación de datos con más de un objeto 46

4.1. Tipos de oclusión. 66
4.2. Arquitectura del método propuesto 67
4.3. Tipos de regiones. 71
4.4. Representación de los cuadrantes usados para impulsar el algoritmo Húngaro. 80

5.1. Apariencia de los objetos de las bases de datos. 88
5.2. Escenarios sintéticos para la comparación del método propuesto. 89
5.3. Ejemplo de las cualidades de la configuración del filtro morfológico elegido. 97
5.4. Preprocesamiento de las bases de datos. 99
5.5. Escenarios con regiones de oclusión con objetos no identificados. 103
5.6. Tasa de falsas alarmas por bases de datos. 108
5.7. Comparación de tiempo de los algoritmos utilizados. 114

5.8. Tiempo en ms de procesamiento entre MHT y el método propuesto (seguimiento + asignación de etiquetas). 118
5.9. Métrica FAR para escenarios sintéticos. 119
5.10. Métrica MOTA para 3 escenarios con diferentes densidades de ruido. 120
5.11. Métrica MOR para 3 escenarios sintéticos modificados. 123
5.12. Métrica MOR para 3 escenarios con diferentes densidades de ruido. 125
5.13. Métrica MLR para escenarios sintéticos modificados. 127

Lista de Tablas

3.1. Tabla comparativa de los principales algoritmos relacionados . . 62

5.1. Caracteristicas de las bases de datos. 87
5.2. Resultados de la evaluación de los algoritmos de modelado del fondo 96
5.3. Parámetros del algoritmo evolutivo 98
5.4. Configuración y contribución de los filtros morfológicos. 99
5.5. Detección de objetos para bases de datos reales. 101
5.6. Comportamiento del método propuesto con diferentes ϵ. 102
5.7. Tiempo de procesamiento del método por cada ϵ 102
5.8. Características de los escenarios. 104
5.9. Regiones obtenidas por el método propuesto. 104
5.10. Comparación entre regiones 105
5.11. Características de las oclusiones en las bases de datos 108
5.12. Métrica ReR para base de datos reales. 109
5.13. Tasa de oclusiones perdidas para los diferentes métodos 110
5.14. Comparación entre métodos con métrica MOTA 112
5.15. Métrica OSPA para los diferentes métodos 113
5.16. Otras características de las bases de datos 115
5.17. Tiempo en ms de cada algoritmo 115
5.18. Tasa de asignaciones para algoritmo Húngaro con impulso 115
5.19. Tasa de asignaciones incorrectas sin tomar en cuenta la oclusión 116
5.20. Tiempo de procesamiento en ms del algoritmo Húngaro con impulso 117
5.21. Tasa de oclusiones pérdidas para base de datos sintéticas. 121

5.22. Características de escenarios para evaluar la oclusión entre objetos 122
5.23. Evaluación de los algoritmos con escenarios sintéticos 124

Agradecimientos

Por el apoyo no condicionado que me obsequiaron para realizar este trabajo, agradezco a:

- A mi padres, que me han criado, con todo el sacrificio que conlleva esa palabra. Su amor y cariño me permitieron seguir y terminar la maestría.
- A mi hermana, que me apoya y cuida.
- A Leonardo, que ha sido un sol en mi vida, iluminando la rutina.
- A Anahi, mujer libre, valiente y capaz, que despierta mi mejor persona, me ayuda con mis temores y metas, enamora mis días, y comparte mi ritmo.
- Al Dr. Leopoldo, que me ayudo enormemente a construir este trabajo, gracias por siempre recibirme, respaldarme y aguantar mis atrasos.
- A mis sinodales, por ayudar a corregir este trabajo y por el tiempo que pasaron revisándolo.
- A mis compañeros de maestría, por empujarme a esforzarme cada día más, compartir noches en vela para hacer trabajos que parecían imposibles.
- A Teporingos FC, por permitirme ser parte de ese proyecto deportivo.
- A Efrain, Voro, Neto, Luis, Valde, Cruz y Mau; por regalarme su compañía en los buenos y malos momentos.
- A mis amigos Gonzalo, Jorge y Arturo, que me ayudaban a olvidar el estrés con su compañía.

- A mis maestros, que con el catalizador de mi vida profesional.
- Al CONACYT por apoyarme económicamente a crecer profesionalmente.

Abstract

One of the main areas of computer vision is object tracking, which has awakened the interest of many researchers, mainly due to their diverse applications, for industry and science. For several decades there has been innovation in object tracking, as a result, many algorithms have emerged. Modern algorithms are able to solve many complex problems, however, to track small objects in occluding states no satisfactory results are obtained yet.

Many tracking algorithms are inappropriate for multiple tracking due to the computational complexity, the dimensionality problem and the lack of a structure capable to represent multiple states and observations. Furthermore, the small targets have not enough pixels associated to generate a good object representation.

During the last decade have emerged techniques that allow multiple targets tracking in noisy environments. Probability Hypothesis Density Filter (PHD) proposed by Mahler, is the main technique for solving multiple targets tracking in a unified framework. The performance of the PHD filter is significantly better compared to other traditional methods such as the Kalman filter, the Particle Filter and JPDAF.

The main problem in the PHD filter and their different approaches is that does not support data association, so problems such as occlusion or target identification can not be handled by this algorithm. In this work an approximation of PHD filter (GM-PHD) is adapted for handling occlusion, by setting detection probabilities and object's distributions automatically. A module for label assignment to identify each target is also incorporated.

The PHD filter proposed is evaluated in different scenarios both synthetic and real. In synthetic scenarios, the filter performance is evaluated against its closest competitors (CGM-PHD, MHT) . In addition, in the videos with real data, small targets are used in situations of high noise density and occlusion, to evaluate the algorithm on the multiple small targets tracking problem.

Resumen

Una de las principales ramas de visión por computadora es el seguimiento de objetos. Esto ha despertado el interés de muchos investigadores debido a sus diversas aplicaciones, tanto para la industria como para la ciencia. Durante varias décadas se ha innovado en esta área, como resultado diversos algoritmos han surgido. Los algoritmos modernos son capaces de resolver muchos problemas complejos, sin embargo, para el seguimiento de objetos pequeños en estado de oclusión todavía no se obtienen resultados satisfactorios.

Muchos algoritmos de seguimiento son inapropiados para el seguimiento múltiple, debido a la complejidad computacional, la dimensionalidad del problema y la falta de una estructura capaz de representar múltiples estados y observaciones. Por otra parte, los objetivos pequeños no tienen suficientes pixeles asociados para generar una representación concluyente.

En la última década han surgido técnicas que permiten el seguimiento de múltiples objetivos en ambientes con ruido. El filtro *Probability Hypothesis Density* (PHD), propuesto por Mahler [**?**], es la principal técnica para resolver el seguimiento de múltiples objetos en un marco unificado multi objetivo. El rendimiento del filtro PHD es perceptiblemente mejor a otros métodos tradicionales como el filtro de Kalman, el filtro de partículas y JPDAF.

El problema principal de PHD y sus diferentes variantes, es que no soportan la asociación de datos, por lo tanto, la oclusión o identificación de objetivos no se pueden manejar en estos algoritmos. En este trabajo se adapta una variante con solución cerrada y práctica de PHD (GM-PHD) para el manejo de la oclusión, configurando tanto las probabilidades de detección y las distribuciones de los

objetos automáticamente. También se incorpora un módulo de asignación de etiquetas para identificar a cada objetivo.

La adaptación del filtro PHD propuesto se evalúa en diferentes escenarios tanto sintéticos como reales. En los escenarios sintéticos se compara el rendimiento del filtro contra sus más cercanos competidores (CGM-PHD, MHT). Por otro lado, en los videos, se utilizan objetivos pequeños en situaciones de alta densidad de ruido y oclusión para evaluar el algoritmo en el problema del seguimiento de múltiples objetivos pequeños.

Parte I

Introducción

Capítulo 1

Introducción

1.1. Introducción General

El permanente y creciente desarrollo tecnológico, así como, el avance gradual en el *hardware* de computadoras y equipos de adquisición de video, han permitido la incorporación de sistemas de visión por computadora en nuestra vida diaria.

Hechos como el ataque terrorista del 11 de septiembre del 2001 en los Estados Unidos, incitó a gobiernos de todo el mundo a invertir en tecnología para sistemas de seguimiento. La Agencia de Seguridad Nacional (NSA) ha promovido frecuentemente proyectos y competencias sobre vigilancia automatizada para aeropuertos, aduanas, estaciones de transporte y espionaje. Por otro lado, en el 2013 en la ciudad de Puebla, se implementó la foto infracción vehicular, que funciona mediante un sistema de seguimiento y reconocimiento de placas de automóviles. Así, un sin número de aplicaciones en diversos campos respaldan la importancia del seguimiento de objetos, el cual no sólo tiene un impacto en la industria, sino también en la ciencia.

El seguimiento de objetos se define como la acción de determinar la posición de un tipo de objeto en una secuencia de imágenes, de forma continua y fiable; de tal forma que el seguidor asigne una etiqueta consistente a cada objeto en la escena. En este documento se aborda el problema del seguimiento de múltiples objetos pequeños en una secuencia de imágenes digitales (video).

El seguimiento de múltiples objetivos (SMO) es un tema de interés en muchas áreas de la ciencia y con un sin número de aplicaciones. Sin embargo, existen diversos problemas para realizar el SMO. La complejidad computacional aumenta en comparación al seguimiento de un solo objeto, además, es necesario un método que permita diferenciar a cada objeto en la escena aun en estados de oclusión.

A la fecha, los enfoques y algoritmos de seguimiento de objetos no han sido completamente eficaces para la detección de múltiples objetos pequeños cuando se desconocen las posiciones y velocidades, es por ello que esto es un problema abierto. Muchos algoritmos robustos utilizados para el seguimiento de un solo objetivo se opacan ante el problema de múltiples objetos.

El seguimiento de múltiples objetos pequeños, es un problema complejo por dos razones principales:

1. Un objeto pequeño se representa por un reducido conjunto de píxeles (entre 10 y 100 [**?**]), y por lo tanto no contiene suficiente información para ser eficientemente representado.

2. Los objetos pequeños se confunden con el ruido presente en la imagen, siendo prácticamente imposible su detección en escenas no controladas.

Una de las aplicaciones más importantes que involucra objetivos pequeños, es el seguimiento de insectos. Varios sistemas han sido desarrollados [**?**] [**?**] [**?**], la mayoría, utilizan el conocido filtro de Partículas para predecir las trayectorias de los objetivos. Sin embargo fallan cuando se siguen múltiples insectos en vuelo.

Actualmente el comportamiento de los insectos se mide manualmente, donde un observador humano con la ayuda de un video determina comportamientos con base en su experiencia. Este proceso es lento, laborioso y en gran medida ineficaz. Muchas veces los observadores pueden perder comportamientos de insectos que son demasiados rápidos o demasiados lentos. Una importante herramienta para el desarrollo de estas aplicaciones, es un método que permita seguir e identificar a cada insecto, aunque estos se ocluyan entre sí.

El filtro *Probability Hypothesis Density* (PHD) ha sido recientemente reconocido como una herramienta eficaz para resolver los problemas de SMO. A pesar que no existen muchas aplicaciones para visión por computadora con este filtro, en problemas sintéticos se ha demostrado su robustez. Un problema importante de los métodos basados en el filtro PHD es la ausencia de un mecanismo de asociación de datos, por lo tanto, no se contempla ni la oclusión, ni la asignación de etiquetas. En este trabajo se empleará un método basado en el filtro PHD, *Gaussian Mixture Probability Hypotheis Density* (GM-PHD), al cual se le incorporarán mecanismos para manejar la oclusión y asignación de etiquetas.

1.2. Planteamiento del problema

A pesar de que numerosos algoritmos y métodos se han propuesto sobre el seguimiento de objetos en visión por computadora, todavía tienen dificultades para trabajar en el seguimiento de múltiples objetos de una misma clase. En ambientes semicontrolados es un problema difícil, debido a diferentes factores como: los grados de oclusión de los objetos en la imagen, el movimiento rápido o brusco del objeto, la estructura no rígida del objeto, el movimiento de la cámara, el ruido en la imagen y los cambios dinámicos en el fondo; entre otros. Es por eso que el seguimiento se realiza generalmente con conocimiento previo de las condiciones de la escena y/o la forma del objeto, limitando así, el problema a aplicaciones particulares.

El seguimiento de objetos es uno de los problemas críticos dentro de los procesos de visión artificial donde la exactitud y el tiempo de respuesta son de gran importancia. Aplicaciones en tiempo real requieren menores tiempos a costa de una disminución en la exactitud. En aplicaciones industriales generalmente se tiene que garantizar una exactitud grande. Sin embargo, lograr una calibración confiable entre el tiempo y la exactitud, es realmente difícil. Para aplicaciones reales de seguimiento múltiple se restringen tanto el tiempo como la exactitud, con base en un análisis de las características del problema.

En la figura **??**, se muestran los principales problemas y condiciones en las que se desenvuelve el método propuesto. Se dividen en 3 principales áreas:

1. Movimiento
   - Oclusión: el método propuesto maneja la oclusión entre objetos.
   - Movimientos discontinuos: los objetos de las bases de datos podrán parar o reanudar su movimiento en periodos desconocidos de tiempo.
   - Cámara estática: el sensor de adquisición para los videos, permanece en un mismo lugar.
2. Ambiente
   - Luz: la luz en la escena no es uniforme.
   - Velocidad: el método propuesto podrá utilizarse para seguir múltiples objetos, alcanzando para las bases de datos propuestas 20 frames por segundo en promedio, tiempo aceptable para aplicaciones de vigilancia visual [**?**].
   - Fondo: el fondo en la escena no es uniforme y contiene ruido.
   - Límites de escena: los objetos pueden salir del campo de visión de la cámara.
3. Características del objeto:
   - Tipo: el tipo y forma del objeto son conocidos.
   - Tamaño: el tamaño del objeto no sobrepasa los 30 pixeles.
   - Número: se seguirán múltiples objetivos, entre 10-100 (mínimo y máximo de objetivos en las bases de datos utilizadas).

Generalmente el seguimiento de objetos consta de 3 pasos principales. En el primero se adquieren y procesan las imágenes de cada frame para coadyuvar a la detección. El segundo paso detecta la posición de los objetos en la imagen y los representa mediante alguna estructura. Por último, de acuerdo a la representación y características extraídas en el segundo paso se sigue al objeto

Figura 1.1: Planteamiento del problema

a través de los diferentes cuadros del video. En este trabajo se realizaron los 3 pasos para evaluar el método propuesto, sin embargo nos centraremos en el seguimiento de objetos. A continuación se describen los pasos necesarios para el seguimiento de múltiples objetos:

1. Diseño de un algoritmo de detección: Se requiere seleccionar un conjunto determinado de características de los objetos, ya sean puntos de interés, información de color o modelado del fondo; de tal forma que se pueda representar al objetivo eficazmente. Cada método de seguimiento requiere un mecanismo de detección, para discriminar el ruido [**?**].

2. Diseño de un algoritmo para el seguimiento: Una vez que se obtuvieron todas las observaciones del sensor y fueron procesadas por un detector, es tarea del seguidor realizar la correspondencia del objetivo de un fotograma a otro.

3. Diseño de un algoritmo que resuelva la oclusión y permita etiquetar a

los objetos: Puede ser que un objetivo de interés no se distinga cuando esté parcial o totalmente ocluido por cualquier objeto en la escena. Por otro lado, para reconocer que un objeto es el mismo a través de las diferentes imágenes que componen el video, es necesario un método de asociación de datos o asignación de etiquetas.

1.3. Hipótesis

Basándose en los resultados obtenidos por [**?**] [**?**] [**?**], podemos inducir que para resolver el problema de seguimiento múltiple es factible utilizar PHD, en concreto, el algoritmo *Gaussian Mixture Probability Hypotheis Density* (GM-PHD), incorporando mecanismos para manejar la oclusión y permitir la asignación de etiquetas. Para el desarrollo de este trabajo se formularon las siguientes preguntas de investigación:

- ¿Por qué usar GM-PHD?
- ¿Cómo se puede modelar el filtro GM-PHD para objetivos pequeños?
- ¿Cómo caracterizar la oclusión?
- ¿Cómo manejar la oclusión en GM-PHD?
- ¿Cómo asignar etiquetas de los objetivos a través de las observaciones en GM-PHD?
- ¿Cómo medir el seguimiento de varios objetivos con estados de oclusión?

Se construyeron las siguientes hipótesis para responder a las preguntas anteriores:

1. GM-PHD alcanza una precisión competitiva en el seguimiento de múltiples objetos para aplicaciones de visión por computadora con alta densidad de ruido, en comparación a algoritmos especializados en seguimiento múltiple, como MHT.
2. La información espacial de los objetivos brindara información suficiente en GM-PHD para caracterizar la oclusión.

3. Si se adaptan las probabilidades de detección y supervivencia de los objetivos a través de las observaciones, se puede desarrollar un método para manejar la oclusión dentro del filtro GM-PHD.

4. Los filtros basados en PHD eliminan el ruido eficazmente, por lo tanto, la asignación de etiquetas se podrá construir como un módulo independiente.

5. Mediante el esquema representado por la figura **??**, se puede tratar el problema del seguimiento de múltiples objetivos en GM-PHD.

6. Se puede evaluar la propuesta realizada utilizando un conjunto de videos, su *ground truth* y las métricas para múltiples objetivos (OSPA, MOTA, MOT).

7. El método propuesto independientemente del detector y el método de asignación, permitirá trabajar con 20 frames por segundo en promedio, para aplicaciones que tengan entre 10 a 50 objetos. Los mecanismos propuestos para la oclusión no repercutirán dramáticamente en el tiempo de ejecución.

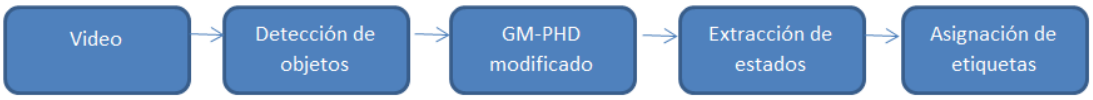


Figura 1.2: Método general para seguimiento de múltiples objetos [**?**].

1.4. Objetivos

1.4.1. Objetivo General

Se diseñará y evaluará un nuevo mecanismo que permita efectuar el seguimiento de múltiples objetos en estado de oclusión y de la misma clase, estimando sus posiciones en imágenes subsecuentes.

1.4.2. Objetivos Específicos

- Definir la extracción y representación del objeto de tal forma que coadyuve a la detección de objetos pequeños.
- Definir estrategias que permitan el manejo de oclusión, independientemente del detector, para el seguimiento de múltiples objetos.
- Elaborar un método que permita el seguimiento de objetos pequeños en estado de oclusión.
- Evaluar el desempeño del método propuesto en escenarios de oclusión y ruido, para el seguimiento de múltiples objetos.

1.5. Motivación

El seguimiento de objetos juega un papel importante en la visión por computadora. Se ha utilizado pertinentemente en una amplia gama de aplicaciones, tales como la interacción humano-computadora, la vigilancia automatizada, la supervisión del tráfico, la realidad aumentada, el seguimiento de objetos astronómicos, la localización de objetivos militares, la vigilancia visual, la compresión de video, entre otras.

Se han propuesto e implementado para el seguimiento múltiple, una variedad de algoritmos [?] [?] [?] [?] [?], los cuales en situaciones críticas pierden al objeto en la escena y/o realizan falsas correspondencias. Además, estos métodos son inapropiados para objetos pequeños, debido a que los objetos no tienen prácticamente estructura asociada a ellos y el ruido en la imagen dificulta su detección, generando falsos positivos.

Existen diversos problemas que vuelven un reto el seguimiento de múltiples objetos:

- Cambios de posición: Varía el aspecto del objeto, cuando se proyecta en la imagen.

- Iluminación: La dirección, la intensidad y el color de la luz influyen en el aspecto del objeto.
- Ruido: El proceso de adquisición usualmente produce señales con ruido, dependiendo de la calidad del sensor.
- Oclusiones: Otros objetos, sean de interés o no, pueden entorpecer la visión de un objeto de interés

El seguimiento automático de objetos con las características anteriormente descritas es uno de los grandes retos de la visión por computadora.

1.6. Contribuciones

Este trabajo de tesis tiene como finalidad presentar una adaptación de GM-PHD para el manejo de oclusión, la cual permite también, seguir múltiples objetos pequeños en ambientes con ruido, estimando su posición en imágenes subsecuentes. El método propuesto es funcional en escenarios con características similares a las descritas en la problemática y es robusto a ambientes con alta densidad de ruido (20 %-70 %). Su desarrollo se basa en los novedosos algoritmos *Multi-Tracking*, que llevan como teoría principal los conjuntos aleatorios finitos y la recursión Bayesiana. Las principales contribuciones de este trabajo son la incorporación de mecanismos en el filtro GM-PHD para:

- Contrarrestar el problema de la oclusión de objetivos pequeños.
- El seguimiento de múltiples objetivos pequeños en ambientes con alta densidad de ruido.
- El manejo de la oclusión, mediante la probabilidad de detección adaptable de cada objetivo.
- Coadyuvar a la eliminación del ruido, mediante una distribución adaptable de aparición de nuevos objetivos y supervivencia.
- Un mecanismo independiente de asignación de etiquetas para GM-PHD.

1.7. Organización del documento

El presente documento, está estructurado en 5 partes:

Parte 1: Introducción, motivación y objetivos del proyecto.

Parte 2: Marco teórico de las principales áreas utilizadas en este trabajo.

Parte 3: Análisis del trabajo relacionado de las diferentes técnicas existentes y significativas para el presente trabajo.

Parte 4: Descripción del diseño e implementación del algoritmo propuesto.

Parte 5: Experimentos y resultados.

Parte 6: Conclusiones y trabajo futuro.

Parte II

Marco Teórico

Capítulo 2

Marco Teórico

En este capítulo se provee al lector con la información necesaria para entender el diseño de este trabajo de tesis. En las primeras secciones se incorporan temas dedicados a la caracterización, representación y detección de objetos. Con el propósito de brindar un panorama general de este extenso tema e introducir las técnicas ocupadas en el método propuesto. A modo de introducción para el seguimiento de objetos, se comparten los conceptos mas importantes utilizados en el seguimiento de múltiples objetos. Posteriormente se comparten las principales técnicas para el filtrado de uno y múltiples objetivos. Finalmente, se explican las principales técnicas para la asociación de datos y asignación de etiquetas.

2.1. Características y representación de objetos

En esta sección, se introducirán las técnicas usadas para caracterizar un objeto. Posteriormente se analizarán las representaciones de los objetos empleadas por el proceso de seguimiento.

A modo de introducción se define un objetivo en visión por computadora, como cualquier conjunto de pixeles bien delimitado que es importante para el problema a resolver. Los objetivos pueden ser representados por su forma, color, apariencia, tamaño, entre otros [**?**]. La mejor elección para representar un objeto se deduce de un análisis del problema, por ejemplo, en un escenario con tonos homogéneos o una intensidad homogénea resultaría mejor representar a los objetos basándose en alguna característica que no sea color.

2.1.1. Principales características de los objetos

La selección de las características y la forma de representar los objetos en problemas de seguimiento es un paso crítico, una mala decisión puede desencadenar un mal funcionamiento de un buen algoritmo. Las representaciones de objetos en conjunto con las características permiten a los objetos en una imagen diferenciarse de los otros. Entre más singular sea una propiedad o característica, más significativa y deseable. En algunos problemas no es posible distinguir a los objetos con una sola característica, por lo que se utiliza una combinación de ellas [**?**].

En [**?**] se menciona que las principales características utilizadas son:

- Color: El color es la característica más usada en la detección. Existen diversos espacios que permiten describir el color en un objeto, el común es RGB, sin embargo no siempre es el más eficaz, en aplicaciones como reconocimiento de piel o detección de rostros, suele ser preferido un espacio uniforme como HSV o HSL [**?**].

- Bordes: Los bordes son características binarias eficaces en imágenes con contrastes entre regiones. En comparación con el color, los bordes son más robustos a cambios de iluminación.

  > En la práctica, los bordes se definen generalmente como conjuntos de puntos en la imagen donde existe una fuerte magnitud del gradiente. Es por ello, que muchos algoritmos de detección de bordes, buscan encadenar puntos con un alto gradiente entre sí para formar una descripción más completa. Estos algoritmos suelen colocar algunas restricciones en las propiedades de un borde, tales como: la forma, la suavidad, y el valor del gradiente [**?**].

- Esquinas o puntos de interés: Los términos de esquinas y puntos de interés son empleados indistintamente en visión por computadora. Son características puntuales en una imagen, que tienen una estructura de dos dimensiones (x,y). Estas características surgieron a partir del análisis de los bordes, donde el gradiente cambia repentinamente. Cabe señalar,

que los puntos de interés no necesariamente son parte de los bordes (por ejemplo un pequeño punto brillante sobre un fondo oscuro, es un punto de interés, pero no sería considerado borde). Los algoritmos más representativos de estas características son SIFT, SURF y ASIFT [**?**] [**?**] [**?**].

- Regiones o *Blobs*: Las regiones proporcionan una descripción de la imagen basándose en características similares. Los descriptores de Blobs suelen solo contener un punto (centro de gravedad o centroide), por lo que suelen ser considerados como operadores de puntos de interés. Los Blobs relacionan pixeles basándose en una característica específica para formar regiones. El estudio y desarrollo de estos detectores es importante para brindar información sobre áreas que no se pueden obtener mediante detectores de bordes o detectores de esquinas. Los detectores de regiones se usan como paso previo para el reconocimiento de objetos o seguimiento de objetos.

- Textura:

  > La textura mide las variaciones de intensidad en la superficie de los objetos [**?**].

  La textura es una propiedad asociada con las superficies tales como suavizado, rugosidad, granularidad, regularidad. En visión por computadora, la textura es la repetición espacial de patrones en una superficie. Los descriptores de texturas son robustos a cambios en la escena como tamaño e iluminación, sin embargo no es común utilizarlos. La textura se puede describir mediante estadísticas de patrones, usando el espectro de Fourier o con filtros de multiresolución. Entre los descriptores más utilizados se encuentran el descriptor de Law y *Gray-Level Cooccurrence Matrices*.

2.1.2. Principales representaciones de objetos

Una vez elegida una o más características del objeto, éste se tiene que representar mediante alguna estructura. Existen un sin número de representaciones, las cuales al igual que las características pueden fusionarse, sin embargo no es tan común debido al aumento de la complejidad. Algunas de las representaciones más utilizadas para realizar el seguimiento de objetos son:

- Representación por medio de puntos: El objeto es representado por un punto que puede ser su centroide o centro de masas. Es la representación más adecuada si el problema solo requiere conocer la posición del objeto. Generalmente se ocupa para objetos pequeños cuya forma no es representativa [**?**].

- Representación por silueta o contorno: En algunas ocasiones se necesita más de un punto para representar un objeto, entonces, se recurre a una representación de silueta. Esta representación se define a partir del límite del objeto y el fondo (contorno), donde la silueta es el área interior del contorno. Generalmente se utiliza para objetos cuya forma cambia a través de los cuadros, denominados objetos no rígidos.

- Representación por modelos esqueléticos: Es una representación más compleja y suele utilizarse para representar objetos articulados. Existen técnicas basadas en grafos, puntos de interés y contornos, las cuales extraen los esqueletos. Una vez obtenido el esqueleto, por medio de clasificadores se recupera la clase de objeto a la que pertenece el esqueleto seleccionado.

- Representación por plantillas: Esta representación brinda información de la apariencia del objeto mediante un conjunto de plantillas. La representación por plantillas se utiliza cuando el objeto no cambia su perfil y/o apariencia. En el último lustro se ha utilizado en gran medida la representación *Sparse* [**?**], la cual representa a un objeto por medio de una combinación lineal de un número de plantillas registradas. El problema fundamental de esta representación es elegir una distancia para relacionar las plantillas con el objeto.

- Representación por modelos de apariencia: Esta representación puede ser construida con 3 componentes, modelos de formas, modelos de apariencia o transformación de la pose. Se requiere una fase de entrenamiento para aprender el modelo [**?**]. Esta estructura compleja no solo sirve para caracterizar a un objeto, sino también para saber que representa el objeto en determinado estado. Generalmente se modelan clases de objetos, y se buscan similitudes (distancia, características) con los objetos en la imagen.

2.2. Algoritmos de detección de objetos

Esta sección se centra en brindar un panorama general de la amplia área de detección de objetos. En las secciones anteriores se mencionaron las técnicas elementales para representar y generar características de los objetos en la imagen, ahora se realiza una clasificación de los principales métodos para la detección de objetos.

Todo sistema de seguimiento de objetos requiere un paso para la detección del objeto de interés. Un gran éxito se ha logrado en un medio controlado, pero el problema sigue sin resolverse en ambientes semi-controlados, cuando los objetos se colocan en entornos ruidosos, donde puede existir oclusión [**?**]. La detección de objetos es un problema desafiante y útil en visión por computadora. En general, la detección de objetos se encarga de la identificación de la presencia de varios objetos en una imagen. Un análisis profundo de las técnicas y algoritmos en detección de objetos requeriría un trabajo más extenso.

En esta sección se definirá una clasificación de los algoritmos base para la detección de objetos en las siguientes ramas:

1. Algoritmos de segmentación.
2. Algoritmos de modelado del fondo.

2.2.1. Algoritmos de segmentación

Cuando hablamos de segmentar una imagen, nos referimos a localizar regiones de interés y etiquetarlas en alguna clase. Generalmente estas regiones se definen en dos grupos: fondo y objetos de interés. En [**?**], la segmentación de una imagen se define así:

> La segmentación es el proceso de dividir una imagen digital en varias partes (grupos de píxeles) u objetos. El objetivo de la segmentación es simplificar y/o cambiar la representación de una imagen en otra más significativa y más fácil de analizar. La segmentación se usa tanto para localizar objetos como para encontrar los límites de estos dentro de una imagen.

El problema principal de los algoritmos de segmentación, es localizar el umbral óptimo para lograr una separación confiable de las regiones. Para ello es necesario un análisis de las características del objeto y la escena. En esta sección se abordarán técnicas de umbralización y agrupamiento.

Técnicas de umbralización

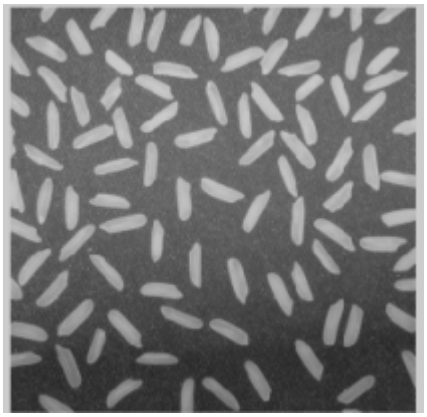
La umbralización es ampliamente empleada en visión por computadora, cuando existen diferencias marcadas entre los objetos con respecto al fondo de la escena. En [**?**] se establece que los principios que rigen a la umbralización son la similitud entre píxeles pertenecientes a un objeto y sus diferencias respecto al resto. Por tanto, la escena debe caracterizarse por un fondo uniforme y objetos similares.

Los algoritmos que se encuentran en esta clasificación se basan en definir uno o varios umbrales, permitiendo así a los objetos definirse dentro de un rango de valores de pixeles. El problema principal de este tipo de técnicas reside en encontrar ese rango óptimo de localización. Entre las técnicas de umbralización más representativas se encuentran:

- Auto umbralización: Se trabaja sobre una imagen de un solo canal, utilizando múltiples umbrales. Se construye un histograma en escala de grises de la imagen. Los valles y picos se extraen del histograma, y se utilizan iterativamente como parámetros para encontrar un umbral óptimo.

- Umbralización con precisión subpixel: Encuentra las regiones de interés con una precisión subpixel; de manera estricta esta técnica no obtiene las regiones, sino las líneas que separan las regiones con un valor menor o mayor que un umbral definido. Para la extracción, la imagen de entrada se considera como una superficie en la que los valores se interpolan entre los centros de los píxeles individuales. No es una técnica común, por la complejidad de su implementación, pero puede ser útil cuando los objetos en la imagen son pequeños y se requiere una mayor precisión.

- Umbralización con método de Otsu: El problema de la umbralización reside en la determinación del valor umbral que segmenta la imagen. El



(a) Imagen original

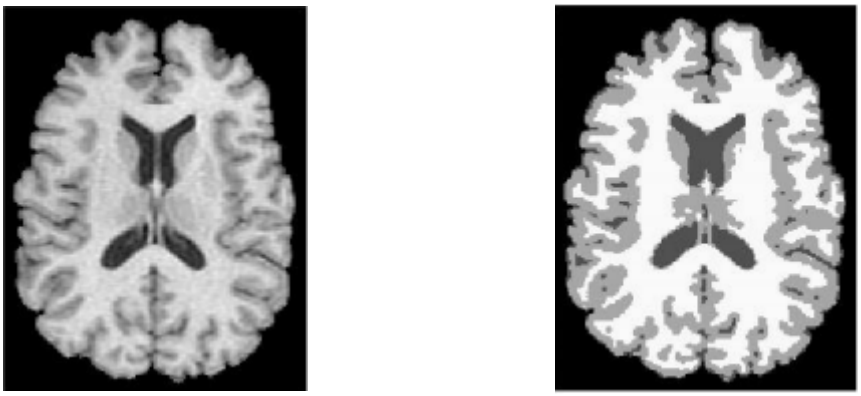
(b) Imagen umbralizada

Figura 2.1: Ejemplo de umbralización con Otsu

método Otsu determina automáticamente el umbral óptimo de segmentación, de tal forma que la dispersión en cada región sea lo más pequeña posible, y lo más alta posible entre segmentos diferentes. Se realiza una relación entre varianzas y se busca un valor umbral para que la relación sea máxima. Este método proporciona el umbral óptimo (*threshold*) para la segmentación de la imagen, en [**?**] se caracteriza este valor óptimo bajo el criterio de máxima varianza entre fondo (*background*) y objeto (*foreground*). Se calcula la varianza entre todas las posibles divisiones, y se toma el umbral que presenta la máxima varianza entre clases. Existen otras técnicas basadas en Otsu las cuales innovan en la búsqueda del umbral, utilizando heurísticas [**?**].

Técnicas de agrupamiento

Las técnicas de agrupamiento son procedimientos en los cuales un conjunto de pixeles se dividen de acuerdo a un criterio. Los criterios de clasificación son distancia o similitud. Generalmente se utiliza una matriz de correlación o se maximiza una propiedad entre los diferentes grupos. Los algoritmos de agrupamiento llevan a cabo esencialmente la misma función que los métodos clasificadores, pero sin utilizar datos de entrenamiento, por lo tanto, son métodos no supervisados. Los métodos de agrupamiento iteran entre segmentar la imagen y caracterizar las propiedades de cada clase. En este sentido, los



(a) Imagen original. **(b)** Imagen resultado.

Figura 2.2: Ejemplo de la técnica K-means para imágenes.

métodos de agrupamiento se entrenan usando los datos en la imagen.

Entre las técnicas representativas de esta clasificación se encuentran:

- *K-means*: En esta técnica se eligen k pixeles con base en un método de búsqueda, los cuales se emplearán como centros. Posteriormente, se asigna cada pixel de la imagen a la región que minimiza la varianza entre el pixel y el centro de la región. Por último se calculan nuevamente los centros de las regiones, con la media de sus pixeles. Así sucesivamente hasta que la clasificación de regiones varié por debajo de un umbral. Una de las principales características de esta técnica es que a diferencia de la umbralización encuentra varias clases de objetos, figura **??**.

- *Mean-Shift*: Es una técnica de agrupación no paramétrica que no requiere conocimiento previo del número de grupos, y no limita su forma. Esta técnica representa los pixeles dentro de clases mediante una función de densidad, generalmente asociada al color. En *Mean-Shift* se encuentran los máximos de una función de densidad para resolver en la imagen, la probabilidad de que un pixel sea parte de una región. Esta técnica encuentra los puntos estacionarios de la función de densidad, y poda estos puntos reteniendo sólo los máximos locales [**?**].

- *Watershed*: La imagen se trata como una superficie topológica. Se utiliza el gradiente de la imagen para construir la superficie. Los mínimos se tratan como pixeles pertenecientes a una misma región y los máximos

son los límites de las regiones. Un problema inherente a esta técnica es que posiblemente segmente en muchas partes la imagen, algunas veces convirtiendo una región de interés en dos o más. Vea figura **??**.

2.2.2. Algoritmos de modelado del fondo

El modelado del fondo o detección del primer plano, son un conjunto de técnicas de visión por computadora, usadas para la detección de objetos. Sus aplicaciones pueden ser vigilancia, captura de movimiento óptico, aplicaciones multimedia, edición de video, entre otras. El modelado del fondo se utiliza a menudo para detectar objetos en movimiento en un video adquirido por una cámara estática. Los algoritmos pueden tener dos enfoques principales [**?**]:

1. Modelos de actualización temporal: El modelo del fondo se calcula a partir de *frames* anteriores, para ello es necesaria la existencia de movimiento.
2. Modelos entrenados: El modelo del fondo se calcula primero a partir de datos de entrenamiento donde no estén presentes los objetos de interés.

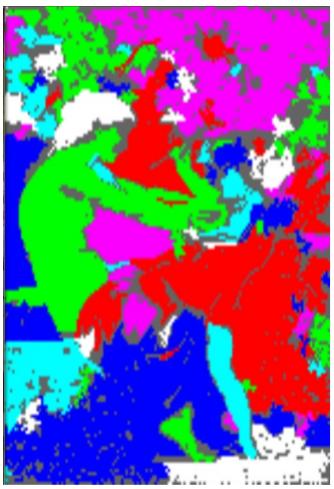
Generalmente para el modelado del fondo, se construye un modelo del fondo B_t asociado al plano de la imagen I_t en el instante t. Posteriormente, en la etapa de detección utilizando el modelo de fondo, se decide que pixeles pertenecen al fondo (valor 1) y cuales pixeles pertenecen al fondo (valor 0), dando como resultado una imagen binaria BW_t.

En esta sección se abordarán las técnicas principales del modelado del fondo, las cuales se clasificarán en dos clases: básicas y estadísticas.

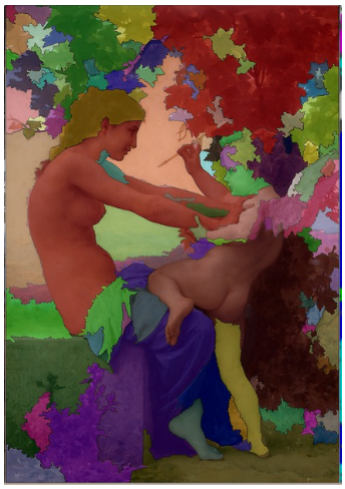
Técnicas Básicas

Este tipo de técnicas en su mayoría están basadas en la diferencia entre fotogramas, utilizan información de las imágenes anteriores para decidir si existe un objeto en movimiento. En algoritmos simples basta con calcular la

(a) Imagen original.



(b) Imagen con segmentación Watershed.



(c) Watershed para detectar piel.

Figura 2.3: Ejemplo de segmentación con Watershed.

diferencia pixel a pixel del modelo de fondo con cada plano y aplicar un umbral:

$$|I_t(x,y) - B_t(x,y)| < Th \tag{2.1}$$

Donde I_t es el frame o imagen en el tiempo t, B_t es el modelo del fondo en el tiempo t y Th es un umbral de decisión.

- Diferencia con modelo estático: La forma más simple de modelar el fondo, es mediante una imagen de fondo que no incluya ningún objeto en movimiento, el modelo entonces será estático. Sin embargo en algunos escenarios, el fondo no está disponible y pueden cambiar constantemente, la iluminación, objetos que se introducen o se retiran de la escena, el ruido de la adquisición, etc. Mediante **??** se define el modelo del fondo para este tipo de técnicas.

  $$B_t = I_r \tag{2.2}$$

  Donde I_r es una imagen estática del fondo y la asignación representa que el modelo B_t se convertirá en la imagen estática I_r.

- Diferencia de frames: Se comprueba la diferencia entre el fotograma actual y su antecesor. Si los píxeles han cambiado drásticamente en una misma área, se deduce que es un objeto en movimiento. Esta técnica en particular no es muy común, debido a que la detección podría generar falsos positivos cuando las condiciones en el escenario cambian dramáticamente.

  $$B_t = I_t - I_{t-1} \tag{2.3}$$

  En **??** se representa como el modelo del fondo en el tiempo t, se convertirá en la diferencia entre el frame en el tiempo t menos el frame en el tiempo $t-1$. En el tiempo $t = 0$ el modelo será I_0.

Técnicas Estadísticas

En este tipo de algoritmos el modelo del fondo está definido por una función de probabilidad condicional. Como en el caso de las técnicas básicas se aplica un umbral para clasificar cada pixel. La detección de un objetivo se logra cuando la probabilidad de un pixel en el tiempo t, dado un modelo del

fondo, supera un umbral Th determinado.

$$P(I_t(x,y)|B_t(x,y)) < Th(x,y) \tag{2.4}$$

También se puede utilizar la varianza de cada pixel, para definir el fondo:

$$\frac{P(I_t(x,y)|B_t(x,y))}{\sigma(x,y)} < Th(x,y) \tag{2.5}$$

De las ecuaciones **??** y **??** una serie de métodos se pueden derivar.

- *Running Gaussian Average*: Si se considera que el fondo no varía considerablemente con el tiempo, y la mayor variación que existe es debida al ruido que introduce la cámara, se puede modelar cada pixel del fondo como una distribución Gaussiana. Esta primer técnica permitió realizar un modelo de fondo independiente para cada pixel de la imagen $I_t(x,y)$ con un ajuste a una función de probabilidad Gaussiana.

- *Gaussian Mixture Model*: Por otro lado, para modelar fondos multimodales, en los que cada pixel de la imagen puede tomar diversos valores, se modela cada canal de color de cada pixel como una mezcla de K Gaussianas. Modelar el fondo con mezcla de Gaussianas ha sido explorado en numerosos artículos [**?**] [**?**], obteniendo buenos resultados en un tiempo aceptable.

  Se puede usar un modelo del fondo a partir de mezclas de Gaussianas (GMM) cuando las imágenes tienen una ligera variabilidad. Por ejemplo, un pixel se puede encontrar en un área, donde un objeto en movimiento aparece sobre un fondo uniforme, entonces el pixel toma en ocasiones el color del objeto y en otras el color del fondo. Este cambio de color es modelable, mediante una GMM de 2 Gaussianas. Donde, una Gaussiana se ajustaría cuando el pixel sea color del objeto y la otra cuando sea color del fondo.

  Formalmente una mezcla de Gaussianas es una función de probabilidad de densidad paramétrica, representada como una suma ponderada de k

(a) Imagen original. **(b)** Imagen resultado de GMM.

Figura 2.4: Modelado del fondo con Mezcla de Gaussianas.

Gaussianas, definida por la ecuación **??**.

$$p(I(x,y)|\theta) = \sum_{j=1}^{k} \omega_j(x,y)\phi(I(x,y)|\mu_j(x,y), \sum_j (x,y)) \tag{2.6}$$

Los parámetros de la mezcla de la Gaussiana son estimados generalmente con el algoritmo *Expectation-Maximization* (EM) o *Maximum A Posteriori* (MAP) [**?**].

En la ecuación **??** ω son los pesos de cada una de las componentes de la mezcla, y cumplen que $\sum_{j=1}^{k} \omega_j(x,y) = 1$ y $\omega_j(x,y) > 0$.

Por otro lado la función ϕ, es la Gaussiana asociada a cada pixel en (x,y).

Para estimar si un pixel $I_t(x,y)$ pertenece al fondo, es necesario umbralizar la verosimilitud de un dato en su GMM. De tal forma que se elija su clase, como se muestra en **??** [**?**]:

$$BW_n(x,y) = \begin{cases} 0 & \text{si } p(I_t(x,y)|\theta(x,y))) > Th \\ 1 & \text{si } p(I_t(x,y)|\theta(x,y))) <= Th \end{cases} \tag{2.7}$$

2.2.3. Algoritmos de aprendizaje supervisado

La detección de objetos se puede realizar aprendiendo diferentes vistas de objetos a partir de un conjunto de referencia. Aprender las diferentes vistas de los objetos exime de la obligación de almacenar un conjunto completo de plantillas. En la detección de objetos, los algoritmos de aprendizaje se componen de las características del objeto y una clase de objetos asociados, donde ambas se definen de forma supervisada.

Los métodos de aprendizaje requieren grandes colecciones de ejemplos de cada clase y su entrenamiento requiere un tiempo considerable, sin embargo la detección generalmente es rápida. Entre los métodos más utilizados en detección de objetos se encuentran las redes neuronales, *adaptative boosting*, máquinas de soporte vectoriales, Haar-like feautures, etc.

- Maquinas de soporte vectorial: Como clasificador, las máquinas de soporte vectorial (SVM) se utilizan para separar a una clase de la otra. El margen de una región, se define por la distancia entre la región y los puntos de datos más cercanos. Los puntos de datos que se encuentran en el límite del margen se llaman vectores de soporte. En el contexto de la detección de objetos, estas clases corresponden a la clase de objeto (muestras positivas) y la clase que no es objeto (muestras negativas).

  Como aplicación de SVM para la detección de objetos, [**?**] utilizó SVM para la detección de peatones y rostros en imágenes. Las características utilizadas para discriminar entre las clases se extrajeron mediante la aplicación de wavelets de Haar a los conjuntos de ejemplos de entrenamiento positivo y negativo.

- *Adaptative boosting*: Los métodos de boosting son algoritmos que en el enfoque supervisado conocen la solución a priori y utilizan esto para adaptar su comportamiento. El boosting se basa en la pregunta ¿Puede un conjunto de clasificadores débiles crear un solo clasificador fuerte?.

  Un clasificador débil es definido como un clasificador que se correlaciona levemente con la clasificación final. En cambio, un clasificador fuerte es un clasificador que se correlaciona fuertemente con la clasificación. *Adap-*

tative Boosting (AdaBoost) es muy popular y quizás el más significativo, pues históricamente fue el primer algoritmo que podía adaptarse a los clasificadores débiles [**?**].

Figura 2.5: Adaboost.

2.3. Conceptos preliminares de seguimiento de objetos

2.3.1. Estado del objetivo

En el filtrado para el seguimiento de múltiples objetivos, el estado de un objeto se refiere a los parámetros desconocidos que el filtro está destinado a determinar (predecir y estimar). Podría ser la velocidad, aceleración, posición, dirección, etc. También podría ser una combinación de dos o más de estos parámetros. En este trabajo de tesis, el estado del objetivo se refiere a la posición del objeto en una imagen con respecto al tiempo y su etiqueta.

2.3.2. Función de distribución de probabilidad

Dada una variable aleatoria X, su función de distribución de probabilidad es:

$$F_X(x) = P(X \leqslant x) \tag{2.8}$$

Si y solo si, cumple con:

- F_X es una función monótona no decreciente.
- F_X es una función continua por la derecha.

- $\lim_{x \to -\infty} f(x) = 0$ y $\lim_{x \to +\infty} f(x) = 1$

2.3.3. Teorema de Bayes

El teorema de Bayes se utiliza sobre todo para inferencia estadística, consiste en la relación entre las probabilidades condicionales de dos variables. La probabilidad condicional se define como:

$$P(A|B) = \frac{P(A, B)}{P(B)} \tag{2.9}$$

Donde $P(A|B)$ denota la probabilidad del evento A, una vez que se conoce la probabilidad del evento B. Entonces, el teorema de Bayes se define como:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \tag{2.10}$$

2.3.4. Estimación Bayesiana

La estimación Bayesiana es el núcleo del problema de filtrado recursivo, donde el objetivo es obtener una caracterización de la incertidumbre sobre el valor de un parámetro, dado un conjunto de observaciones de un experimento [**?**]. Cuando los parámetros tienen propiedades dinámicas que cambian con el tiempo, o cuando nuevas mediciones están disponibles, el proceso de estimación se realiza de forma recursiva con el fin de incorporar nueva información del proceso. La estimación Bayesiana analiza las observaciones y calcula un valor estimado según el grado de creencia planteado en una hipótesis. Esto implica que al tener mayor cantidad de datos disponibles se podrá obtener resultados más satisfactorios.

El estado desconocido de un parámetro se denotará x_k, donde el subíndice k indica el tiempo discreto actual (aunque el filtrado se puede realizar en tiempo continuo, las mediciones de la cámara, se producen en momentos discretos). Entonces, en un determinado tiempo k, se busca estimar x_k con una alta precisión. La estadística Bayesiana, permite incorporar información externa y modificar la probabilidad ante la ocurrencia de otro evento mediante la utilización del teorema de Bayes.

$$p(x_k|Z_k) = p(x_k|z_k, Z_{k-1}) \tag{2.11}$$

$$p(x_k|Z_k) = \frac{p(z_k|x_k)p(x_k|Z_{k-1})}{p(z_k|Z_{k-1})} \tag{2.12}$$

Donde, Z_k representa una secuencia de mediciones del tiempo 1 al tiempo k. Una medición puede ser un vector de medidas, por ejemplo, posiciones, velocidades y aceleraciones. La secuencia de mediciones es entonces un conjunto ordenado, como se observa en **??**.

$$Z_k = \{z_1, ..., z_k\} \tag{2.13}$$

En el filtrado de un solo objeto, la función de distribución de probabilidad posterior (pdf), se usa para inferir el estado del objetivo oculto.

2.3.5. Proceso de Markov

En la teoría de la probabilidad, un proceso de Márkov, es un fenómeno aleatorio dependiente del tiempo para el cual se cumple una propiedad específica: en un proceso de Márkov (sin memoria), la probabilidad condicional sobre el estado presente, futuro y pasado del sistema son independientes. Por otro parte, el término cadena de Márkov se usa para dar a entender que un proceso de Márkov tiene un espacio de estados discreto (infinito o numerable).

2.3.6. *Random Finite Sets* (RFS)

En la última década, se constituyó una revolución teórica y práctica de varios algoritmos de filtrado multiobjetivo, gracias a la introducción de los conjuntos finitos aleatorios (RFS). Los cuales ofrecen un medio para integrar los diferentes aspectos de la fusión de datos y capacidad para trabajar en un solo marco probabilístico diferentes aspectos importantes del problema multiobjeto.

Los conjuntos finitos aleatorios (RFS), introducidos por Mahler [**?**] [**?**] [**?**], junto con las estadísticas para conjuntos finitos (FISST). Son herramientas que permiten la formulación Bayesiana para múltiples objetivos. Los RFS han generado un gran interés en los últimos años, debido en gran parte a que son la base de los métodos de filtrado *Probability Hypothesis Density* (PHD).

Un conjunto aleatorio finito es un conjunto desordenado y finito de elementos distribuidos de acuerdo a una distribución de probabilidad, donde no sólo los elementos del conjunto son aleatorios sino también la cardinalidad del conjunto. La formulación de un RFS permitió de forma natural modelar problemas de seguimiento de múltiples objetos y constituye la base de la que la formulación Bayesiana para múltiples objetos.

De manera formal un RFS se define por medio de la definición **??**, en [**?**] [**?**].

Definición 2.3.1 *Dado un espacio separable de Hausdorff* E *y sea* $\mathcal{F}(E)$ *la colección de todos los subconjuntos finitos de* E*. Un RFS* Ξ *en* E *es definido como una mapeo:*

$$\Xi : \Omega \to \mathcal{F}(E) \tag{2.14}$$

Donde Ω *es una muestra del espacio con un probabilidad medible* P*.*

La probabilidad P produce una ley para el RFS Ξ, la cual está definida por cualquier subconjunto de Borel $\mathcal{T}$ de $\mathcal{F}(E)$.

$$P(\mathcal{T}) = P(\{\omega \in \Omega : \Xi(\omega) \in \mathcal{T}\}) \tag{2.15}$$

Otra alternativa para definir la ley producida por P es una función de creencia β_Ξ y el subconjunto cerrado S de E.

$$\beta_\Xi(S) = P(\{\omega \in \Omega : \Xi(\omega) \subseteq S\}) \tag{2.16}$$

En la práctica $\Xi = \mathbb{R}^n$, donde n es la dimensión del vector de estados x_k.

$$\Xi_k = \begin{cases} \emptyset & \text{si no existe ningún objetivo} \\ x_k & \text{si solo existe un objetivo} \\ \{x_k^{(1)}, x_k^{(2)}\} & \text{si existen dos objetivos} \\ \vdots & \\ \{x_k^{(1)} ..., x_k^{(n)}\} & \text{si existen n objetivos} \end{cases} \tag{2.17}$$

Una propiedad importante es que los elementos de un RFS pueden estar desordenados y aún así representar satisfactoriamente los problemas de seguimiento múltiple.

$$\Xi_k = \{x_k^{(1)}, x_k^{(2)}\} = \{x_k^{(2)}, x_k^{(1)}\} \tag{2.18}$$

El problema del seguimiento múltiple entonces se puede definir dentro de las estadísticas para RFS (FISST), donde:

- X_k es el RFS de los objetivos en el tiempo k.
- Z_k es el RFS de las mediciones del tiempo 1 hasta k

Se busca, entonces, calcular $p(X_k|Z_k)$, como en el filtro Bayesiano de un solo objetivo, solo que ahora el conjunto de estado se comprende de múltiples objetivos. En el paradigma de filtrado Bayesiano, la distribución de probabilidad del estado en el instante k dado todas las observaciones hasta el instante k es de gran importancia. Por ello se necesita el concepto de un conjunto finito aleatorio para el problema de filtrado multiobjeto en el marco Bayesiano.

2.3.7. Estadísticas de conjuntos finitos

Las estadísticas de conjuntos finitos (FISST) es un enfoque propuesto por Mahler, y se convirtió en un marco de trabajo para solucionar el problema del seguimiento de múltiples objetos. El uso de RFS permite modelar múltiples objetivos y medidas recibidas por los sensores, sin embargo RFS solo es una estructura no un marco de trabajo, las reglas y herramientas para describir, analizar y realizar el seguimiento multiobjeto se encuentran en la estadística de esta estructura (FISST).

Precisamente, FISST proporciona un conjunto de herramientas matemáticas que permite la aplicación directa de la inferencia Bayesiana para problemas multiobjetivo. Siendo la clave para una formulación rigurosa de la estimación de objetivos múltiples como un problema de filtrado Bayesiano.

FISST se basa en conceptos probabilísticos no convencionales y sus relaciones con la probabilidad convencional no es clara. En particular, no se ha establecido la validez del filtro FISST Bayes. Desde un punto de vista teórico, los conceptos centrales de FISST son la derivada e integral ya que son la base para el filtrado Bayesiano.

En un sistema multiobjetivo, X_k es el conjunto finito aleatorio de estados de todos los objetivos en el tiempo k. Si existen $M(k)$ objetivos en R^{n_x}, entonces $X_k = \{x_{k,1}, x_{k,2}, ..., x_{k,M(k)}\} \subseteq R^{n_x}$. El espacio de estados multiobjetivo es la colección $\mathcal{F}(R^{n_x})$ y el espacio de mediciones es la colección $\mathcal{F}(R^{n_z})$.

Para realizar la recursión Bayesiana para el seguimiento multiobjetivo, se requiere una noción de integración sobre los espacios $\mathcal{F}(\mathbf{R}^{n_x})$ y $\mathcal{F}(\mathbf{R}^{n_z})$.

2.4. Algoritmos de seguimiento

Un algoritmo de seguimiento generalmente consta de dos partes, el filtrado y la asignación de etiquetas. El objetivo del seguimiento de objetos es generar la trayectoria de un objeto a través del tiempo mediante la localización de su posición en cada fotograma del video. Las tareas de detección y el establecer una correspondencia entre objetos se pueden realizar conjuntamente o por separado del seguidor.

Por otra parte el filtrado se puede descomponer esencialmente en dos pasos: predicción y actualización. En la etapa de predicción, el estado futuro es obtenido, por medio de ecuaciones sobre el estado actual, y en la etapa de actualización, las mediciones del estado actual son utilizadas para mejorar las predicciones futuras.

A modo de introducción y con el propósito de simplificar el seguimiento de objetos, se presentara primero el filtrado para un solo objetivo.

2.4.1. Filtrado para un solo objetivo

El problema más simple y representativo del seguimiento, se compone de un solo objeto en un área de interés, donde se estima recursivamente su estado, mediante el uso de la información procedente de un proceso de observación. Los objetos de interés se denominan generalmente como *targets*, y la zona de observación generalmente se llama región de interés. El proceso de observación puede consistir en datos procedentes de los radares, sonares, cámaras, satélites, micrófonos, etc.

Los retos del filtrado para el seguimiento de objetos son que la información procedente de los sensores no es perfecta: los objetivos pueden no ser detectados durante uno o más pasos de tiempo y las observaciones en problemas reales son observaciones ruidosas.

Filtro de Bayes

El filtro de Bayes es una recursión que consiste en dos pasos, predicción y actualización. A través del tiempo, la ecuación de Chapman-Kolgomorov se utiliza para propagar la densidad posterior.

La regla de Bayes, ecuación **??**, es usada para obtener la pdf posterior, la cual es proporcional al producto de la probabilidad de las mediciones dado el estado actual, $p(z_k|x_k)$ conocido como modelo de observaciones, y la probabilidad del estado actual dadas las mediciones anteriores, $p(x_k|Z_k{-}1)$. En general la ecuación **??** describe que tan probable es el estado x_k dada las mediciones obtenidas del sensor Z_k.

En el filtrado Bayesiano la dinámica del objetivo cumple la propiedad de Markov, no contiene "memoria". Es por ello que se necesita propagar la densidad anterior, $p(x_k|Z_{k-1})$, para poder ser utilizada en la ecuación **??**. Esto se logra mediante la ecuación de Chapman-Kolmogorov, **??** [**?**].

$$P(x_k|Z_{k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|Z_{k-1})d_{x_{k-1}} \tag{2.19}$$

La integral en **??** propaga la función de densidad, $p(x_{k-1}|Z_{k-1})$, definida en $k-1$, a través del modelo de movimiento $p(x_k|x_{k-1})$. Usando las ecuaciones **??** y **??** podemos expresar la densidad posterior, $p(x_k|Z_k)$, en función de la pdf anterior, el modelo de movimiento y el modelo de mediciones.

Por otra parte, el paso de actualización consiste en incorporar la información sobre las observaciones en el tiempo actual, ecuación **??**.

$$P(x_k|Z_k) = \frac{p(Z_k|x_k)p(x_k|Z_{k-1})}{p(Z_k|Z_{k-1})} \tag{2.20}$$

El filtro Bayesiano permite la construcción de la función de probabilidad de densidad (pdf) posterior en el instante k. Sin embargo, las ecuaciones son insolubles en la práctica, salvo en casos especiales y muy limitados. Por lo tanto, se han propuesto soluciones aproximadas basadas en técnicas de discretización espacial e integraciones numéricas. Sin embargo, debido a la complejidad del problema, estos enfoques son útiles sólo cuando la dimensión de los espacios de estado y de observaciones son relativamente bajos.

Filtro de Kalman

El filtro de Kalman, se describe en [**?**] como una solución recursiva para el problema del filtrado lineal de datos discretos, el cual pronostica un nuevo estado a partir de la estimación previa, añadiendo un término de corrección proporcional al error de predicción, de tal forma que esté último es minimizado. El filtro de Kalman ha sido objeto de una extensiva investigación y aplicación, particularmente en el área de la navegación autónoma, en rastreo de misiles y en economía.

Las ecuaciones que se actualizan en el tiempo son ecuaciones de predicción, mientras que las ecuaciones que incorporan nueva información pueden considerarse como ecuaciones de corrección. Así el filtro de Kalman funciona por medio de un mecanismo que pronostica el nuevo estado, su incertidumbre y la corrige.

Para la realización del filtro de Kalman se utilizan modelos estado-espacio. La representación estado-espacio de un sistema lineal captura la dinámica de un vector de observaciones Z_k de orden $n \times 1$ en términos de un posible vector de estados X_k de orden $m \times 1$. Un modelo puede reescribirse en términos de una ecuación de proceso o estado y de una ecuación de medida u observación.

En **??** y **??** se representa el modelo de movimiento y medición respectivamente, los cuales son una descripción alternativa en estado-espacio de los modelos de movimiento y mediciones definidos en la sección anterior.

$$x_k = F_{k-1} x_{k-1} + v_{k-1} \tag{2.21}$$

Donde:

x_k representa el vector de estado en el instante k.

F_{k-1} es la matriz de transición del vector de estado en $k-1$.

x_{k-1} representa el vector de estado al momento $k-1$.

v_{k-1} representa el error en el modelo de movimiento en el instante $k-1$.

$$z_k = H_k x_k + w_k \tag{2.22}$$

Donde:

z_k son las mediciones del proceso u observaciones derivadas.

H_k es la matriz que relaciona los vectores de estado del sistema con las medidas.

x_k representa el vector de estado en el instante k.

w_k es el error asociado a las observaciones.

La ecuación **??** indica que el nuevo vector de estado es modelado como una combinación lineal del vector de estado anterior y de algún proceso de error. Por otra parte, la ecuación **??** describe cómo las medidas u observaciones son derivadas desde los vectores de estado [**?**].

Como en el filtro Bayesiano, el filtro de Kalman opera recursivamente con una etapa de predicción y una etapa de corrección. La etapa de predicción está caracterizada por las ecuaciones siguientes:

$$x_{k|k-1} = F_{k-1} x_{k-1|k-1} \tag{2.23}$$

$$P_{k|k-1} = F_{k-1} P_{k-1|k-1} F_{k-1}^T + Q_{k-1} \tag{2.24}$$

Donde $x_{k|k-1}$ es el estado estimado, F_{k-1} la matriz del sistema, Q_{k-1} es la matriz de ruido y $P_{k|k-1}$ es la matriz de covarianza.

Por otra parte, la etapa de corrección está definida por las siguientes ecuaciones:

$$\widetilde{z}_k = z_k - H_k x_{k|k-1} \tag{2.25}$$

$$S_k = H_k P_{k|k-1} H_k^T + R_k \tag{2.26}$$

$$K_k = P_{k|k-1} H_k^T S_k^{-1} \tag{2.27}$$

$$x_{k|k} = x_{k|k-1} + K_k \widetilde{z}_k \tag{2.28}$$

$$P_{k|k} = (I - K_k H_k) P_{k|k-1} \tag{2.29}$$

El vector $\widetilde{z}_k$ denota la diferencia entre las observaciones o mediciones recibidas z_k y las mediciones predichas $H_k x_{k|k-1}$. S_k es la matriz de covarianza de $\widetilde{z}_k$, H_k es la matriz de mediciones y K_k es una matriz llamada ganancia de Kalman.

El filtro de Kalman puede operar solo en modelos lineales. Para resolver este problema se han propuesto varias extensiones como el *Extended Kalman Filter* (EKF) y *Unscented Kalman Filter* (UKF) [**?**].

Filtro de Partículas

El filtro de partículas supone una nueva manera de representar y generar recursivamente una aproximación del estado de la función de densidad de probabilidad (pdf). La idea central es representar la pdf como un conjunto de muestras aleatorias, más que como una función sobre el espacio de estados.

Fue propuesto en 1993 por N. Gordon, D. Salmond y A. Smith como una implementación Bayesiana recursiva. Básicamente, el filtro de partículas se compone de un conjunto de muestras (partículas) y unos valores, o pesos, asociados a cada una de esas muestras. Las partículas son estados posibles del proceso, que se pueden representar como puntos en el espacio de estados de dicho proceso. El Filtro de Partículas se compone de cuatro etapas principales:

1. Inicialización.

2. Actualización.

3. Estimación.

4. Predicción.

En concreto, el Filtro de Partículas representa la densidad a posterior mediante un conjunto discreto de N partículas $(m_1, ..., m_N)$y sus probabilidades asociadas $(\Pi_1, ..., \Pi_N)$ Inicialmente, el conjunto de partículas se escoge a partir de la distribución $p(x_0)$. Si no existe información a priori, entonces las partículas se distribuyen uniformemente por el espacio de estados.

Posteriormente, en cada instante de tiempo k, se actualizan las N partículas en función de la acción anterior a_{k-1} y la observación actual Z_k. Para ello, se aplica el modelo de movimiento $p(x_t|x_{k-1}, a_{k-1})$ a cada una de las N partículas, generando un nuevo conjunto de partículas. Las partículas nuevas representan la predicción de la variable de estado.

En un último paso, se remuestrea el conjunto de partículas, extrayendo N partículas del conjunto actual, proporcional al peso de cada una. En este nuevo conjunto, tendrán más probabilidad de desaparecer aquellas partículas para las que no hay evidencia de verosimilitud o, lo que es lo mismo, que tenga menor peso. Una vez construido el nuevo conjunto de partículas, según la probabilidad de éstas se asocia un peso a cada una. Al terminar la etapa de predicción, se obtiene un nuevo conjunto de puntos al que se le vuelve a aplicar la etapa de actualización, repitiendo hasta que terminen las mediciones o desaparezca el objeto, caso en el cual se volvería a la etapa de inicialización.

El algoritmo de actualización de forma general queda como sigue:

1. Propagar cada partícula m_i en el tiempo, utilizando el modelo de movimiento $p(x_t|x_{k-1}, a_{k-1})$ para obtener un conjunto actualizado de partículas.

2. Obtener un nuevo vector de medidas Z y evaluar la densidad de probabilidad posterior. Esto puede ser escrito usando la regla de Bayes **??**.

3. Volver a muestrear y generar un nuevo conjunto bien ponderado de partículas.

4. Repetir los pasos 1 y 3 para instantes sucesivos.

Diferencia entre el filtro de Kalman y el filtro de Partículas

El Filtro de Kalman define un conjunto de ecuaciones matemáticas que proveen una solución recursiva. La potencia del filtro se debe a varios aspectos: soporta estimaciones del pasado, presente e incluso estados futuros, y puede utilizarse cuando la naturaleza precisa del sistema a modelar es desconocida.

El Filtro de Partículas representa una poderosa herramienta para el tratamiento de procesos del mundo real, evitando hacer cualquier tipo de suposiciones sobre las características del proceso. La diferencia fundamental entre ambos algoritmos está en el hecho de que Kalman pueden representar solamente la estimación del estado por una Gaussiana uni-modal, mientras los Filtros de Partículas pueden representar densidades multimodales complejas empleando una gran cantidad de partículas aleatoriamente muestreadas, debido a esto son más precisos pero mucho más complejos y requieren más recursos tanto tecnológicos como temporales.

2.4.2. Filtrado para múltiples objetivos

Ha habido una considerable actividad de investigación en el seguimiento de múltiples objetivos en los últimos 20 años. Este interés está motivado por numerosas aplicaciones y el creciente desarrollo tecnológico.

Los enfoques tradicionales de filtrado para múltiples objetos tratan el problema dividiendo la asociación de la estimación, usando técnicas de filtrado para cada objetivo, por lo tanto se pierde efectividad en el proceso, y aumenta la complejidad.

En la última década se han desarrollado algoritmos de filtrado multiobjeto; uno de ellos es el filtro Probability Hypothesis Density (PHD), el cual emula

en cierta medida la sencillez del enfoque de filtro de Kalman para resolver el problema multiobjeto.

En esta sección se abordará el problema del seguimiento de múltiples objetos, sus bases teóricas y el filtro Probability Hypothesis Density.

Filtro *Probability Hypothesis Density* (PHD)

El filtro PHD es una aproximación al filtro Bayesiano para múltiples objetivos. Se compone de una etapa de predicción y una etapa de actualización, que propagan de forma recursiva el primer momento en forma de un conjunto finito aleatorio (RFS). En PHD se estima el número de objetivos, así como sus estados. La recursividad en PHD se basa en un conjunto de hipótesis [**?**]:

- Cada objetivo puede evolucionar y generar mediciones independientemente.
- El nacimiento de objetivos se supone que está distribuido independiente por Poisson.
- Los objetivos previstos y posteriores se aproximan por Poisson.

Denotando por $\Upsilon_t(x)$ y $\Upsilon_{t|t-1}(x)$ las funciones de intensidad asociados a los estados multiobjeto, la intensidad predicha está dada por:

$$\Upsilon_{t|t-1}(x) = \int_{E_s} [P_{s,t}(u) f_{t|t-1}(x|u) + b_{t|t-1}(x|u)] \Upsilon_{t-1|t-1}(u) du + \mu_t(x) \tag{2.30}$$

Donde $f_{t|t-1}(x|u)$ es la evolución de densidad de un objetivo en el tiempo t, $P_{s,t}(u)$ es la probabilidad de supervivencia de un estado u, $\mu_t(x)$ es la función de intensidad de nacimiento de nuevos objetivos en tiempo t. Por otra parte, la intensidad actualizada o posterior se obtiene mediante:

$$\Upsilon_{t|t}(x) = (1 - P_{d,t}(x)) \Upsilon_{t|t-1}(x) + \sum_{y \in \Upsilon_i} \frac{P_{d,t}(x) g_t(y|x) \Upsilon_{t|t-1}(x)}{h_t(y) + \int P_{d,t}(u) g_t(y|u) \Upsilon_{t|t-1}(u) du} \tag{2.31}$$

Donde $g_t(y|x)$ denota la función de similitud, $P_{d,t}(x)$ la probabilidad de detección en el tiempo t y h_t es la intensidad del ruido en el tiempo t.

Las ventajas y debilidades del filtro PHD se mencionan a continuación:
Ventajas:

- La complejidad computacional es $O(mn)$, donde n es el número de objetivos y m el número de observaciones [**?**].
- Puede resolver acontecimientos de falsas detecciones y falsas alarmas
- Su implementación es flexible y adaptable (varias técnicas existen).
- Se estima el número de objetivos en cada instante de tiempo, pero no se propaga.

Desventajas:

- Puede existir una alta variación en la estimación del número de objetivos entre cortos periodos de muestras.
- Mucha información se pierde al propagar el primer momento.
- No se desempeña bien en situaciones de bajo SNR.
- PHD está basado en la suposición que los objetivos están alejados entre sí, cuando los objetivos están demasiado cerca no se genera medición y un solo seguidor es generado.
- No tiene una asociación de los datos implícita, solo estima el estado del objeto en cada instante de tiempo.

Filtro *Gaussian Mixture-Probability Hypotheis Density* (GM-PHD)

Para un conjunto de problemas de seguimiento multiobjetivo, existe una solución de forma cerrada, GM-PHD. El filtro GM-PHD, es una aproximación mediante mezclas de Gaussianas de PHD. Para la implementación de este filtro, varias supuestos son necesarios:

1. Cada objetivo tiene un modelo dinámico lineal donde la función de transición del estado y las observaciones son Gaussianas:

$$f_{k|k-1}(x_k|x_{k+1}) = \mathcal{N}(x_k; F_k x_{k-1}, Q_k) \tag{2.32}$$

$$g_k(z_k|x_k) = \mathcal{N}(z_k; H_k x_k, R_k) \tag{2.33}$$

2. Las probabilidades de detección y supervivencia son independientes del estado de los objetivos:

$$P_{sv,k}(x_{k-1}) = P_{sv,k} \tag{2.34}$$

$$P_{d,k}(x_k) = P_{d,k} \tag{2.35}$$

3. La intensidad de nacimiento puede ser representada por:

$$\gamma_k(x_k) = \sum_{i=1}^{J\gamma} w_{\gamma,k}^{(i)} \mathcal{N}(x_k; m_{\gamma,k}^{(i)}, P_{\gamma,k}^{(i)}) \tag{2.36}$$

Donde $m_{\gamma,k}^{(i)}$ es la media y $P_{\gamma,k}^{(i)}$ la varianza de la Gaussiana.

El filtro GM-PHD sigue los mismos pasos que el filtro PHD, sin embargo son modelados mediante mezclas de Gaussianas:

1. **Inicialización:** El algoritmo recursivo es inicializado en $k = 0$, con una intensidad inicial v_0, la cual es la suma de todos los componentes de las Gaussianas en $k = 0$. Matemáticamente esta inicialización se presenta en la siguiente ecuación:

$$v_0(x) = \sum_{i=0}^{J_0} w_0^{(i)} \mathcal{N}(x_0; m_0^{(i)}, P_0^{(i)}) \tag{2.37}$$

Donde $m_0^{(i)}$ es el vector de medias de los estados, $P_0^{(i)}$ es la matriz de covarianza y $w_0^(i)$ es el peso de cada componente Gaussiano.

2. **Predicción:** Para cada momento, el filtro calcula la intensidad $v_{(k|k-1)}(x)$ en el tiempo k, la cual es una mezcla de Gaussianas. Utiliza $v_{k-1}(x)$ y sus componentes.

$$v_{(k|k-1)}(x) = \sum_{i=0}^{J_{(k|k-1)}} w_{(k|k-1)}^{(i)} \mathcal{N}(x_k; m_{(k|k-1)}^{(i)}, P_{(k|k-1)}^{(i)}) \tag{2.38}$$

Vale la pena recalcar que el algoritmo en este paso hace uso de la información sobre la media y la matriz de transición de un estado anterior, para calcular la media predicha.

3. **Actualización:** El filtro colecta información del proceso de observación para validar actualizar la predicción y calcular la intensidad actual $v_{k|k}(x_k)$. La cual también está modelada por una mezcla de Gaussianas.

4. **Estimación de estados:** Los estados de los objetivos son determinados por medio de la media de las Gaussianas, umbralizando los pesos.

$$X_k = \{\tilde{m}_k^{(i)}\} \tag{2.39}$$

Donde X_k es el conjunto de estados estimados y $\{\tilde{m}_k^{(i)}\}$ es la media de las Gaussianas umbralizadas.

2.5. Asociación de datos

En esta sección, se discute el enfoque tradicional de la asociación de datos para el seguimiento de objeto. La asociación de datos permite asignar a cada objetivo una etiqueta de identidad. En primer lugar, se presenta una introducción al problema, y luego se realiza un análisis de los principales algoritmos de asociación de datos.

2.5.1. Introducción

En el seguimiento de objetos, suele haber varias mediciones desde el sensor, los objetivos y el ruido. La asociación de datos trata el problema de la selección de la medición más probable que se origina desde un objeto. Si se selecciona la medición incorrecta, o si la observación no se detecta en absoluto, las estimaciones de los estados podrían comprometerse.

Un problema común en la asociación de datos es que el número de objetos a ser rastreados podría exceder de un modo a las mediciones obtenidas, debido a oclusión o fallo del sensor. Este problema se vuelve más complejo cuando el número de objetos es desconocido y no constante. El ejemplo más simple de asociación de datos se muestra en la figura **??**, donde existen más de una medición y un solo objeto.

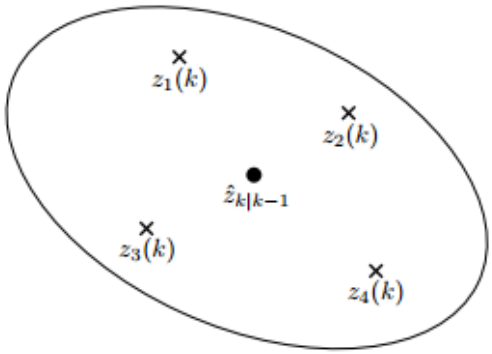


Figura 2.6: Asociación de datos con un solo objeto

Por otro parte, nuevos objetos pueden aparecer y desaparecer en cualquier momento y en cualquier lugar lo que significa que las mediciones que no tienen una asociación pueden ser relevantes. Un algoritmo de asociación de datos tiene que determinar si una observación es correcta o incorrecta y si un seguidor tiene que ser inicializado o un objeto existente continúa.

En el caso más complejo, se toman todas las asignaciones como hipótesis, y se elige una de acuerdo a una métrica. La figura **??** muestra un ejemplo donde en cada instante de tiempo pueden existir muchas observaciones (asignaciones) para los objetivos. En las siguientes secciones se describirán las principales características de los algoritmos que permiten combatir este tipo de problemas.

2.5.2. Vecino más cercano

El algoritmo del vecino más cercano es probablemente el método más sencillo para resolver el problema de asociación de datos. Muchas variantes de este método existen: el vecino más cercano probabilístico, vecino más cercano subóptimo y el vecino más cercano global (GNN). Algunos detalles sobre estos algoritmos se pueden encontrar en los trabajos de Bar-Shalom y Fortmann [1988].

La idea de este método es seleccionar la medición que sea más cercana a la predicha por el filtro, es decir, minimizar la distancia entre los puntos previstos

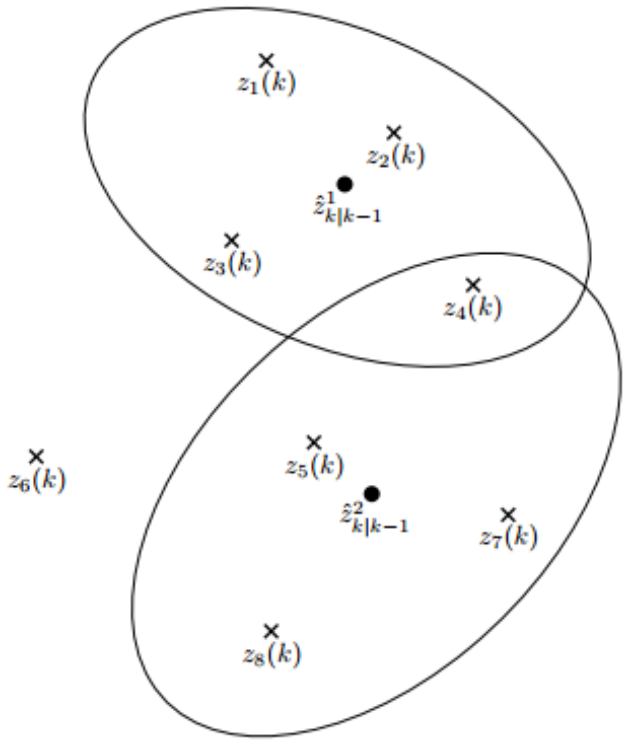


Figura 2.7: Asociación de datos con más de un objeto

y medidos. Con el fin de hacerlo, para cada una de las mediciones se utiliza la Distancia de Mahalanobis **??** [**?**].

$$d^2 = \tilde{z}_k^T S_k^{-1} \tilde{z}_k \tag{2.40}$$

Este filtro se puede aplicar para el seguimiento de cualquier número de objetos conocidos y una ventaja es la complejidad computacional baja. Un inconveniente evidente es que, no siempre el vecino más cercano es la observación correcta. Este algoritmo no es capaz de hacer frente a la aparición o desaparición de objetos. Como resultado, sólo se puede utilizar combinado con modelos de nacimiento y muerte de objetivos.

2.5.3. Asociación de datos probabilista (PDA)

El filtro de asociación de datos probabilista (PDA) es un algoritmo óptimo que utiliza las probabilidades de asociación de las últimas mediciones. En la que todos los candidatos potenciales para la asociación se combinan en una

sola actualización estadística más probable, teniendo en cuenta la distribución estadística de los errores del objetivo y el ruido, y suponiendo que sólo uno de los candidatos es un objetivo, y el resto son falsas alarmas. Una explicación detallada de este filtro se puede encontrar en [**?**].

Esta técnica se utiliza comúnmente para el seguimiento de un solo objetivo en ambientes ruidosos y muestra un rendimiento razonable. PDA, sin embargo, como el método vecino más cercano, asume algún conocimiento previo de la existencia del objetivo.

2.5.4. Asociación de datos probabilista conjunta (JPDA)

JPDA es una extensión de PDA, pero se diferencia en el sentido de que permite la asociación de múltiples objetivos en entornos de ruido. Concretamente, la diferencia es que las probabilidades se calculan de forma conjunta.

Como en PDA, sólo la última medición se utiliza y cada objetivo se supone que tiene su propio modelo de movimiento y medición. Por otra parte, se supone que el número de objetivos tiene que ser conocido. Se utiliza generalmente junto con el filtro de Kalman, porque en sus ecuaciones se calcula la probabilidad conjunta.

Una desventaja obvia de este método es que sólo tiene en cuenta los modelos lineales, ya que utiliza Kalman. Por otro lado, el número de objetivos se tiene que conocer y debe ser constante. A pesar de estos inconvenientes, JPDA ha demostrado su valor en diversos ejemplos y fue el principal algoritmo de asociación de datos por más de una década. Además, este filtro es capaz de tratar con múltiples sensores. Sin embargo, tanto PDA y JPDA no tienen la capacidad para recuperarse de errores causados por asociaciones erróneas en estados anteriores.

2.5.5. Seguimiento por múltiples hipótesis (MHT)

Es una técnica muy utilizada, que mantiene diferentes hipótesis de asociación entre los objetivos y las observaciones, y permite a las mediciones futuras

resolver la incertidumbre. El filtro MHT funciona mediante la realización de una búsqueda exhaustiva sobre *frames* anteriores, obteniendo todas las posibles combinaciones de asociación entre los objetivos y las mediciones.

MHT analiza varias posibilidades calculando sus probabilidades posteriores usando la regla de Bayes. Luego, una medida se asocia a ruido o a un objetivo activo, el conjunto de hipótesis con mayor probabilidad posterior se mantiene y se propaga mediante el uso del filtro de Kalman [**?**]. Como es evidente, es necesario adoptar técnicas para evitar que el número de hipótesis crezca exponencialmente en el tiempo y por esta razón el método resulta exigente.

El mayor inconveniente de MHT es de hecho su naturaleza combinatoria; implementaciones prácticas utilizan distancias de validación para limitar el número de cálculo, y la poda heurística para eliminar las hipótesis de baja probabilidad. Se han propuesto variantes de MHT, como MHT probabilístico (PMHT) que utiliza asociaciones probabilísticas de medición [**?**]. Sin embargo todavía no es práctico en aplicaciones donde el número de objetivos es grande.

2.5.6. Algoritmo Hungaro

El algoritmo Húngaro resuelve el problema de asignación con una complejidad $O(n^3)$. La idea fundamental de resolución responde la siguiente pregunta, ¿Qué estado satisface mejor al objetivo?. Una característica particular del modelo de asignación es que para su resolución no se hace necesario que el número de fuentes sea igual al número de destinos, lo cual es muy común en problemas de seguimiento de objetos.

En general el método Húngaro consiste en generar una matriz de costos $n * m$, que contendrá la distancia Mahalanobis entre los estados generados en k y $k-1$. Posteriormente se iterará para encontrar el costo mínimo de asignaciones en la matriz, para ello se remueven las filas y columnas donde el costo de asignación no minimice la matriz.

Se comprobará si la asignación cumple las siguientes validaciones:

- Los seguidores que no tengan asociado un objetivo no pueden utilizarse para la eliminación de filas o columnas.
- Las mediciones que no estén asociadas se pueden utilizar para generar un seguidor.
- En caso de ambigüedad se ocupará la velocidad y dirección de los objetivos, para resolver la asignación, vea sección **??**.

Para profundizar en los pasos del algoritmo Húngaro para el problema de asignación, se puede estudiar los trabajos [**?**] [**?**] [**?**]

Parte III

Trabajo relacionado

Capítulo 3

Trabajo relacionado

Esta capítulo se enfoca a la revisión de los trabajos más importantes del seguimiento múltiple. Específicamente se enfatizarán los métodos de seguimiento visual basados en PHD. El presente trabajo de tesis se centra en dos problemas principales: el manejo de la oclusión en GM-PHD y la asignación de etiquetas en el filtrado basado en PHD, por ello el trabajo relacionado está enfocado en estos problemas.

3.1. Introducción

Un primer enfoque para el filtrado de múltiples objetivos es considerar cada objetivo por separado y seguirlo con un filtro individual. Sin embargo, esto requiere la correcta asociación de los objetivos individuales con su medición, elevando la complejidad del problema. Algunos algoritmos, [**?**] no utilizan filtros estrictamente, más bien mediante características de los objetos, y con ayuda del detector determinan la posición del objeto a través de los diversos *frames*.

Los primeros intentos de conocer el estado de múltiples objetivos mediante un solo marco o estructura de datos, se construyó a partir de vectores en [**?**], marcando una pauta para el desarrollo de nuevos algoritmos de seguimiento multiobjetivo. La representación por medio de vectores, introdujo las virtudes de plantear el problema SMO, desde un punto de vista con múltiples objetos. Se desarrollaron diversos trabajos con base en la utilización de vectores, pero siempre con problemas inherentes a su representación, ya que solo se podía

seguir un número acotado de objetivos y con conocimiento a priori de sus estados iniciales.

Fue en [**?**], donde surgieron los primeros intentos por modelar la colección de objetivos individuales y mediciones como conjuntos finitos aleatorios (RFS). Esta estrategia, sin embargo no mantenía ningún registro de la identidad del objetivo, evitando la asociación de datos e imposibilitando el etiquetado de objetivos. Con la incorporación de los conjuntos finitos aleatorios, se planteó una representación natural para el problema SMO. El primer método de filtrado de múltiples objetos, usando conjuntos aleatorios finitos fue desarrollado por Mahler en [**?**]. Mahler introdujo la teoría para permitir al filtro Bayesiano incorporar el seguimiento de múltiples objetos. Los trabajos presentados por Mahler se convertirían en las principales bases para las estadísticas de los conjuntos finitos aleatorios (FISST) [**?**].

El primer método practico basado en FISST, se obtuvo en 2001 cuando Mahler publica [**?**], trabajo fundamental que lleva la teoría de FISST al *framework* Bayesiano. Con base en una idea novedosa, Mahler implementa la propagación del primer momento de la función de densidad (PHD), como característica suficiente para la realización del seguimiento múltiple. El trabajo de Mahler culmina posteriormente en la construcción del filtro PHD [**?**]. El éxito del filtro en el procesamiento de señales y seguimiento de objetos fue abrumador, promoviendo la investigación en el filtro alrededor del mundo.

El filtro PHD es hoy en día una alternativa simple y eficaz para el filtrado múltiple. Sin embargo, existe la necesidad de evaluar integrales que no tienen forma cerrada, en general, esto hace que las implementaciones del filtro PHD sean extremadamente difíciles. Por ello muchas aproximaciones se han desarrollado [**?**] [**?**] [**?**].

3.2. *Sequential Monte Carlo Probability Hypothesis Density* (SMC-PHD)

Para poder llevar a la práctica el filtro PHD, esté se implementó mediante el uso de métodos secuenciales de Monte Carlo. En [**?**], Vo, Singh, Doucet, y Clark, desarrollaron *Sequential Monte Carlo Probability Hypothesis Density* (SMC-PHD) o *Particle Filter Probability Hypothesis Density* (PF-PHD). Vo, Singh, Doucet, y Clark también establecieron resultados de convergencia para PF-PHD en [**?**].

La principal aportación del trabajo de Vo, Singh y Doucet fue establecer una relación entre el conjunto de derivadas de FISST y la densidad de probabilidad (PHD), modelando a través de partículas los múltiples estados de los objetivos. Diversas implementaciones surgieron a partir de su trabajo [**?**] [**?**] [**?**].

El principal problema de esta implementación, misma que sufre el filtro de Partículas, es la complejidad de modelar las diferentes partículas del conjunto de estados y sus pesos, limitándolo a problemas con un número reducido de objetos. Para resolver estos problemas en [**?**] desarrollaron un enfoque de SMC-PHD, capaz de realizar el seguimiento de múltiples objetivos limitando el número de partículas, y lo que es más importante estimando un número desconocido de objetivos. Este método se llama *Box-Particle PHD Filter*, el cual reduce el número de partículas utilizando una muestra aleatoria de una región rectangular, llamada también caja de partículas (*Box-particle*). Con esta estructura Schikora et al. disminuyeron la dimensionalidad del problema, ocupando información espacial para submuestrear las partículas.

3.2.1. SMC-PHD manejo de oclusión y asignación de etiquetas

Maggio et al. en [**?**] desarrollaron un algoritmo basado en SMC-PHD, el cual resuelve parcialmente el problema de la asociación de datos. En este trabajo se incorpora un mecanismo de remuestreo para limitar el número de partículas que se generará por cada objetivo, la cantidad de partículas remuestreadas estará relacionada con el peso de las partículas asociadas al objetivo en el

pasado, $k-1$. Por otro lado para proporcionar información de la etiqueta, el autor utiliza la información del color para asociar los resultados actuales con los resultados anteriores. Se utilizó un procedimiento de asociación de datos basado en grafos, el cual es un módulo independiente del método.

En [**?**] se manejó el problema de la asignación de etiquetas en videos de personas. En este trabajo se fusionaron dos detectores para permitir al algoritmo SMC-PHD decidir si los objetivos son de interés. Un detector proporciona información de nuevas personas que se incorporan en la escena, mientras que el segundo, recolecta propiedades y las asigna a cada persona individual para identificarlas. Sobre la base de las observaciones híbridas, se realizó un mecanismo para extraer múltiples estados con una etiqueta generada por el segundo detector.

En [**?**] se propone añadir una etiqueta a cada partícula, la cual registrará la identidad del destino y puede ser heredada. El proceso de evolución de las partículas y la estimación del estado para un objetivo individual se puede obtener mediante el uso de partículas con una misma etiqueta, generalmente se utilizan técnicas de agrupamiento para identificar la posición del objetivo. Uno de los problemas principales de esta técnica es que requiere que los objetivos estén bien separados, no maneja la oclusión.

[**?**] propuso dos métodos para la identificación de objetivos, cuya complejidad es menor que los filtros MHT y JPDA, en su trabajo se realiza un agrupamiento de partículas asociadas a cada objetivo, la cual se propaga al estado siguiente.

[**?**] describe una técnica para la asociación de los objetivos para el filtro SMC-PHD basado en la coincidencia entre picos de los pesos de las partículas, usando técnicas de optimización para reducir al mínimo el costo de asociación.

3.3. Filtro *Gaussian Mixture-Probability Hypotheis Density* (GM-PHD)

[?] [?] [?] propusieron mediante el uso de mezclas gaussianas un filtro basado en PHD. Esta implementación, bajo ciertos supuestos simplificadores, mejora en gran medida la eficiencia computacional del filtro, permitiendo su aplicación en escenarios reales.

En el último lustro GM-PHD ha sido ampliamente utilizado como la principal aproximación al filtro Bayesiano multiobjetivo. Las primeras implementaciones de GM-PHD fueron propuestas por Vo y Ma en su trabajo [?]. Concluyendo que aunque el filtro es menos robusto al seguir objetivos no lineales, en comparación con SMC-PHD, pero proporciona una complejidad computacional mucho menor.

[?], demostraron una fuerte convergencia L1 uniforme para el filtro GM-PHD. Concluyendo que proporciona una solución de forma cerrada al problema de filtrado Bayesiano para múltiples objetivos.

[?], presentaron un algoritmo del filtro de GM-PHD para el seguimiento de múltiples objetivos en un escenario con alta densidad de ruido. El algoritmo fue capaz de estimar el número de objetivos, seguir sus trayectorias a través del tiempo, operar con detecciones pérdidas y encontrar las trayectorias en el pasado, una vez que un objetivo fue identificado.

Esto fue una mejora de la implementación inicial del filtro de GM-PHD que proporcionaba estimaciones para el conjunto de estados en cada punto en el tiempo, pero no garantizaba la continuidad del seguidor. En [?] y [?] se demostró que las trayectorias de los objetivos se pueden determinar directamente a partir de la mezcla de Gaussianas.

En [?] se desarrolló una evaluación y comparación entre GM-PHD y MHT. Demostrando que GM-PHD supera a MHT en la capacidad para operar en

regiones de alta densidad de ruido, reduciendo su complejidad significativamente. Sin embrago, el algoritmo de GM-PHD por sí solo no maneja la asociación de datos, algo que si realiza MHT con una complejidad grande.

Las aplicaciones prácticas de este método han incluido el seguimiento de vehículos en diferentes terrenos y el seguimiento de objetivos con radares y sonares [**?**] [**?**] [**?**]. Investigadores de la Agencia de Investigación de Defensa de Suecia han utilizado GM-PHD para el seguimiento de personas en videos digitales.

3.3.1. GM-PHD manejo de oclusión y asignación de etiquetas

Dado que el filtro GM-PHD no proporciona métodos explícitos de asociación de datos y etiquetado de objetivos, muchos esfuerzos se han dedicado a la investigación y desarrollo de la extracción de los objetivos eficientemente para realizar un seguimiento con técnicas de asociación.

Panta , Vo , y Singh [**?**] propuso dos arquitecturas diferentes del filtro PHD, que proporcionan estimaciones del seguidor con valores de objetivos individuales, permitiendo realizar un etiquetado basado en alguna técnica de asignación. Se evaluó su método con la propagación de múltiples hipótesis del algoritmo MHT, obteniendo resultados similares con una complejidad menor. Wang , Jing y Hu [**?**] también propusieron una técnica de asociación de datos basada en MHT, pero utilizando el filtro de SMC-PHD para eliminar medidas inverosímiles.

Wu et al. [**?**] propone una arquitectura similar a la desarrollada en [**?**] y [**?**]. Wu et al. utiliza un detector basado en el modelado del fondo junto con el algoritmo GM-PHD para reducir el ruido y estimar el estado de los objetos. Posteriormente para seguir a objetivos que aparecen en la escena, se juntan entre sí, se separan o desaparecen de la escena; incorpora un método de asignación de etiquetas, el algoritmo de subasta (similar al algoritmo Húngaro). El principal problema de [**?**], es que en el método que implementó no se contempla ni la caracterización de la oclusión ni mecanismos para garantizar la continuidad del seguidor.

Por otra parte PHD tampoco cuenta con un manejador de oclusiones. Esta característica del filtro no se ha estudiado minuciosamente, existiendo pocas referencias de adaptaciones de GM-PHD para el manejo de oclusión. La baja precisión del algoritmo GM-PHD, comparado a SMC-PHD, frente a objetivos que sigan movimientos no lineales, limito por varios años su utilización al seguimiento simple de objetos.

Los primeros esfuerzos para armar al algoritmo GM-PHD frente al problema de oclusión se publicaron en 2012, donde [**?**] definen la suposición *one-to-one*, que permite caracterizar a un objetivo en un estado de oclusión mediante dos proposiciones:

1. Una observación es generada desde un objetivo.

2. Un objetivo puede estar asociado a una medición.

La suposición *one-to-one* se basa en la identificación de la oclusión cuando el peso de algún objetivo sobrepasa el máximo, 1. La oclusión se maneja con una renormalización de los pesos de los estados, permitiendo así al seguidor permanecer en la escena.

Por otra parte, [**?**] proponen un método basado en GM-PHD que estima la probabilidad de oclusión para cada objetivo de forma dinámica e incorpora esta información para modificar su probabilidad de detección. Para ello se utiliza una distancia de oclusión entre los objetivos y dependiendo su valor se resuelve la probabilidad de detección.

En 2013 [**?**], presentan una implementación de GM-PHD con un manejador de oclusión basado en teoría de juegos. Este método caracteriza la oclusión con base en la información espacial, delimitando regiones. Entonces, un algoritmo no cooperativo basado en teoría de juegos es aplicado, modelando las regiones de oclusión como jugadores que compiten por maximizar las utilidades espaciales, con base en una estrategia definida por el ámbito del problema.

Por otra parte, con el fin de contrarrestar los problemas de oclusión, se evaluaron diferentes funciones de densidad de probabilidad, para los modelos de nacimiento del objetivo [**?**] [**?**]. Con ello se proporcionó una fuente de referencia para caracterizar al filtro GM-PHD. [**?**] utilizaron mezclas de Gaussianas, y estimaron automáticamente las intensidades a través del tiempo. Esto sirvió como referencia en muchos otros trabajos, y ahora se incorpora como parte fundamental del filtro.

3.4. Discusión

Las implementaciones de los filtros PHD con manejo de oclusión y etiquetado de objetivos son recientes, aún no se tiene caracterizado un algoritmo que sea lo suficientemente robusto para poder usarse como referencia teórica.

El problema de la oclusión se resuelve desde dos perspectivas generales, con mecanismos externos al filtro PHD o incorporando mecanismos en sus ecuaciones que permitan mantener al seguidor asociado con un objetivo en estado de oclusión. Por otra parte, los métodos de asignación de etiquetas o asociación de datos se desarrollan como módulos independientes del algoritmo PHD.

A continuación y utilizando la tabla comparativa **??** de los principales trabajos relacionados. Se exploran los pros y contras de cada algoritmo representativo:

1. Este algoritmo utiliza un módulo independiente para la asociación de datos, y puede recuperarse de oclusiones que duran un corto periodo de tiempo. Sin embargo, utiliza información del color de los objetivos para asociarlos de un *frame* a otro, por lo tanto no puede manejar la oclusión de objetivos de la misma clase.

2. El etiquetado es robusto porque utiliza información de dos detectores, uno basado en características globales de la escena y el otro en características particulares del objeto. Este algoritmo maneja la oclusión parcial, sin utilizar mecanismos en el filtro PHD, lo que hace que el algoritmo dependa totalmente de los detectores.

3. Aunque la idea general de este algoritmo es reducir el número de partículas, utiliza una etiqueta en cada partícula elevando la complejidad. Este método no soporta la oclusión de ningún tipo.

4. En este trabajo se implementa por primera vez un método donde el mecanismo de asociación es independiente, parte de la idea original que el filtro PHD reduce el ruido y falsas detecciones para permitir a la asociación trabajar en espacios de menor dimensionalidad. Este método no contempla ni oclusión, ni etiquetado.

5. Es uno de los pocos métodos que ataca el problema de la oclusión como tema central, utiliza un método que caracteriza la oclusión mediante el peso de los seguidores. El problema principal de este método es que no identifica todas las oclusiones y puede generar falsas alarmas sí existe mucho ruido alrededor del objetivo.

6. Primer trabajo que caracteriza la oclusión en los filtros PHD mediante regiones de interés, es sencillo de implementar. Este método solo toma en cuenta la oclusión entre objetivos y falla en oclusiones largas, no se ha evaluado en problemas reales.

7. Se añade al filtro GM-PHD un método de asignación de etiquetas, no se especifica ningún mecanismo para manejar la oclusión.

Tabla 3.1: Tabla comparativa de los principales algoritmos relacionados

Algoritmo base	Características	Oclusión/Etiquetado
MHT	Múltiples hipótesis, recursión Bayesiana	SI/SI
SCM-PHD [1]	Mecanismo de remuestreo	NO/SI
SCM-PHD [2]	Fusión de 2 detectores	SI/SI
SCM-PHD [3]	Cada partícula tiene una etiqueta	NO/SI
GM-PHD [4]	Método de asociación independiente	NO/SI
GM-PHD [5]	Renormalización de los estados	SI/NO
GM-PHD [6]	Probabilidad de detección dinámica	SI/NO
GM-PHD [7]	Algoritmo de subasta	NO/SI
GM-PHD *	Intensidad de nacimiento, supervivencia y detección adaptable	SI/SI

[1] [?], [2] [?], [3] [?], [4] [?], [5] [?], [6] [?], [7] [?], * Método propuesto

Parte IV

Diseño

Capítulo 4

Diseño de GM-PHD adaptado

4.1. Introducción

El método descrito en esta sección es una adaptación del algoritmo GM-PHD para manejar el problema de la oclusión y asignación de etiquetas en el seguimiento de múltiples objetos. Para ello, se diseñaron dos mecanismos que constituyen un nuevo método. Estos mecanismos permiten soportar problemas de oclusión y asignación de etiquetas a objetivos en escenarios tanto sintéticos como reales.

En el algoritmo GM-PHD, no se consideran las asociaciones entre objetivos, por lo que la identidad del objetivo es desconocida. Por otra parte, en GM-PHD existe una observación asociada a un seguidor siempre y cuando el objetivo sea detectado. Sin embargo, cuando existe oclusión, los objetivos involucrados se fusionan y simplemente desaparecen de la escena. Por lo tanto, el rendimiento del filtro GM-PHD en estados de oclusión decrece, propagando errores en pasos posteriores a través de su recursividad.

Como primer paso en el método propuesto, se caracteriza el estado de un objetivo en oclusión. Posteriormente se incorporan mecanismos en el algoritmo GM-PHD, que adaptan las probabilidades de detección, la intensidad de supervivencia y nacimiento. En este enfoque se identifican 3 tipos de oclusión:

- Cuando uno o más objetivos de interés se ocluyen entre sí, parcial o totalmente dentro de la escena, figura **??**.

(a) Oclusión entre objetivos.

(b) Oclusión por ruido.

(c) Oclusión por objeto en la escena.

Figura 4.1: Tipos de oclusión.

- Cuando uno o más objetivos de interés son ocluidos por el ruido en la escena, figura **??**.
- Cuando un objeto en la escena, que es considerado fondo, ocluye a uno o más objetivos de interés, figura **??**.

En el método propuesto para el manejo de la oclusión se adaptan a través del tiempo las probabilidades de nacimiento, aparición espontánea y detección de los objetivos. Con el propósito de conservar activos los seguidores asociados a los objetivos ocluidos.

Por otra parte, para identificar a cada objeto a través de los diferentes frames se emplea un mecanismo de asignación de etiquetas. Este método es independiente del algoritmo GM-PHD, y utiliza la información procesada del estado anterior y actual de GM-PHD. A continuación se describe el método propuesto, analizando cada uno de sus componentes. En primer lugar se presenta la arquitectura del método.

Cabe aclarar que en este trabajo se realizó tanto el método de detección como la extracción de características, sin embargo la contribución principal

está directamente relacionada con el seguidor. Todas las adaptaciones, con excepción del módulo de asignación de etiquetas, se desarrollaron dentro del algoritmo PHD, con el fin de que este no dependa del detector.

4.1.1. Arquitectura del método

La arquitectura del método propuesto consta de 5 etapas principales. En la figura **??** se muestra la relación y secuencia de estas etapas. A continuación se describen de manera general cada etapa junto con su función.

El método recibe como entrada un conjunto de observaciones en el tiempo k, Z_k, obtenidas de la detección de objetos en un video o mediante datos sintéticos. El parámetro de salida del método en el tiempo k, es un conjunto de estados de los objetivos etiquetados X_k.

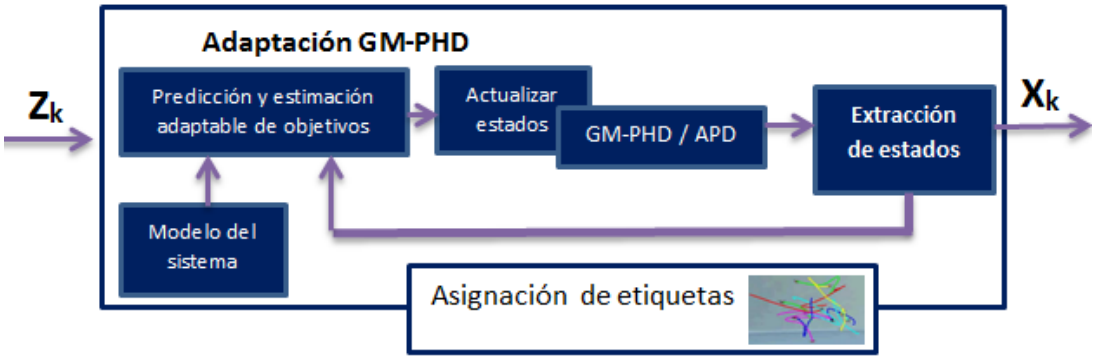


Figura 4.2: Arquitectura del método propuesto

En el modelo del sistema se inicializan los parámetros del algoritmo GM-PHD: el tiempo de muestreo, la dinámica de los objetivos, el modelo de mediciones, las probabilidades de detección y supervivencia; y la distribución de nacimiento de objetos.

En la predicción y estimación adaptable de los objetivos, se busca propagar las condiciones del escenario modificando las distribuciones de supervivencia y nacimiento mediante información de los objetivos en la escena.

En la etapa de actualización de estados, se garantizará la propagación de estados con objetivos ocluidos. Se adapta la probabilidad de detección para cada objetivo y se propaga a través de su recursividad.

La extracción de estados consta de una poda de objetivos con una probabilidad de detección baja. En este paso se diseñaron mecanismos para podar estados de objetivos que se encuentren en estado de oclusión.

La etapa de asociación de datos se presenta como un módulo independiente de GM-PHD. El cual sirve para asignar etiquetas a los objetivos, añadiendo capacidad al filtro para identificar objetos de la misma clase. Se eligió el algoritmo Húngaro de asignación, el cual trabaja con los estados extraídos y la asociación anterior.

4.1.2. Caracterización de la oclusión

La oclusión es un fenómeno o proceso que surge en el escenario o campo de visión de la cámara, cuando un objetivo de interés es obstruido por otro objeto. Intuitivamente un ser humano puede prever la oclusión con base en un análisis de la dirección, velocidad, tipo de objeto, etc. Sin embargo, emular en un programa esté proceso, es complejo.

Una paso importante para el método propuesto y así mismo para coadyuvar a la asignación de etiquetas, es la caracterización de los objetivos en estado de oclusión, o lo que es lo mismo, determinar precisamente cuando un objetivo se encuentra en un estado ocluido.

En la mayoría de trabajos relacionados, la caracterización de la oclusión está fuertemente ligada al proceso de detección de objetos, sin embargo, esto crea una dependencia hacia una sola clase de detectores. Por otro lado, realizar la caracterización de la oclusión independientemente del detector, limita el proceso al uso de información espacial de los objetivos (no hay otra forma de obtener información para caracterizarla).

Para los diferentes tipos de oclusión mencionados en la introducción, la información espacial juega un papel muy importante en la adaptación realizada de GM-PHD. Como se mostró en el trabajo relacionado, la principal tendencia para caracterizar la oclusión en los filtros basados en PHD es la utilización de regiones [**?**] [**?**] [**?**]. En esta propuesta se definieron tres regiones para este propósito:

1. Región de posible oclusión: Cada seguidor en la imagen tendrá asociada una región de posible oclusión, la cual se determina mediante un análisis de las dimensiones de cada objeto. La región de posible oclusión se representará como un círculo que utiliza al seguidor como centro y tiene una circunferencia de por lo menos el tamaño del objeto, figura **??**.

2. Región de oclusión: Cuando dos o más regiones de posible oclusión se interceptan, se genera una región de oclusión. El estado de los objetivos se considera ocluido, si y solo si, su seguidor desaparece en esa área, figura **??**. Las regiones de oclusión se forman por la unión de varias circunferencias, correspondientes a las regiones de posible oclusión de los seguidores. Cuando se crea una región de oclusión, el detector generalmente identificará un solo objeto con dimensiones más grandes, esto permitirá definir el centro de la región de oclusión en el centro del objeto identificado por el detector. Estrictamente una región de oclusión solo contiene seguidores u objetivos ocluidos, ya que se constituye de seguidores que no sobrevivieron a la intercepción de dos o más regiones de posible oclusión.

3. Región de oclusión con un objeto no identificado: En un escenario independientemente del ruido, existen objetos que son considerados como parte del fondo pero que pueden ocluir a objetos de interés haciendo que desaparezcan de la escena. Los límites de estos objetos caracterizarán a una región de oclusión con un objeto no identificado, figura **??**. Estas regiones se calcularán mediante un análisis espacial de las muertes repentinas de los seguidores, y su cálculo estará estrechamente relacionado con las ecuaciones de GM-PHD.

De forma práctica, la región de oclusión se define mediante la utilización de las distancias entre objetivos. El estado del seguidor, se puede calcular mediante el

uso de una distancia d, en este caso euclidiana. La ecuación **??** determina si un seguidor se encuentra en estado ocluido.

$$E(x_k^i, x_k^j) = \begin{cases} 1 & \text{si } d(x_k^i, x_k^j) \leqslant \epsilon \text{ y } P_{S_k} \leqslant P_s(x_k^i) \text{ ó } P_s(x_k^j) \\ 0 & \text{de cualquier otra forma} \end{cases} \tag{4.1}$$

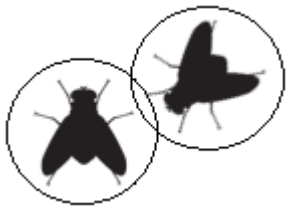
Se dice que un seguidor está en estado de oclusión con otro seguidor $E(x_k^i, x_k^j) = 1$, si su región de posible oclusión está traslapada en algún punto con otra (la distancia sea menor o igual a un ϵ) y la probabilidad de supervivencia del seguidor P_{S_k} sea menor al umbral mínimo definido para seguir activo en el siguiente frame. El umbral ϵ para determinar la distancia de oclusión se calcula a partir del tamaño de los objetos en el frame actual. En la sección **??** se realiza un análisis del umbral seleccionado (tamaño promedio de los objetos en el frame actual).

Por otra parte, un seguidor asociado a un objetivo de interés puede desaparecer dentro de los límites de la imagen sin entrar en una región de oclusión, por lo que puede ser considerado ruido incorrectamente. Para lidiar con este problema, se define una región donde la probabilidad de supervivencia y nacimiento de los seguidores se adapta de tal forma que el seguidor pueda recuperarse fácilmente. Esta región a diferencia de la región de oclusión se calcula por medio de información espacial de las muertes de los seguidores y no por una distancia.

La región de oclusión con un objeto no identificado, se construye a partir de la fusión de regiones de posible oclusión. La principal diferencia con la región de oclusión es que se calcula dinámicamente durante toda la secuencia de imágenes o datos sintéticos. Cada vez que un seguidor desaparece dentro de la escena por una oclusión con un objeto desconocido, la información espacial de su última vista se utiliza para construir o redefinir una región.

Dado un conjunto R de todas las regiones de oclusión con un objeto no identificado y la región de posible oclusión de un objeto cuyo seguidor desaparecerá sin entrar en una región de oclusión. Se verifica que la intersección de la región de posible oclusión con las regiones de R sea diferente de $\varnothing$, si es así se

(a) Región de posible oclusión.



(b) Región de oclusión.

(c) Región de oclusión con un objeto no identificado.

Figura 4.3: Tipos de regiones.

actualizan los límites de las regiones cuya intersección no sea vacía (si más de una región de R es actualizada, se fusionan). En caso que la intersección entre la región de posible oclusión y todos los elementos de R sea $\varnothing$, la región de posible oclusión se añade al conjunto y forma una nueva región de oclusión con un objeto no identificado.

4.1.3. Modelado del sistema

En el modelado del sistema se inicializan tanto los parámetros del filtro GM-PHD, como la dinámica de los objetivos, las observaciones y para los escenarios sintéticos, el área de vigilancia (tamaño, ruido, objetivos y oclusiones). El modelo del sistema está definido en un tiempo discreto donde el periodo de muestreo es $T_s = 1$ y depende directamente del número de *frames* que contiene el video.

Cada objetivo y el sensor, siguen un modelo dinámico lineal Gaussiano. Donde el estado x_k de cada objetivo está dado por el conjunto $\{(x_k^i, y_k^i), \mathfrak{d}, \mathfrak{v}\}$ que representa el centroide del objeto en el tiempo k, la dirección del objeto

y su velocidad. Por otra parte, las observaciones z_k están representadas por un par ordenado (x, y). El modelado del sistema comprende dos funciones: la función de transición multiobjetivo, ecuación **??** y la función de verosimilitud multiobjetivo, ecuación**??**.

$$f_{k|k-1}(x_k|x_{k+1}) = \mathcal{N}(x_k; F_k x_{k-1}, Q_k) \tag{4.2}$$

$$g_k(z_k|x_k) = \mathcal{N}(z_k; H_k x_k, R_k) \tag{4.3}$$

Donde $\mathcal{N}$ es una distribución Gaussiana de la forma $\mathcal{N}(; m, P)$, m es la media y P la covarianza de la distribución. Por otra parte F representa la matriz de transición de estados y H es la matriz de observaciones. Q la matriz del ruido del sistema y R es la matriz de ruido de las observaciones.

$$F = H = \begin{bmatrix} 1 & T_s & 0 \\ 0 & 1 & T_s \\ 0 & 0 & 1 \end{bmatrix}$$

$$Q = R = \begin{bmatrix} \frac{T_s^2}{2} & T_s & 0 \\ 0 & \frac{T_s^2}{2} & T_s \\ 0 & 0 & \frac{T_s^2}{2} \end{bmatrix}$$

Modelo de movimiento

Se asume que la colección de mediciones se realiza en el tiempo k y $k+1$. El comportamiento del objetivo es modelado por:

- El movimiento de los objetivos individuales, ecuación **??**, que representa la probabilidad que un objetivo tenga un estado x en el tiempo $k+1$, dado su estado en k.

- $P_s(x_k)$ es la probabilidad de supervivencia, o que un objetivo persista en $k+1$, si tiene un estado x_k.

- $\gamma_k(x_k)$ es la densidad de probabilidad que representa que un conjunto X_k de nuevos objetivos aparezcan en $k+1$.

Se modela la distribución de nacimiento espontaneo de objetivos con una mezcla de Gaussianas, ecuación **??**. La cual está distribuida uniformemente en toda la imagen. Esta distribución en el método propuesto cambiará a lo largo del muestreo, añadiendo los objetivos con estados ocluidos.

$$\gamma_k(x_k) = \sum_{i=1}^{J\gamma} w_{\gamma,k}^{(i)} \mathcal{N}(x_k; m_{\gamma,k}^{(i)}, P_{\gamma,k}^{(i)}) \tag{4.4}$$

Donde $m_{\gamma,k}^{(i)}$ es la media y $P_{\gamma,k}^{(i)}$ la varianza de la Gaussiana y determinan la forma de la intensidad de nacimiento.

Por otro lado, la probabilidad de supervivencia es independiente del estado y en la inicialización se define como una constante.

$$P_{s,k}(x_0) = 0{,}99 \tag{4.5}$$

El valor elegido para la probabilidad de supervivencia de los objetivos se fijó en 0.99, de acuerdo al trabajo [**?**], donde se realizó un análisis que indica que ese valor permite a los seguidores asociados a objetivos de interés sobrevivir sin que el ruido se propague.

Modelado de las observaciones

Dado un solo sensor, el espacio de todas las mediciones es modelado mediante:

- $P_{d,k}(x_k)$, la probabilidad que un objetivo sea generado por una observación.
- Si un objetivo genera un conjunto de observaciones, la función de probabilidad, ecuación **??**, es la probabilidad que las observaciones sean z.
- En $k+1$, la probabilidad que N falsas alarmas sean generadas sigue una distribución de Poisson.

La probabilidad de detección es independiente del estado y en la incialización se define como constante.

$$P_{d,k}(x_k) = 0{,}98 \tag{4.6}$$

El valor elegido para la probabilidad de detección de los objetivos se fijó en 0,98, de acuerdo al trabajo [**?**]. En ese trabajo se realizaron un conjunto de experimentos, donde se examinaron diferentes valores de la probabilidad de detección en 2 escenarios, con la métrica CPEP (*circular position error probability*), llegando a la conclusión de que ese valor permite a los objetos de interés ser detectados, manteniendo el nivel de falsos positivos en su más baja tasa.

4.1.4. Adaptación de la probabilidad de detección

Cuando un objetivo se encuentra en estado de oclusión, el filtro GM-PHD lo ignora y no es detectado en *frames* posteriores, debido a que el peso del objetivo ocluido decrece sustancialmente en el tiempo.

Para impedir que el peso asociado a un objeto ocluido disminuya rápidamente, se adaptó la probabilidad de detección $P_{d,k}(x_k^i)$, ajustando su valor de acuerdo a las condiciones y estado del objeto. Cuando el objetivo se encuentra en una región de oclusión, la probabilidad de detección cambiará a 1. Es decir, que el seguidor asociado al objetivo de interés permanecerá en la escena con una probabilidad del 100 %.

La probabilidad de detección una vez que se encuentra en estado de oclusión, disminuirá con un valor χ, el cual se elige de tal forma que el seguidor no muera después de una oclusión, pero que no permanezca durante toda la secuencia si sale del campo de visión del sensor o no vuelve a aparecer. Para ello se calcula a χ con ayuda del tiempo promedio de oclusión de la siguiente forma:

$$\chi_k = \frac{\text{rango máximo de decremento de } P_{d,k}(x^i)}{\text{frames promedio de oclusión} * 100} \tag{4.7}$$

Donde el dividiendo representa el rango máximo de disminución de la probabilidad de detección. Permite que la probabilidad de detección de cada objetivo mantenga un valor mínimo (esto no garantiza que el seguidor sobreviva). El valor mínimo de la probabilidad de detección se utiliza para que objetivos que tengan una duración de oclusión atípica puedan ser detectados; por ejemplo, un objetivo que se mantenga ocluido durante 10 *frames* cuando el promedio es 5.

La elección de este valor mínimo, depende de las condiciones del problema (la desviación estándar de los *frames* promedio de oclusión). En los experimentos realizados el valor mínimo se fijó en 0,7, lo que permite al seguidor mantenerse si su movimiento no es abrupto y las observaciones generadas son consistentes.

El valor de χ_k para objetivos que no estén en estado de oclusión es 0 (no repercute en absoluto), en caso contrario se utilizara la ecuación **??**. Sin embargo cuando se desconoce la duración promedio de las oclusiones, se asigna como parámetro de entrada del algoritmo. La probabilidad de detección se aplicará para cada objetivo en la escena dependiendo de su estado, el cálculo de esta probabilidad está definido por la ecuación **??**.

$$P_{d,k}(x^i) = P_{d,k-1}(x^i) - \chi_k \tag{4.8}$$

4.1.5. Adaptación de la distribución de nacimiento y supervivencia de objetos

Se sabe que los seguidores en GM-PHD pueden estar asociados a tres tipos de objetos [**?**]:

1. objetivos que existen en $k-1$ y continúan en k.

2. objetivos que se generan a partir de las mediciones de k.

3. objetivos que son engendrados a partir de objetivos en $k-1$.

De acuerdo a estos tipos, los estados de los objetivos se verifican de 3 formas diferentes en PHD: los objetivos que sobreviven, los objetivos que nacen espontáneamente y los objetivos que son engendrados.

Por otro lado, el algoritmo GM-PHD se compone de tres partes fundamentales: la recursión, la poda de estados y la extracción de múltiples estados. La recursión a su vez, se conforma de dos fases: la predicción de estados y la actualización de los estados.

La predicción está dada por la siguiente intensidad:

$$v_{(k|k-1)}(x) = v_{(S,k|k-1)}(x) + v_{(\beta,k|k-1)}(x) + \gamma_k(x) \tag{4.9}$$

Donde $\gamma_k(x)$ es la intensidad de nacimiento espontaneo y se define como:

$$\gamma_k(x) = \sum_{i=1}^{J\gamma} w_{\gamma,k}^{(i)} \mathcal{N}(x_k; m_{\gamma,k}^{(i)}, P_{\gamma,k}^{(i)}) \tag{4.10}$$

Y la intensidad de supervivencia, se define como:

$$v_{(S,k|k-1)}(x) = P_s(x_{k+1}) \sum_{i=1}^{J_k} w_{k|k}^{(i)} \mathcal{N}(x_k; m_{k+1,k}^{(i)}, P_{k+1,k}^{(i)}) \tag{4.11}$$

Las ecuaciones **??** y **??** tienen una distribución Gaussiana asociada, la cual podrá ajustarse de tal forma que la información recolectada de todas las regiones que caracterizan la oclusión se incorpore en la predicción.

Las adaptaciones en la intensidad de nacimiento y la intensidad de supervivencia, añadiendo información espacial, permitirán al seguidor continuar en la escena aun cuando este no exista, coadyuvando así a la resolución del problema de oclusión.

Incorporación de información espacial en GM-PHD

Mediante la modificación de la distribución Gaussiana, ecuación **??**, en concreto de la media $m_k^{(i)}$ que contiene información espacial del centro de la región y la varianza $P_k^{(i)}$ que guarda la información sobre los límites [**?**], se puede incorporar información espacial tanto en la intensidad de nacimiento, ecuación **??**, como en la intensidad de supervivencia, ecuación **??**. Para ello, se modifican las Gaussianas asociadas a los estados (ecuación **??**), añadiendo información sobre el tamaño y posición de regiones de oclusión y regiones de oclusión con objetos no identificados.

$$\mathcal{N}(x; m_k^{(i)}, P_k^{(i)}) \tag{4.12}$$

$$\sum_{i=1}^{J_k} w_k^{(i)} \mathcal{N}(x; m_k^{(i)}, P_k^{(i)}) \tag{4.13}$$

Donde $\sum_{i=1}^{J_k} w_k^{(i)}$ es la sumatoria de la probabilidad de detección o supervivencia de los objetivos en el tiempo k.

Por cada región de oclusión con un objetivo de interés, se construyen dos vectores $\tilde{m}_k$ y $\tilde{P}_k$, los cuales contienen información del centro y límites de las regiones de oclusión. Se construyen Gaussianas a partir de estos vectores, las cuales de incorporan a la mezcla de Gaussianas **??**, para permitir al filtro actualizar las intensidades de supervivencia. Para las regiones de oclusión con un objeto no identificado se construyen $\tilde{m}_k^1$ y $\tilde{P}_k^1$, que permitirán del mismo modo actualizar la intensidad de nacimiento. Se añaden entonces, regiones con alta probabilidad de nacimiento y supervivencia que se propagan a través del tiempo. Parte fundamental de este trabajo.

Distribución de nacimiento adaptable

Uno de los principales problemas del filtro PHD en escenarios reales, es que requiere el estado inicial de los objetivo como parámetro. Para superar este obstáculo comúnmente se igualan las primeras mediciones al estado inicial, $X_0 = Z_1$, o se configura manualmente el conjunto de estados X_0 [**?**] [**?**] [**?**]. Con estos mecanismos, y sobre todo en escenas con mucho ruido, se propagan errores a través del tiempo.

En aplicaciones prácticas se limita la capacidad del filtro, ya que es necesario conocer el estado X_0 para poder calcular los siguientes. En esta propuesta se ataca este problema con un mecanismo de ajuste gradual, lo que permite discriminar los objetivos del ruido. Entonces la intensidad de nacimiento en el tiempo 0 se establece como:

$$v_{0|0} = 0 \tag{4.14}$$

Y los estados subsecuentes están dados por la siguiente ecuación [**?**]:

$$v_{k+1|k}(x) = P_s(x_{k+1}) \sum_{i=1}^{J_k} w_{k|k}^{(i)} \mathcal{N}(x_k; m_{k+1,k}^{(i)}, P_{k+1,k}^{(i)}) \tag{4.15}$$

Teniendo en cuenta que las observaciones en el tiempo k, Z_k, pueden surgir de objetivos que sobrevivieron, objetivos que nacieron o se engendraron, el filtro GM-PHD es capaz de diferenciar el ruido de los objetivos sobrevivientes y los nuevos objetivos, si y solo si sus intensidades se determinan correctamente antes del filtrado. Cuando $k = 0$, solo es posible encontrar nuevos objetivos, es por eso que se ocupa solamente la intensidad de supervivencia, ecuación **??**.

Por otro lado cuando k es mayor que 0, la distribución o intensidad de nacimiento, ecuación **??**, sirve para caracterizar regiones donde la probabilidad de nacimiento es considerablemente mayor. Si se utiliza una distribución uniforme en toda la imagen (como generalmente se hace), se mantiene la suposición de que la probabilidad de nacimiento es la misma para cualquier región, lo que no es verdad.

Como se mencionó en la introducción, la intensidad de nacimientos espontáneos es modelada por una mezcla Gaussiana, cuyos componentes se distribuyen en toda la región de vigilancia. Aunque esta hipótesis es válida para algunas aplicaciones específicas, puede ser demasiado restrictiva.

Se propone entonces, incorporar la información obtenida de las regiones de oclusión y otras zonas predefinidas como regiones de nacimiento de objetos para robustecer el seguimiento de objetivos nuevos. La simple idea que motiva a este enfoque es que los objetivos son más propensos a aparecer alrededor de regiones donde han nacido objetivos anteriormente y alrededor de las regiones de oclusión con objetos no identificados.

Se establece una intensidad de nacimiento del filtro PHD, con información a priori espacial, modificando la ecuación **??** con la media y varianza definidas en las ecuaciones **??** y **??**, donde los vectores m^1_{k-1} y P^1_{k-1} contienen información de las regiones de oclusión. La regiones de nacimiento de objetos se actualizarán con la información de los objetivos recién nacidos anteriores.

$$m^{(k)}_{S,k|k-1} = F_{k-1}\tilde{m}^1_{k-1} \tag{4.16}$$

$$P^{(k)}_{S,k|k-1} = Q_{k-1} + F_{k-1}\tilde{P}^1_{k-1}F^T_{k-1} \tag{4.17}$$

Distribución de supervivencia adaptable

Cuando una región de oclusión aparece en el escenario, los objetivos que se encuentran en ella cambiarán a un estado ocluido y sus seguidores desaparecerán eventualmente porque ya no tienen asignado a ningún objetivo. De acuerdo a la caracterización de la oclusión propuesta, para manejar eficazmente

la oclusión en GM-PHD, es necesario garantizar la supervivencia de seguidores en regiones de oclusión con otros objetivos de interés.

Adaptar la probabilidad de supervivencia no garantiza la supervivencia del seguidor asociado a un objetivo ocluido, ya que se necesitaría normalizar los pesos de las Gaussianas para que se detectado en k+1, tal y como se realizó para el método *Competitive Gaussian Mixture Probability Hypothesis Density* (CGM-PHD) en [**?**]. Esto limita, el número de oclusiones que pueden realizarse simultáneamente. Entonces, se propone utilizar las regiones de los objetivos en estado de oclusión para actualizar la intensidad de supervivencia, para ello se usan los estados obtenidos en el tiempo k, en lugar de $k-1$. Esta distribución nos permite definir para cada seguidor una probabilidad de supervivencia, la cual determina la existencia del seguidor en los *frames* subsecuentes.

Al modificar la intensidad de supervivencia, ecuación **??**, de un objetivo en estado de oclusión se obliga al seguidor a mantenerse en la escena hasta que cambie su estado o se alcance un umbral de muerte χ. En las ecuaciones **??** y **??** podemos notar la construcción de la media y varianza modificada para la ecuación **??**, cabe notar que tanto los vectores $\tilde{m}$ y $\tilde{P}$ contienen la información de las regiones y dependen del estado actual, es decir se nutren de las modificaciones hechas en el estado k, sobre las regiones de oclusión.

$$m^{(k)}_{S,k|k-1} = F_{k-1}\tilde{m}_k \tag{4.18}$$

$$P^{(k)}_{S,k|k-1} = Q_{k-1} + F_{k-1}\tilde{P}_k F^T_{k-1} \tag{4.19}$$

4.1.6. Asignación de etiquetas

Como se ha mencionado antes los algoritmos PHD no consideran la asociación de datos para mantener etiquetado a los objetivos. Sin embargo se puede incorporar un algoritmo de asignación de etiquetas al filtro, de tal manera que trabaje con el conjunto de estados extraídos en cada *frame*.

En general, estos filtros reciben como entrada los datos de los objetivos que tengan asociado un seguidor y los objetivos etiquetados del estado $k-1$. En

el presente trabajo el algoritmo Húngaro es utilizado para la asignación de etiquetas.

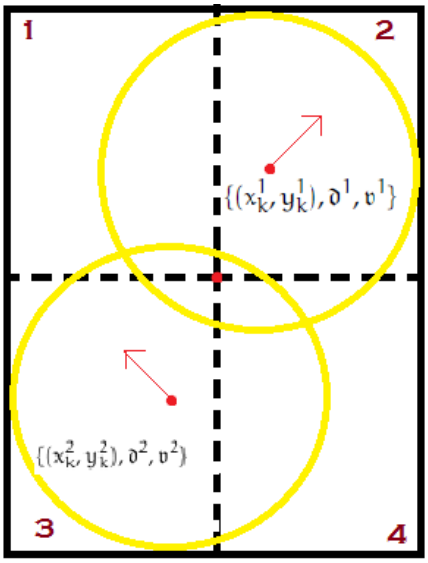


Figura 4.4: Representación de los cuadrantes usados para impulsar el algoritmo Húngaro.

El algoritmo Húngaro no tiene en cuenta el desorden y ruido espacial, pero el costo computacional es menor que otras técnicas intensivas como la asociación de datos en MHT. La elección de un algoritmo de asociación de etiquetas como módulo independiente, está motivada por el hecho de que no se necesita manejar el desorden ni el ruido en esa etapa, ya que han sido tratados por el filtro GM-PHD.

Aunque el algoritmo Húngaro provee un buen desempeño en la asignación de etiquetas, cuando existe oclusión, los objetivos se fusionan y se pierde la información sobre sus posiciones. Entonces el algoritmo es propenso a fallar, debido a que en el transcurso de la oclusión, no existe información de las posiciones del objetivo.

Para ayudar al algoritmo de asignación de etiquetas a mejorar su desempeño durante la oclusión, el estado X_k estará formado por la dupla de posición, la

dirección anterior y la velocidad promedio de los objetivos en las etiquetas:

$$X_k = \{(x_k^i, y_k^i), \vartheta^i, \upsilon^i\} \tag{4.20}$$

Donde la velocidad se calcula por una diferencia de las posiciones del objetivo en k y $k-1$, y la dirección θ se calcula usando las relaciones trigonométricas del triángulo rectángulo en el teorema de Pitágoras.

$$\upsilon = x_k - x_{k-1} \tag{4.21}$$

$$\theta = arctan(\frac{y_2 - y_1}{x_2 - x_1}) \tag{4.22}$$

En situaciones de ambigüedad, donde el algoritmo de asignación de etiquetas encuentra una misma etiqueta para dos objetivos, o no se cumpla alguna validación desarrollada en la sección **??**, la dirección y velocidad promedio serán fundamentales para dar mayor prioridad a objetivos en cuadrantes donde sea más probable que aparezca el objetivo asignado a la etiqueta. El cuadrante más probable se calcula tomando como punto de referencia la última posición del objetivo. Por otra parte, la velocidad nos permitirá establecer cuál es la distancia probable del objetivo (solo se utiliza si la región de oclusión es grande, 3 a 4 objetos). Una representación de los cuadrantes utilizados para impulsar el algoritmo Húngaro, a partir de la velocidad y la dirección, se puede visualizar en la figura **??**.

La figura **??**, nos muestra dos objetivos en estado de oclusión. Las regiones de posible oclusión de cada objeto representadas por el circulo amarillo y la región de oclusión representada por el rectángulo negro. En esta figura, se observa la división de la región de oclusión y de acuerdo a la dirección, el primer objetivo aparecería alrededor del cuadrante 2 y el segundo objetivo alrededor del cuadrante 1. La figura **??** representa un problema simple de asignación de etiquetas, sin embargo cuando los objetivos no siguen movimientos lineales, es muy difícil identificarlos y asignarles su etiqueta correctamente, ya que no podemos determinar sus posiciones dentro de la región de oclusión.

Parte V

Resultados y Experimentos

Capítulo 5

Resultados y Experimentos

En este capítulo se presentan los resultados obtenidos a partir esta tesis y su comparación con los métodos existentes. Se presentan tablas con base en las evaluaciones, analizando los resultados. En primer lugar se realiza una descripción de las bases de datos que se utilizaron. Posteriormente se definen de forma general las métricas de evaluación utilizadas en los experimentos. A continuación se describen los experimentos sobre las bases de datos y los escenarios sintéticos. Finalmente se presentan los resultados alcanzados demostrando el desempeño del método presentado en este trabajo.

5.1. Detalles de implementación

Todos los mecanismos descritos en la parte del diseño fueron implementados en MATLAB R2013a (8.1.0.604), con la ayuda de la librería MexOpenCv, que es una colección de funciones mex para la biblioteca OpenCV.

Los algoritmos CGM-PHD y el método propuesto se diseñaron a partir de una implementación de Bryan Clarke del filtro GM-PHD, el cual está disponible en la página de intercambio de archivos de MATLAB. Por otra parte el algoritmo *Multiple Hypothesis Tracking* se obtuvo de una librería desarrollada en Java por David Antunes, desde su página principal.

En los experimentos todos los algoritmos fueron ejecutados en la misma computadora y compilados de la misma forma. A continuación se muestran las características de la computadora utilizada:

- Sistema Operativo: Windows 8.1, 64 bits (x64).
- Procesador: Intel Core i5-3337U, 1.8 GHz.
- Memoria RAM: 6 GB.

5.2. Descripción de base de datos

Para evaluar el método desarrollado en este trabajo se utilizó un conjunto de bases de datos. Videos que contienen varios insectos en vuelo. Las características de los insectos en los videos permite resaltar las virtudes del método propuesto, ya que están representados por un conjunto pequeño de pixeles, menos de 30, y se mueven rápidamente en la escena, generando oclusiones.

La tabla **??** muestra una descripción de las bases de datos con base en sus características. Las principales características presentadas en esta tabla son: el número de objetos que contiene en promedio cada video de la base de datos, el tamaño promedio en pixeles de los objetos, el número de videos que contiene cada base de datos, y por último los retos que presenta cada base de datos: múltiples objetos, existencia de oclusión o existencia de ruido. Una descripción del origen de las bases de datos, se registra en la siguiente lista:

- Caltech modificado: Esta base de datos se construyó para evaluar un software de seguimiento del Instituto Tecnológico de Carolina, Ctrax: The Caltech Multiple Walking Fly Tracker. Consta de 8 videos de moscas de fruta, que fueron obtenidos por una cámara infrarroja. Se eligieron 4 videos, donde existe oclusión entre los objetos, para evaluar el método presentado en este trabajo.
- Ants: Base de datos de hormigas en un contenedor y en su ambiente. Esta base de datos sirve como referencia para evaluar el algoritmo MHT, obtenido de la Multiple Hypothesis Library.
- videos Random: Esta base de datos, se construyó a partir de una serie de videos obtenidos de http://www.shutterstock.com/. Se utiliza para evaluar el problema de la aparición aleatoria de objetos en la escena.

Permitiendo analizar los problemas que surgen en nuestro método con el cambio drástico de la probabilidad de aparición de objetos, ya que los objetivos en esta base de datos aparecen y desaparecen en cortos periodos de tiempo.

- Aves: Es una de las bases representativas para el problema de *Aircraft tracking*, sirve para evaluar los cambios repentinos en la escena, los cuales generan ruido.
- Extremas: Esta base de datos también se construyó a partir de videos obtenidos de http://www.shutterstock.com/, con la finalidad de analizar el desempeño del filtro en escenarios con multitudes, donde existen oclusiones largas y constantes. Dos del total de videos de este conjunto son simulaciones de múltiples objetos, no contienen ruido.

Tabla 5.1: Caracteristicas de las bases de datos.

BD	Objetos	Tamaño	Videos	MOT/O/R
Caltech modificado	20 moscas	8 pixeles	4	SI/SI/NO
Ants	20 hormigas	20 pixeles	3	SI/SI/SI
Videos Random	15 abejas	30 pixeles	2	SI/SI/SI
Aves	15 aves	25 pixeles	3	SI/SI/SI
Extremas	30-100 insectos	15 pixeles	4	SI/SI/SI

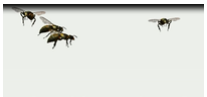
R: Ruido O:Oclusión MOT:Múltiples objetos

En la figura **??** se puede observar de forma general la apariencia y tamaño de los objetos de cada base de datos. En particular se eligieron secciones de *frames* de 200 pixeles de largo por 92 pixeles de ancho. Cabe resaltar que en el caso de la figura **??** se eligió uno de los dos videos con modelos sintéticos.

Por otro lado, para evaluar el método propuesto y compararlo con otros métodos se utilizan 3 escenarios sintéticos, vea figura **??**. El primer escenario se puede encontrar en la página del grupo de investigación *Advanced Learning Evolutionary Algorithms* del *Institut de Mathématiques de Bordeaux* (INRIA), figura

(a) Objetos de Caltech modificado.

(b) Objetos de Ants.



(c) Objetos de videos Random.

(d) Objetos de Aves.

(e) Objetos de Extremas.

Figura 5.1: Apariencia de los objetos de las bases de datos.

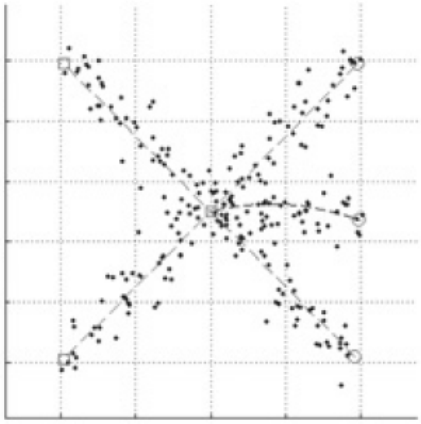
??, mientras que los otros dos escenarios representados por las figuras **?? ??** están definidos en [**?**] [**?**] respectivamente.

Las características de los escenarios sintéticos se mencionan a continuación:

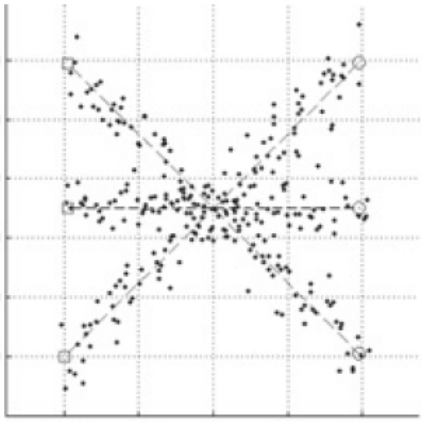
- Escenario 1, figura **??**: En este escenario 3 objetivos se cruzan en la mitad del escenario. Dos objetivos entran en escena en el tiempo 1 y termina en el tiempo 50, por otra parte un tercer objetivo nace cuando los objetivos se encuentran en oclusión en el tiempo 25. El escenario tiene una altura de 1000 ($[0, 1000]$) y un ancho de 2000 ($[-1000, 1000]$) unidades.

- Escenario 2, figura **??**: En este escenario tres blancos se cruzan en el punto [0,500] (centro del escenario), los tres objetivos entran a escena en el tiempo 1 y salen de escena en el tiempo 50. El escenario tiene una altura de 1000 ($[0, 1000]$) y un ancho de 2000 ($[-1000, 1000]$) unidades.

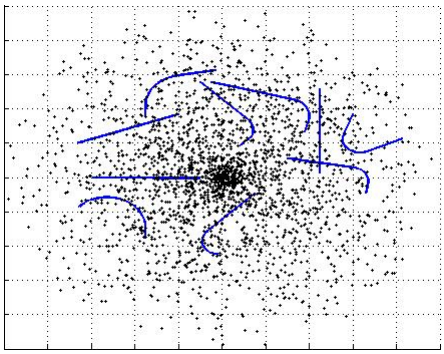
- Escenario 3, figura **??**: En este escenario 11 objetivos son generados, 2 objetivos nacen cada 5 unidades de tiempo. Durante cada paso de tiempo 50 falsas alarmas de forma aleatoria se generan (ruido).



(a) Escena 1.



(b) Escena 2.



(c) Escena 3.

Figura 5.2: Escenarios sintéticos para la comparación del método propuesto.

Los escenarios de las figuras **??** y **??** permiten evaluar el método propuesto en un marco de referencia con su más cercano competidor [**?**]. Mientras que **??** nos permite realizar una evaluación en un ambiente con mucho ruido. En el caso de **??** y **??** el ruido no es estrictamente el mismo que se utiliza en [**?**] y [**?**]. Se acondicionó un ruido aleatorio que no superará 50 unidades de la observación del objetivo real.

5.3. Construcción del *Ground Truth*

Se construyeron 3 *ground truth* para evaluar la detección, el seguimiento de objetos y la asignación de etiquetas. Estos puntos de referencia se realizaron manualmente. Se implementó un programa que por cada *frame* de cada video de las bases de datos, permite:

1. Seleccionar el contorno de los objetos para guardar los pixeles que los representan.
2. Seleccionar un punto del objeto (aproximadamente el centro), y guardar su posición.
3. Añadir una etiqueta al punto elegido en 2. La asociación de los objetivos se realizó de forma manual, mediante una inspección visual del movimiento de los objetos.

El programa realizado genera, a partir de la recolección de la información, un marco de evaluación para la detección, el seguimiento y la asignación de etiquetas.

5.4. Métricas de evaluación

La evaluación del rendimiento de detección de objetos y el seguimiento de múltiples objetos es una de las principales tareas para validar la exactitud y robustez de un algoritmo de seguimiento. Por ello, en 2006 el *National Institute of Standards and Technology* (NIST), realizó el taller de *Classification of Events, Activities, and Relationships* (CLEAR). El cual tuvo como objetivo establecer un marco de evaluación universal para el seguimiento múltiple. En este taller

se presentaron diversas métricas que permiten comparar objetivamente las características del algoritmo de seguimiento, centrándose en su precision para estimar la localización de objetos y su habilidad para consistentemente etiquetar objetos. En esta sección se examinarán las principales métricas que permitirán evaluar la propuesta desarrollada.

5.4.1. Métricas para algoritmos de seguimiento

Para medir el desempeño de los algoritmos de seguimiento múltiple, se utilizan las siguientes métricas:

- Tasa de falsas alarmas (FAR): Se emplea para evaluar el efecto del ruido en el experimento, se define de la siguiente forma.

$$\mathrm{FAR} = \frac{\text{número de falsas estimaciones}}{\text{total de estimaciones}} \tag{5.1}$$

- Tasa de exhaustividad (ReR): Se utiliza para examinar la capacidad del filtro para seguir objetivos en condiciones de oclusión.

$$\mathrm{ReR} = \frac{\text{número de estimaciones correctas}}{\text{total de objetivos verdaderos}} \tag{5.2}$$

- Tasa de asignaciones pérdidas (MLR): Es definida como el porcentaje promedio de asignaciones de etiquetas incorrectas.

$$\mathrm{MLR} = \frac{\text{suma del tiempo etiquetas pérdidas}}{\text{total de tiempo de objetivos}} \tag{5.3}$$

- Tasa de oclusiones pérdidas (MOR): Evalúa la capacidad del algoritmo para manejar la oclusión.

$$\mathrm{MOR} = \frac{\text{oclusiones pérdidas}}{\text{total de oclusiones}} \tag{5.4}$$

5.4.2. *Multiple object tracking accuracy* (MOTA)

MOTA representa todos los errores del seguimiento realizados por el rastreador. Es similar a las métricas utilizadas en otros ámbitos, porque define una medida simple pero funcional del proceso de seguimiento y la detección de objetos. MOTA puede verse como una conjunción de 3 tipos de errores [**?**]:

El error de pérdidas en la secuencia, entre el número total de objetos en todos los *frames*:

$$\tilde{m} = \frac{\sum_t m_t}{\sum_t g_t} \tag{5.5}$$

El error de falsos positivos:

$$\tilde{f_p} = \frac{\sum_t f_{pt}}{\sum_t g_t} \tag{5.6}$$

El error de falsas correspondencias:

$$\tilde{mme} = \frac{\sum_t mme_t}{\sum_t g_t} \tag{5.7}$$

Donde m_t, f_{pt} y mme_t son las pérdidas en la secuencia, falsos positivos y falsas correspondencias en el tiempo t respectivamente.

Entonces MOTA es una unión de las 3 métricas anteriores, descrita en **??**.

$$\text{MOTA} = 1 - \frac{\sum_t (m_t + f_{pt} + mme_t)}{\sum_t g_t} = 1 - (\tilde{m} + \tilde{f_p} + \tilde{mme}) \tag{5.8}$$

5.4.3. *Optimal Subpattern Assigment*

En [**?**] [**?**] [**?**], se menciona que una métrica para evaluar múltiples objetos debe cumplir los siguientes puntos esenciales:

- ser una métrica definida en conjunto finitos,
- tener una interpretación física natural,
- capturar errores de cardinalidad y errores de estado de manera significativa,
- ser fácilmente computable.

En el presente trabajo para evaluar los diferentes experimentos se recurre a *Optimal Subpattern Assigment* (OSPA), métrica que cumple estas características. Fue presentada, en [**?**] y [**?**], y se basa en la métrica Wasserstein.

La métrica OSPA es una distancia entre dos conjuntos cualquiera de múltiples objetos, la cardinalidad de los conjuntos es m y n respectivamente.

OSPA está definida por $d^{(c)}(x,y) = min(c, d(x,y))$ donde $x, y \in W$, ventana de observación cerrada y acotada; Π_k denota las permutaciones de $\{1, 2, 3..., k\}$. Entonces para $p \geqslant 1$, $c > 0$, $X = \{x_1, ..., x_m\}$, $Y = \{y_1, ..., y_n\}$, la métrica OSPA, está defina mediante la ecuación **??**.

$$\tilde{d}_p^{(c)}(X,Y) = \begin{cases} 0 & \text{si } m = n = 0, \\ (\frac{1}{n}(min_{\pi \in \Pi_n} \sum_{i=1}^{m} d^{(c)}(x_i, y_{\pi(i)})^p + c^p(n-m)))^{\frac{1}{p}} & \text{si } m \leqslant n, \\ \tilde{d}_p^{(c)}(Y,X) & \text{si } m > n, \end{cases} \tag{5.9}$$

Donde c es el parámetro de corte entre dos objetivos y p es el parámetro de orden. La elección de los parámetros c y p sigue la configuración de la métrica original, OSPA [**?**]. El parámetro punto de corte, c, determina la ponderación relativa atribuida a la componente de error de cardinalidad contra el componente de error de distancia máxima. Valores altos de c tienden a enfatizar los errores de cardinalidad y valores bajos enfatizan errores de distancia.

El error de cardinalidad verifica un número desigual de objetivos en los conjuntos de múltiples objetos, lo que significa que c puede interpretarse como una medida asignada a los seguidores perdidos o falsas detecciones. El parámetro de orden, p, controla estimaciones atípicas (que no están cerca del *ground trouth*). Un mayor valor de p aumenta la sensibilidad a los valores atípicos.

La métrica OSPA captura todos los aspectos de la diferencia entre dos conjuntos de estados: la cardinalidad, el error de localización y el error de etiquetado. El código fuente MATLAB para el cálculo de OSPA está disponible en la siguiente página http://randomsets.eps.hw.ac.uk/index.html.

En el contexto de la evaluación del desempeño multiobjeto, podemos utilizar la distancia OSPA para definir dos errores. Estos son los errores de localización y errores de cardinalidad. Para $p < \infty$ los errores están definidos por:

$$e_{p,loc}^{(c)}(X,Y) = (\frac{1}{n} min_{\pi \in \Pi} \sum_{i=1}^{m} \tilde{d}_p^{(c)}(x_i, y_{\pi(i)})^p)^{\frac{1}{p}} \tag{5.10}$$

$$e^{(c)}_{p,card}(X,Y) = (\frac{c^p(n-m)}{n})^{\frac{1}{p}} \tag{5.11}$$

Donde $e^{(c)}_{p,loc}(X,Y)$ y $e^{(c)}_{p,card}(X,Y)$ son los errores de localización y carnidalidad respectivamente. Si $m \leqslant n$, entonces $e^{(c)}_{p,loc}(X,Y) = e^{(c)}_{p,loc}(Y,X)$ y $e^{(c)}_{p,card}(X,Y) = e^{(c)}_{p,card}(Y,X)$. La descomposición de la métrica OSPA en componentes separados por lo general no es necesaria para la evaluación del desempeño, pero puede proporcionar valiosa información adicional.

5.4.4. Métricas para la detección de objetos

Con el fin de proporcionar una perspectiva cuantitativa sobre la calidad de la detección del fondo, varias métricas son utilizadas. Estas medidas son, la tasa de falsos negativos (FNR) y la tasa de falsos positivos (FPR), Recall y Precisión que se definen como:

$$FNR = \frac{N_{fbad}}{N_{fgt}} \tag{5.12}$$

Donde N_{fbad} es el número de pixeles incorrectamente clasificados como objeto y N_{fgt} es el número de pixeles del *Ground Truth* que son objeto.

$$FPR = \frac{N_{bbad}}{N_{bgt}} \tag{5.13}$$

Donde N_{bbad} son los pixeles del fondo clasificados incorrectamente y N_{bgt} son los pixeles del $\mathtt{GroundTruth}$ que pertenecen al fondo.

$$Recall = \frac{N_{fgood}}{N_{fgt}} \tag{5.14}$$

Donde el número de pixeles identificados correctamente como objeto es N_{fgood}, y N_{fgt} es el número de pixeles clasificados como objetos en el *Ground Truth*

$$Precision = \frac{N_{fgood}}{N_{fall}} \tag{5.15}$$

Donde N_{fall} son los pixeles clasificados como objetos por el algoritmo.

5.5. Experimentos y Resultados

En esta sección se establecieron las principales características de los experimentos realizados y sus funciones. Se fundamentó que se tienen dos fuentes de datos para los experimentos: los datos sintéticos, creados a partir de 3 escenarios, y los datos extraídos de videos de múltiples objetivos.

5.5.1. Detección de objetos

Se realizaron un conjunto de experimentos relacionados con la detección de objetos en videos con múltiples objetivos. En el primer paso, para mostrar el comportamiento del método propuesto, se realizó una evaluación de la detección con base en los datos definidos en la sección **??**, y las métricas para la detección de objetos.

El método propuesto se comparó con otros métodos (MHT, GM-PHD, CGM-PHD) que atacan el problema del seguimiento de múltiples objetos con estados ocluidos, para ello, se usó la detección de objetos como punto de referencia.

Selección de algoritmo para modelado del fondo

El detector es una pieza fundamental de cualquier algoritmo de seguimiento. La elección de un detector basado en el modelado del fondo se desprendió de la necesidad de enfatizar el proceso de seguimiento. Los detectores de modelado de fondo son algoritmos que procesan la información de la imagen rápidamente en comparación con otros detectores, aunque, tienden a generar falsos positivos y ruido.

Aunque no fue un objetivo de este trabajo evaluar el método propuesto con diferentes detectores. Un algoritmo de detección robusto reduciría el ruido y limitaría el problema, dejando en segundo plano las virtudes de esta propuesta.

Para los experimentos se eligieron 5 diferentes tipos de modelado de fondo (*Adaptive Background Learning*, *Gaussian Mixture Model*, GMG, *Pixel-Based Adaptive Segmenter*, *Gaussian Average*). Todas las implementaciones se obtuvieron de la

librería *bgslibrary* de OpenCV desarrollada por Andrews Sobral. La elección de estos algoritmos se basó en el trabajo [**?**] donde se midió el porcentaje promedio de uso de CPU, el porcentaje promedio de memoria y el porcentaje promedio de tiempo de ejecución sobre un conjunto de 2 videos.

Para seleccionar el algoritmo de modelado de fondo adecuado para el método propuesto, se aplicó cada algoritmo de modelado al conjunto de los videos de las bases de datos definidas en la tabla **??**. Se midió el tiempo promedio de procesamiento por cada frame, el *accuracy*, su *recall* y su medida F1. Se ejecutaron los diferentes algoritmos consecutivamente en una misma máquina, 5 veces y se promedió cada métrica.

Tabla 5.2: Resultados de la evaluación de los algoritmos de modelado del fondo

Modelado del fondo	Accuracy	Precision	Recall	F1	ms/frame
PBAS Hofmann et al (2012)	**0.662**	0.684	**0.665**	**0.668**	542
GMM Baf et al (2008)	0.657	**0.691**	0.625	0.657	86
GMG Godbehere et al (2012)	0.606	0.605	0.607	0.606	194
Adaptive Background Learning	0.433	0.545	0.521	0.448	324
Gaussian Average	0.451	0.475	0.511	0.447	**54**

En la tabla **??** se muestra la evaluación de los 5 principales métodos de modelado de fondo para la detección de objetos. Se utilizaron estos algoritmos, porque las características de las bases de datos se adaptaban bien (cámara estática, cambios graduales y sobre todo objetos pequeños). Aunque el desempeño de los algoritmos es bastante similar, existen ciertas diferencias sutiles que en conjunto con las características del método propuesto, permiten elegir uno para fungir como detector principal.

En la tabla **??** se observa que el método PBAS (*Pixel-Based Adaptive Segmenter*) consiguió mejores porcentajes en las métricas de detección, sin embargo, el procesamiento de cada *frame* requiere demasiado tiempo. Debido a esto, se eligió GMM como detector principal, ya que el tiempo de procesamiento es bajo y el desempeño de detección está un poco por debajo de PBAS. Otro punto a

Figura 5.3: Ejemplo de las cualidades de la configuración del filtro morfológico elegido.

favor del método GMM es que requiere mucho menos memoria y funciona bien con objetos pequeños, a diferencia de GMG y *Adaptative Background Learning* que en algunos casos, incluso, discriminaron algunos objetos de interés cuando su *blob* fue pequeño.

Configuración filtros morfológicos

Otra parte importante de la detección de objetos, es la elección y configuración de filtros morfológicos, encargados de apoyar a la detección para identificar con una mayor precisión a los objetos. El buen desempeño de una configuración del filtro morfológico está fuertemente relacionado con la forma de los objetos.

La selección de una estructura y un operador morfológico adecuada ayudó a que las observaciones de los objetivos no se dividieran o se mezclaran entre sí, figura **??**. En nuestro método el filtro morfológico también sirvió para reducir el número de regiones de oclusión, ya que, si un objetivo es identificado como dos objetivos, se crearían dos regiones de posible oclusión, posiblemente traslapadas, identificando a ese objetivo en estado de oclusión. Este problema disminuiría dramáticamente el desempeño del método propuesto.

Se diseñó un experimento que mide la trascendencia de varias configuraciones de filtros morfológicos en las métricas de precisión, *recall* y *accuracy* de la detección de objetos. Para la elección de la mejor configuración se utilizó un algoritmo genético que busca la configuración que maximiza el desempeño del

detector. Cada elemento de la población del algoritmo genético está caracterizado por un vector de números binarios que representan las características estructurales y los diferentes tipos de operadores del filtro morfológico (cruz, diamante, disco, rectángulo, dilatación, apertura, cierre y erosión). Por otra parte, los parámetros de dimensión de las características estructurales se eligieron en un rango de 2-4, ya que valores mayores perjudican la detección de los objetos pequeños en estas bases de datos.

El algoritmo genético se realizó con ayuda de la Toolbox de MatLab para Optimización. La función objetivo está dada por la ejecución de la detección con base en la configuración seleccionada de la población. Los parámetros de configuración del algoritmo genético se pueden visualizar en la tabla **??**.

Tabla 5.3: Parámetros del algoritmo evolutivo

Parámetro	Valor
Tamaño población	200
Población inicial	Aleatoria
Función de selección	Ruleta
Población elite	10
Probabilidad de cruza	0.8
Función de mutación	Uniforme
Criterio de parada	20 generaciones

La tabla **??** nos muestra el porcentaje promedio que aumentó cada una de las métricas de detección con la configuración del filtro seleccionada, cabe señalar que este aumento ya está representado en las tablas **??** y **??**.

En este experimento se constató la importancia de los filtros morfológicos en el desempeño del detector de objetos. El filtro morfológico en el método propuesto coadyuva a la eliminación de ruido y es necesario para unir mediciones que pertenezcan a un mismo objetivo. Con respecto a la tabla **??** observamos el gran impacto en el desempeño que genera un filtro morfológico con una configuración bien seleccionada, figura **??**. Se observó que la conjunción de dos

Tabla 5.4: Configuración y contribución de los filtros morfológicos.

Filtro morfológico	Accuracy	Recall	Precisión
Dilate (square 2)+ Close (rectangle 2-2)	4.18 %	7.99 %	4.59 %
Dilate (diamond 4)+ Close (rectangle 2-2)	**10.95 %**	**13.01 %**	**13.31 %**
Dilate (rectangle 4)+ Close (rectangle 2-2)	1.25 %	2.97 %	0.34 %

Figura 5.4: Preprocesamiento de las bases de datos.

filtros morfológicos elevó en gran medida el desempeño de la detección. La dilatación por un lado, aunque, maximiza el área del objeto, también permitió unir elementos del objeto que se encuentran separados. Por lo que, un valor pequeño de esta operación evita unir objetivos de interés cercanos. Por otra parte, la operación *close*, aplica una dilatación con una erosión y permitió cerrar huecos en los objetos.

Extracción de características de los objetos

Para la realización de los experimentos se consideraron las bases de datos definidas en la sección **??**. En el caso de los escenarios sintéticos, los datos fueron generados como vectores de posiciones, y sirvieron como entrada para el método propuesto. Sin embargo, para los videos se necesitó un preprocesamiento, que permitió determinar las observaciones de objetos.

En la figura **??** se muestran los pasos del preprocesamiento de los videos de la tabla **??** y la detección de objetos. Los cuáles se describen a continuación:

1. Se realizó un preprocesamiento del video, para enfatizar las características de los objetos. Esta fase es opcional, pero ayuda significativamente al

seguidor. Generalmente se ocuparon técnicas de ecualización del histograma y realzamiento de regiones. En concreto, para los experimentos con las bases de datos de la tabla **??**, se utilizó una ecualización automática del histograma, lo que permitió obtener una imagen con una distribución uniforme del histograma.

2. Posteriormente, para identificar los objetos se utilizó una técnica de modelado del fondo basado en mezcla de Gaussianas, la cual se presenta en [**?**], como resultado se obtiene una imagen binaria.

3. Se procesó la imagen binaria mediante filtros morfológicos, los cuáles permitieron resaltar características ocultas de los objetos. Ya que, los filtros morfológicos operan mediante ventanas que recorren la imagen y ejecutan un operador, para seleccionar la forma de la ventana, el tamaño de la ventana y el operador, se realizó un análisis de las diferentes configuraciones. Se implementó un algoritmo evolutivo que selecciona la configuración que minimiza el error de detección.

4. Por último un detector de *blobs*, recuperó los centroides de las regiones consideradas objetos. Los centroides y el largo del *blob* rectangular, fungirán como entradas para el método propuesto en este trabajo.

En la tabla **??** se presentan los resultados de la detección para cada base de datos, con el método de detección elegido y la configuración morfológica obtenida, tablas **??** y **??**.

Tabla 5.5: Detección de objetos para bases de datos reales.

BD	Frames	Objetos	Accuracy	Precision	Recall
Caltech modificado	500	6700	0.634	0.645	0.622
Ants	798	15960	0.565	0.692	0.617
Video Random	193	2547	0.632	0.686	0.586
Aves	548	9042	0.532	0.681	0.673
Extremas	430	23653	0.321	0.423	0.374

5.5.2. Caracterización de la oclusión

Por lo antes mencionado, la caracterización de la oclusión se basó en la definición de tres regiones (región de posible oclusión, región de oclusión y región de oclusión con un objeto no identificado). Estas tres regiones son muy importantes porque definieron el comportamiento del método propuesto, para adaptar las diferentes probabilidades e intensidades de GM-PHD.

Comportamiento de la caracterización de la oclusión para diferentes ϵ

Para la caracterización de la oclusión, se requirió determinar el valor de un umbral ϵ, que permitió determinar si un objeto se encuentra en estado ocluido. Como ya se mencionó, el tamaño de las regiones de oclusión y posible oclusión deben estar relacionadas directamente con el tamaño de los objetivos de interés. En las bases de datos, los objetivos pertenecen siempre a una misma clase, por lo tanto su tamaño no vario significativamente.

Se realizó un experimento donde se propusieron varios ϵ, todos en dependencia del largo del rectángulo de su *blob*, obtenido en la etapa de detección:

1. ϵ_1, es el tamaño del objeto más grande del frame actual.
2. ϵ_2, es el tamaño promedio de los objetos en el frame actual.
3. ϵ_3, es el tamaño promedio de los objetos en todos los frames anteriores.
4. ϵ_4 depende del tamaño de cada objeto, se calcula por cada objetivo.

Tabla 5.6: Comportamiento del método propuesto con diferentes ϵ.

	Tasa de oclusiones pérdidas (MOR)				
Tamaño de ϵ	**Caltech modificado**	**Ants**	**Video Random**	**Aves**	**Extremas**
ϵ_1	1.5 %	0.9 %	0	2.3 %	17 %
ϵ_2	0	0	0.3 %	0.5 %	9 %
ϵ_3	0	1.2 %	1.1 %	1.4 %	25 %
ϵ_4	0	0	0	0.2 %	5 %

En la tabla **??** se observa la evaluación de la tasa de oclusiones perdidas para el seguimiento de objetos con diferentes umbrales ϵ, los cuales permitieron construir regiones de oclusión entre objetivos de interés. Esta experimento, no solo nos permitió seleccionar un umbral que define las regiones de oclusión, sino también visualizar como la caracterización de la oclusión propuesta se desempeña (que tan fiable es la caracterización de la oclusión basada en regiones). Como era de esperar ϵ_4 tiene el mejor desempeño, sin embargo ϵ_2 se acerca mucho a sus capacidades.

Tabla 5.7: Tiempo de procesamiento del método por cada ϵ

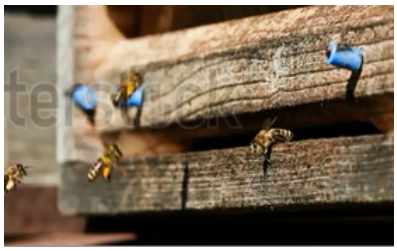
Epsilon	**ms/*frame***
ϵ_1	54
ϵ_2	62
ϵ_3	60
ϵ_4	134

Al usar ϵ_4 se necesita almacenar en un vector los estados de todas las observaciones y el tamaño de cada observación. De otro modo se tiene que calcular en la etapa de filtrado, perjudicando la eficiencia del método. En la tabla **??** podemos notar el tiempo promedio que tarda el método propuesto con cada epsilon, para ϵ_4 este tiempo supera por mucho a los otros.

Por otra parte, aunque el porcentaje de oclusiones pérdidas es pequeño, representa un número grande de oclusiones sin identificar; por ejemplo: en la columna Extremas el ϵ_2 tiene 9 %, que equivale a 84 oclusiones no identificadas vs 5 % de ϵ_4, que representa 46 oclusiones pérdidas.

De acuerdo a estas características se decidió sacrificar entre una menor precisión y un procesamiento más rápido. Por lo antes mencionado, se elige a ϵ_2, ya que es el tiempo de ejecución no incrementó demasiado y el desempeño del método es similar a ϵ_4.

(a) Escenario 1.



(b) Escenario 2.

(c) Escenario 3.

(d) Escenario 4.

Figura 5.5: Escenarios con regiones de oclusión con objetos no identificados.

Construcción de regiones de oclusión con objetos no identificados

Para caracterizar zonas en la imagen, donde desaparezcan objetos, se propusó construir regiones de oclusión con objetivos no identificados. Sin embargo, medir la fiabilidad de la construcción propuesta no fue algo sencillo, en primer lugar se necesitó identificar y definir estas zonas en las bases de datos, para posteriormente compararlas con las generadas por el método propuesto. Las comparaciones dependerán de la forma y tamaño de las áreas.

En las bases de datos existen 4 videos, los cuales contienen regiones de oclusión con un objeto no identificado. En la figura **??** se muestran la región de estos escenarios y en la tabla **??** las características de estos videos y sus regiones. Cabe señalar que el cálculo del tamaño de estas zonas se realizó con base en un rectángulo. La columna "Nacimiento de objetos"de la tabla **??** representa el número de objetivos que nacieron en esa región y que fueron detectados como objetos de interés correctamente.

En la tabla **??** se muestran importantes características de las regiones obtenidas mediante el método propuesto. Se encontraron 6 regiones (algunas fueron identificadas como 2 regiones diferentes), esto se debe principalmente a que

Tabla 5.8: Características de los escenarios.

Escenario	Base de datos	Tamaño de región2cm# de oclusiones2cmNacimiento objetos2cm		
1	Ants	50x782cm	1342cm	76 %2cm
2	Videos Random	45x2342cm	212cm	54 %2cm
3	Ants	76x1542cm	4022cm	74 %2cm
4	Videos Random	95x542cm	122cm	68 %2cm

los objetivos de interés desaparecieron en zonas focalizadas de la región que estaban separadas. En la tabla **??** la columna "Diferencia entre regiones", nos indica que tanto más pequeña es la región creada o que tanto más grande es. En el caso de números negativos significa que la región creada por el método propuesto es más pequeña por ese valor de pixeles en comparación de la extraída mediante la inspección visual; en caso que sea positiva significa que la región es más grande por ese valor de pixeles.

Tabla 5.9: Regiones obtenidas por el método propuesto.

Región	Escenario	Tamaño	Diferencia entre regiones
1	1	62X65	130
2	2	56x124	-2208
3		26x53	
4	3	23x45	-9291
5		31x43	
6	4	54x41	-2916

Por otra parte, la tabla **??** nos muestra la comparación entre las regiones creadas por el método propuesto y las extraídas de los escenarios. De esta tabla se consideraron los siguientes resultados obtenidos:

- Columna pixeles positivos: representa el porcentaje de pixeles de la rergión que fueron clasificados correctamente como región de oclusión con un objeto desconocido.

Tabla 5.10: Comparación entre regiones

Región	Pixeles positivos	Pixeles falsos negativos	Recall	Nacimiento objetos
1	86 %	12 %	87 %	98 %
2	76 %	40 %	66 %	78 %
3	83 %			
4	74 %	55 %	59 %	92 %
5	89 %			
6	95 %	59 %	61 %	94 %

- Columna pixeles falsos negativos: representa el porcentaje de la región de oclusión con un objeto no identificado que no fue descubierta por el método propuesto.
- Columna Recall: representa la métrica recall, $\frac{\text{positivos verdaderos}}{\text{positivos verdaderos + falsos negativos}}$.
- Columna nacimiento de objetos: representa el porcentaje de objetos encontrados satisfactoriamente que nacieron en esa región.

Cabe resaltar que la caracterización de regiones se adapta a través de la muerte repentina de objetivos de interés, por lo tanto, aunque exista una región de oclusión con un objeto no identificado, sino existen desapariciones de objetos de interés en ella, no será descubierta. Esto justifica que en los escenarios encontrados el valor de pixeles falsos negativos sea grande, en otras palabras no todos los objetos de interés recorren toda la zona de oclusión. Por otra parte, los objetivos de interés mueren en los límites de las zonas de oclusión y por lo tanto solo se utiliza la información de esa parte de la región, lo que provoca que el área entre las regiones descubiertas por el método propuesto y las generadas mediante inspección visual difieran en gran medida.

La tabla **??** nos muestra que las regiones descubiertas representan las regiones de oclusión con un objeto no identificado donde desaparecen más objetivos de interés, el alto porcentaje de pixeles positivos se traduce a una mejor adaptación de la probabilidad de nacimiento y en un aumento del porcentaje de detección de objetos que nacieron en esas zonas. Por último se realizó un análisis de los resultados obtenidos por cada escenario:

- Escenario 1: Se descubrió una sola región, la cual obtuvo los mejores resultados, un mayor número de pixeles positivos y un menor número de pixeles falsos negativos, esto se debe a que un gran número de objetivos fueron utilizados para construirla, 147. Además, es una región pequeña en comparación de las otras. Por otra parte, los objetivos de interés que nacieron en esa área fueron detectados correctamente en un 98 % de los casos.

- Escenario 2: Se descubrieron dos regiones, debido a que muy pocos objetos fueron utilizados y la región real es grande. La región 2 descubierta supera el alto de la región real, debido a que los objetos de interés desaparecen en los límites, y las regiones de posible oclusión asociados a ellos quedan fuera de la región, esto sucede en la mayoría de regiones construidas. Los objetivos morían en zonas diferentes de donde nacían, lo que provoca que el porcentaje de nacimiento sea menor para este escenario.

- Escenario 3: Se descubrieron 2 regiones en esta escena, sin embargo a diferencia de las regiones del escenario 2, los objetos de interés morían y nacían en las mismas áreas. Por lo tanto no se necesitó descubrir gran parte de la región para obtener resultados arriba del 90 %. Aunque esta región obtuvo el menor *Recall* de todas, su porcentaje de nacimiento fue uno de los más elevados. Esto se debe a que realmente no importa que se descubra todo el área de oclusión real, sino más bien es primordial que se descubran las zonas donde nacen los objetivos de interés.

- Escenario 4: Se descubrió una sola región, la cual a diferencia de las otras regiones descubiertas tiene la particularidad que los objetivos desaparecen una vez que entran completamente en la región de oclusión. Debido a esto la región descubierta es más pequeña. Por otro lado, en algunas ocasiones nacían dos nuevos objetos ocluyéndose mutuamente, cuando sucedía este fenómeno se tomó como un solo objeto, para definir el porcentaje de nacimiento de objetos. El método propuesto no puede detectar una oclusión que es generada por objetos que nacen.

5.5.3. Seguimiento en videos de objetos pequeños

En este grupo de experimentos, se utilizó para todos los videos, el método propuesto con una misma configuración de extracción de características y detección, desarrollada en la sección **??**.

Manejo de oclusión en escenarios reales

El manejo de la oclusión en escenarios reales es una de las principales contribuciones del método propuesto, para medir el desempeño en cada base de datos se aplican dos métricas: la tasa de exhaustividad (ReR) y la tasa de oclusiones pérdidas (MOR). Se utilizan las bases de datos de la tabla **??** para mostrar el comportamiento de los métodos en el seguimiento de objetos pequeños en estados de oclusión.

En la tabla **??** se muestran características importantes de las bases de datos para determinar sus comportamientos frente al problema de la oclusión (se tomó en cuenta solo las oclusiones entre objetivos de interés). El número de oclusiones generalmente depende del tamaño del escenario y el número de objetos en él. El tiempo promedio que permanece un objetivo en estado de oclusión ayudó a definir el umbral χ que decrementa la probabilidad de detección para un objeto ocluido. El promedio de objetos en oclusión por cada imagen, nos permite conocer la frecuencia en que suceden las oclusiones.

Tabla 5.11: Características de las oclusiones en las bases de datos

BD	Número de oclusiones	Tiempo promedio de oclusión	# de objetos promedio en oclusión
Caltech modificado	163	4 frames	2.13
Ants	668	7 frames	3.12
Video Random	383	6 frames	2.41
Aves	442	7 frames	3.25
Extremas	9351	12 frames	5.51

Por otra parte, la figura **??** nos muestra la evaluación de la métrica de tasa de falsas alarmas (FAR) para cada base de datos. Esta métrica se aplicó en

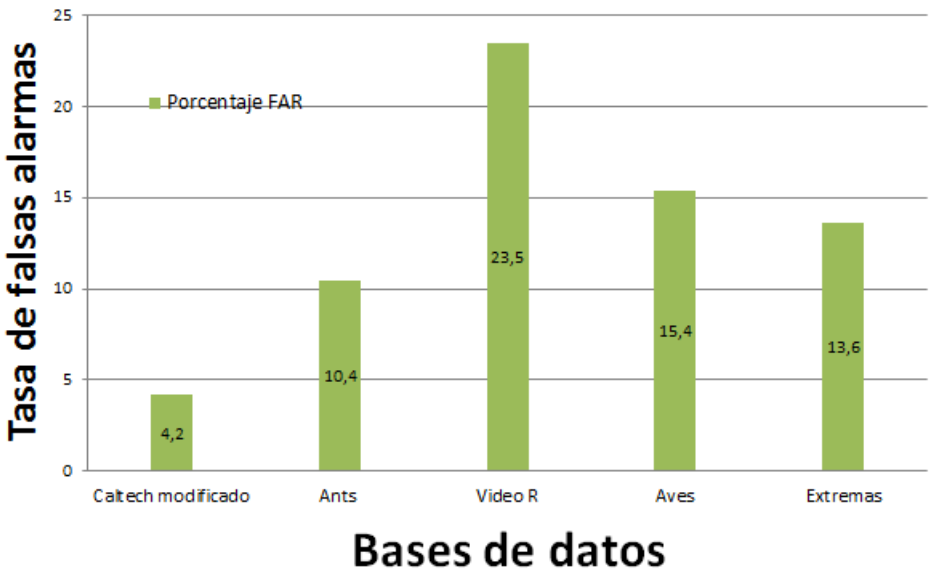


Figura 5.6: Tasa de falsas alarmas por bases de datos.

el método de detección para visualizar el porcentaje de ruido que contiene cada base de datos. Entre mas alto el porcentaje se concluye que el ruido en esa base de datos es mas denso, ya que se reconocen objetos de interés de observaciones que no lo son. Como se mencionó anteriormente la detección basada en modelado del fondo genera ruido en las mediciones, el cual se debe generalmente a un fondo que cambia dramáticamente, como es el caso de algunos escenarios de Ants y Video R.

En la tabla **??**, se evalúa MHT, CGM-PHD y la propuesta de este trabajo con base en la métrica ReR. En esta tabla se muestra la superioridad de la propuesta en comparación a MHT y CGM-PHD. En todas las bases de datos el método propuesto supera a MHT y CGM-PHD, incluso en la base de datos Extremas, el desempeño del método propuesto es mejor, lo cual es importante porque GM-PHD pierde efectividad cuando existen objetivos con movimiento que no son lineales. Por otra parte, la métrica ReR nos da una idea de cómo se comportan los métodos en el seguimiento de objetos, ya que engloba también objetivos que no se encuentran en estado de oclusión, esta métrica permite conocer la tasa de objetivos correctamente seguidos, pero no toma en cuenta ni los errores de distancia, ni de falsos positivos.

Tabla 5.12: Métrica ReR para base de datos reales.

Base de datos	MHT	CGM-PHD	Propuesta
Caltech modificado	0.812	0.834	**0.905**
Ants	0.794	0.726	**0.867**
Video R	0.651	0.873	**0.891**
Aves	0.783	0.764	**0.8653**
Extremas	0.589	0.481	**0.743**
Promedio	0.76	0.78	**0.88**

Si queremos profundizar en el comportamiento de los métodos en el proceso de oclusión, se necesita una métrica que evalué este fenómeno. Se utiliza la métrica de oclusiones perdidas, MOR. En la tabla **??** se puede ver la evaluación de la métrica MOR para cada método. El método propuesto supero a los otros dos y a MHT significativamente con un nivel de significancia del .03, vea apéndice **??**. La tasa de oclusiones perdidas para el método propuesto no es más grande del 1 % en casi todas las base de datos. Por otra parte, con una prueba de correlación lineal de Pearson se comprobó la relación entre el ruido en la escena (FAR) y el desempeño de MHT en la métrica MOR, estas dos variables tienen una correlación positiva con un nivel de significancia del 0.0163.

Si comparamos los resultados de nuestra propuesta con GM-PHD, podemos notar una gran diferencia, pese a ser métodos que se basan en el mismo algoritmo. Esto se debe principalmente a que CGM-PHD utiliza una normalización de pesos y en eventos de oclusión donde interactuan mas de un objetivo, aveces tiende a identificar menos objetivos. Se realizo una prueba de correlación lineal de Pearson, entre el promedio de objetivos ocluidos en las bases de datos y la tasa de oclusiones perdidas, obteniendo que estas variables tienen una correlación positiva con un nivel de significancia de 0.023.

El método propuesto resulto ser competitivo para el manejo de oclusiones en escenarios que asemejan a problemas reales (oclusión, fondo no uniforme, múltiples objetos pequeños, movimientos rápidos, entre otros). Incluso superan-

Tabla 5.13: Tasa de oclusiones perdidas para los diferentes métodos

Bases de datos	Tasa de oclusiones perdidas		
	MHT	**CGM-PHD**	**Propuesta**
Caltech modificado	0.3 %	0.4 %	**0 %**
Ants	2.2 %	1.1 %	**0 %**
Video Random	4.5 %	3.24 %	**0.3 %**
Aves	3.3 %	1.5 %	**0.5 %**
Extremas	19.4 %	21.6 %	**9 %**

do a uno de los algoritmos más utilizados en la práctica, MHT, algo que no pudo lograr la implementación de CGM-PHD en la base de datos Extremas, donde MHT obtuvo un porcentaje bajo por el ruido y el número de oclusiones y CGM-PHD por las oclusiones entre más de dos objetivos mal identificadas.

Seguimiento de objetos

En este experimento se evalúan los métodos MHT, CGM-PHD y el método propuesto con base en las métricas para seguimiento de múltiples objetos OSPA y MOTA. Estas métricas son las más utilizadas actualmente para evaluar algoritmos de seguimiento múltiple.

Para la métrica OSPA, se necesitan definir los parámetros p y c, los cuales impactan en la forma de como OSPA evalúa el desempeño del método. El parámetro de orden p determina la sensibilidad de la métrica en penalizar las estimaciones de valores atípicos. Generalmente el valor de p es 2 [**?**] [**?**], y depende de que tan rigurosa es la métrica con los errores de localización. El parámetro de corte c determina la penalización de los errores de cardinalidad, y su valor refleja la máxima distancia entre el objetivo de interés y el *gorund truth*. En la tabla **??** se puede observar la evaluación de los métodos con la métrica OSPA, con p=2 y los parámetros de corte en 5 y 10.

Cuando c=5, se puede observar que el método propuesto funciona mejor que MHT y CGM-PHD. La distancia media de OSPA para CGM-PHD es aproximadamente dos veces más grande y para MHT es aproximadamente

tres veces más grande, en comparación con el método propuesto. La diferencia entre MHT y el método propuesto se debe principalmente a la disparidad en la cardinalidad de sus estimaciones . Sin embargo, para CGM-PHD la diferencia de la distancia OSPA es aproximadamente 0.7 pixeles, ya que la elección de un pequeño corte indica a la métrica de no penalizar los errores de cardinalidad duramente y por lo tanto refleja principalmente errores de localización.

Por otra parte, cuando c=10, los errores de localización y cardinalidad están balanceados. Tanto la magnitud de los errores y la magnitud de la diferencia entre los errores han aumentado, debido a la elección de un punto de corte mayor, que tiene el efecto de equilibrio entre los errores de localización y de cardinalidad. En la distancia OSPA se tomaron en cuenta las estimaciones que no estuvieran alejadas más de 10 pixeles del centro del objetivo. En esta tabla podemos notar una gran ventaja del método propuesto frente a MHT, sin embargo en comparación con CGM-PHD la diferencia disminuye, lo que indica que la principal diferencia entre estos dos métodos es debido a errores de localización y no de cardinalidad.

En la tabla **??** se muestra la métrica MOTA para los diferentes métodos evaluados, los porcentajes del método propuesto contra CGM-PHD, demuestran que la adaptación propuesta del algoritmo GM-PHD se desempeña bien en problema de seguimiento con varios objetivos que emulan aplicaciones de vigilancia visual:

- Para la primera base de datos los tres algoritmos tuvieron un buen desempeño, superior al 90 %, los objetivos en esta base de datos se diferencian bien y existen menos oclusiones en comparación a las otra bases de datos.
- El método propuesto consiguió los mejores resultados en casi todas las bases de datos utilizadas. A diferencia de OSPA, la métrica MOTA toma en cuenta las perdidas en la secuencia y falsas correspondencias.
- En el caso de la base de datos Extremas, el método propuesto ocupo el segundo lugar, debajo de MHT. Esto se debe principalmente a que GM-PHD se desempeña bien en escenas con ruido, pero cuando existen dinámicas no lineales del movimiento pierde efectividad. Por otra parte,

MHT recupera mejor seguidores perdidos que GM-PHD, a un costo de elevar la complejidad del algoritmo.

Tabla 5.14: Comparación entre métodos con métrica MOTA

Bases de datos	MOTA		
	MHT	CGM-PHD	Propuesta
Caltech modificado	0.934	0.944	**0.955**
Ants	0.724	0.680	**0.842**
Video R	0.784	0.642	**0.808**
Aves	0.757	0.632	**0.830**
Extremas	**0.722**	0.472	0.682
Promedio	0.784	0.654	**0.823**

Tabla 5.15: Métrica OSPA para los diferentes métodos

Bases de datos	OSPA, p=2, c=10			OSPA, p=2, c=5		
	MHT	CGM-PHD	Propuesta	MHT	CGM-PHD	Propuesta
Caltech modificado	4.3	2.3	**1.8**	1.3	1.6	**0.3**
Ants	3.13	**2.1**	**2.1**	2.13	2.1	**0.7**
Video R	**2.2**	2.58	2.42	2.2	2.13	**0.64**
Aves	2.12	2.4	**1.86**	2.85	2.12	**0.53**
Extremas	4.36	4.1	**3.1**	3.34	2.1	**1.4**
Promedio	3.222	2.696	**2.256**	2.364	1.584	**0.714**

Comparación de tiempos entre algoritmos

Se realizó un experimento sobre el conjunto de bases de datos descritas en la sección **??** y con las características descritas en la sección **??**, para examinar la tasa de frames por segundo de los diferentes métodos utilizados. Para la fase de detección y extracción de características se utilizó la misma configuración para todos los algoritmos.

Tabla 5.16: Otras características de las bases de datos

BD/Características	Número de Cuadros	Tamaño imagen	Total de objetos (repetición)
Caltech modificado	500	1024x1024	6700
Ants	798	720x480	15960
Video Random	193	596x336	2547
Aves	548	596x336	9042
Extremas	430	596x336 (2), 720x582 (2)	23653

En la tabla **??** se pueden observar tres características de las bases de datos que influyen en el tiempo de ejecución de los algoritmos. Se toma en cuenta el número total de objetos, porque en algunos videos los objetivos salen del campo de visión del sensor (cámara).

En la tabla **??** podemos observar el tiempo en ms que tarda cada algoritmo en ejecutarse, y en la figura **??** cómo se expresa la relación de tiempo en *frames* por segundo. De estas dos tablas se puede deducir que las adaptaciones propuestas para GM-PHD, no influyen drásticamente en el tiempo. El método propuesto compite y supera al método CGM-PHD, el cual es una modificación que maneja la oclusión mediante la normalización de pesos de los objetivos en estado de oclusión. Comparando al algoritmo GM-PHD contra el método propuesto, se observa una ligera ventaja de este, pero hay que tomar en cuenta que GM-PHD no maneja oclusión de ningún tipo.

Por otro lado, se sabe que la complejidad computacional de los filtros PHD es mucho menor que MHT [**?**] [**?**], esto se enfatiza en la figura **??**. En general, el método propuesto puede manejar problemas complejos con un promedio de 20 fps independientemente del detector y el algoritmo de asignación de etiquetas. En arquitecturas de seguimiento con características semejantes, es un tiempo que se ajusta a problemas de vigilancia visual [**?**] [**?**].

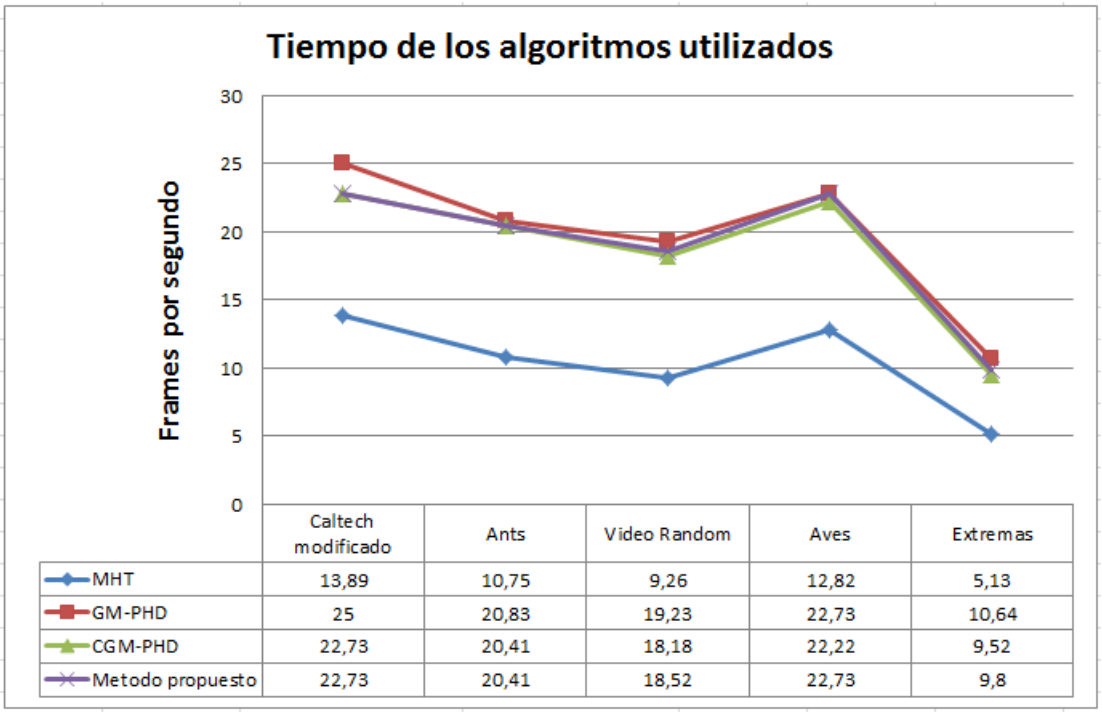


Figura 5.7: Comparación de tiempo de los algoritmos utilizados.

Tabla 5.17: Tiempo en ms de cada algoritmo

BD/Algoritmo	MHT	GM-PHD	CGM-PHD	Método propuesto	Detector
Caltech modificado	72.00	40.00	44.00	44.00	72.00
Ants	93.00	48.00	49.00	49.00	84.00
Video Random	108.00	52.00	55.00	54.00	72.00
Aves	78.00	44.00	45.00	44.00	81.00
Extremas	195.00	94.00	105.00	102.00	113.00

Asignación de etiquetas

En esta sección se examina como el algoritmo Húngaro se desenvuelve en conjunto con el método propuesto. También se mide cuanto influye la dirección y la velocidad para mejorar el proceso de asignación. La implementación de este mecanismo se baso en el programa desarrollado por Yi Cao en 2011, [**?**], su implementación es 3.5 veces mas rápida en comparación de la implementada por el software de MatLab.

Se utiliza el algoritmo Húngaro sin ninguna modificación, para posteriormente comparar los datos obtenidos con la propuesta de la sección **??**. Cabe señalar que se utiliza la misma configuración en todos los experimentos.

Tabla 5.18: Tasa de asignaciones para algoritmo Húngaro con impulso

	Tasa de asignaciones incorrectas (MLR)	
BD	**Húngaro**	**Húngaro con impulso**
Caltech modificado	**12.2 %**	**12.2 %**
Ants	41.4 %	**34.7 %**
Videos Random	47 %	**43 %**
Aves	31.5 %	**24.1 %**
Extremas	69.4 %	**54.6 %**

En la tabla **??** se observa como el algoritmo Húngaro con el impulso de dirección y velocidad promedio supera el rendimiento del algoritmo Húngaro. La métrica utilizada no solo verifica las pérdidas de etiqueta, sino que también la propagación del error, por ejemplo si un objetivo pierde su etiqueta o se le asigna otra errónea en un tiempo menor a otro objetivo, la primera influirá mayormente en incrementar la tasa de asignaciones incorrectas.

Se realizó un segundo experimento, en el cual no se consideraron las asignaciones erróneas generadas por oclusiones. En la tabla **??**, se muestra cómo es de esperar, que el desempeño del algoritmo incrementara, porque en las regiones de oclusión no existe información de las posiciones de los objetivos, y

ellos se pueden desplazar dentro de la región de oclusión sin ser eficientemente detectados.

Tabla 5.19: Tasa de asignaciones incorrectas sin tomar en cuenta la oclusión

	Tasa de asignaciones incorrectas (MLR)	
BD	**Húngaro**	**Húngaro con impulso**
Caltech modificado	**0 %**	**0 %**
Ants	**24 %**	**24 %**
Videos Random	**35 %**	**35 %**
Aves	24 %	**17 %**
Extremas	**36 %**	41 %

En la tabla **??** podemos notar cómo el método propuesto, no tiene una eficiencia mayor que el método Húngaro normal, en cambio en la base de datos Extremas parece perjudicarlo. Esto se debe, a que los objetivos cambian bruscamente su trayectoria, perjudicando tanto al algoritmo GM-PHD, como a la propuesta realizada para la asignación de etiquetas.

Por ultimo, en la tabla **??** se observa el tiempo de procesamiento en milisegundos del algoritmo Húngaro con impulso y sin impulso. En esta tabla se observa un incremento en el tiempo de procesamiento debido al impulso, el incremento va de 1 a 11 ms. Sin embargo, incluso con ese pequeño incremento se supera a MHT en todas las bases de datos, como se observa en la figura **??**.

Tabla 5.20: Tiempo de procesamiento en ms del algoritmo Húngaro con impulso

	Tiempo de procesamiento en ms	
BD	**Hungaro**	**Hungaro con impulso**
Caltech modificado	8.2	6.3
Ants	3.1	4.5
Video Random	12	17.1
Aves	15.5	16.4
Extremas	51.5	62.1

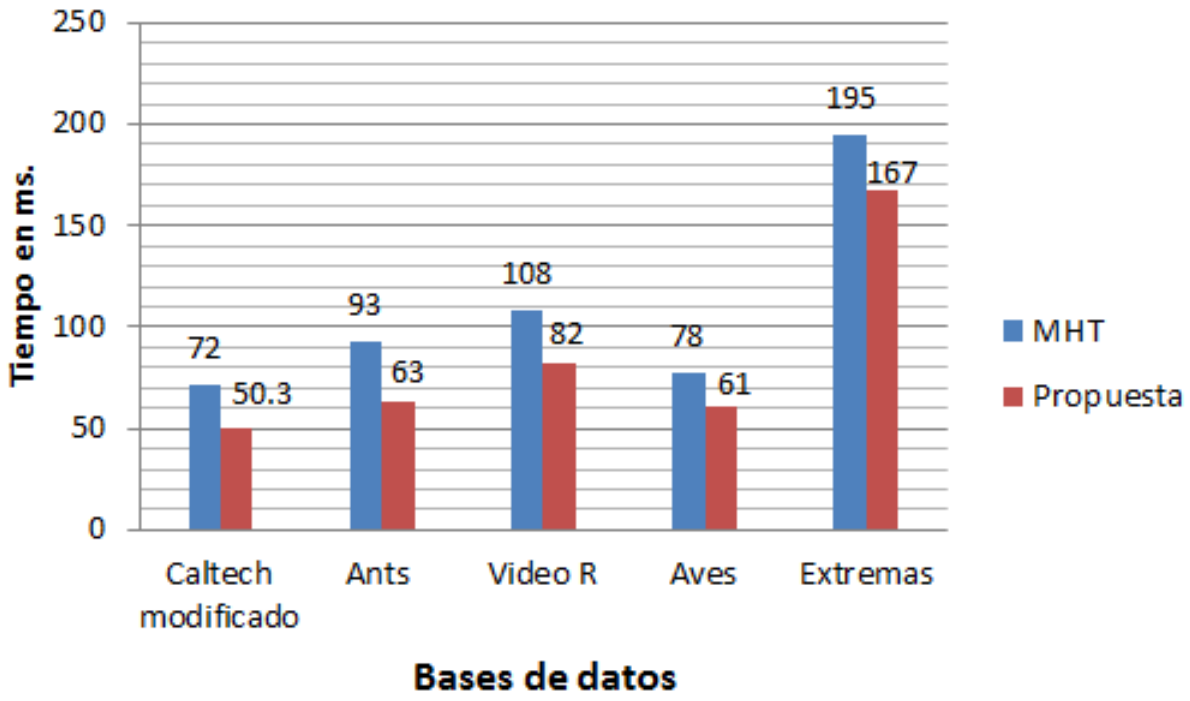


Figura 5.8: Tiempo en ms de procesamiento entre MHT y el método propuesto (seguimiento + asignación de etiquetas).

5.5.4. Seguimiento en escenarios sintéticos

Los escenarios sintéticos a diferencia de los datos obtenidos de videos, tienen algunos parámetros que pueden variar. Esto permite que se construyan a partir de ellos escenarios nuevos para analizar el comportamiento de los diferentes métodos de seguimiento. Los parámetros variables de los escenarios propuestos son los siguientes:

- Densidad del ruido: Este parámetro permite acondicionar el ruido en la escena en dependencia de los objetivos utilizados. Para los experimentos realizados se utilizaron 3 diferentes densidades de ruido 10, 25 y 50 unidades por objetivo de interés en la escena.

- Tamaño del escenario: El tamaño del escenario tienen una correlación directa con la densidad del ruido en la escena, cuando el escenario es más pequeño el ruido es más denso alrededor de los objetivos y complica el seguimiento.

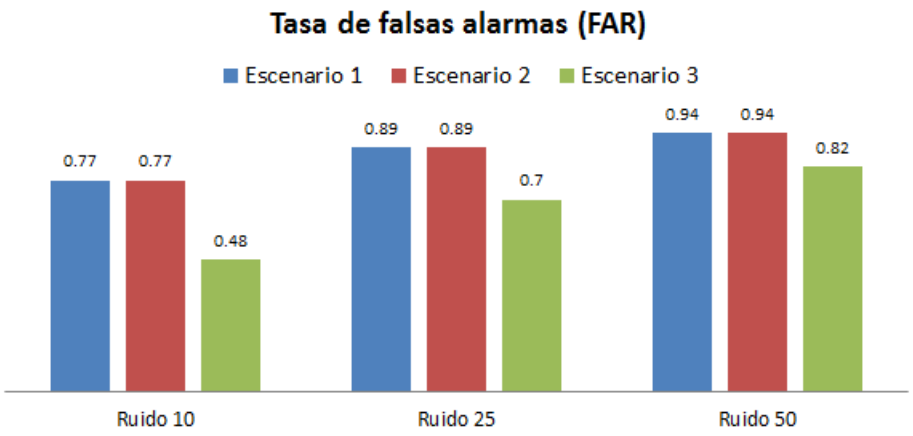


Figura 5.9: Métrica FAR para escenarios sintéticos.

- Oclusión de objetivos: Este parámetro permite modificar el tiempo que vive un objetivo, muere o entra en una zona de oclusión sea con otro objetivo de interés o con algún otro objeto.

Manejo de ruido en el método propuesto

Como se mencionó en la introducción, a partir de los escenarios base se pueden construir otros, cambiando parámetros como el ruido y el tamaño del escenario. En esta sección se realizan varios experimentos para evaluar el desempeño del método propuesto en escenarios a los cuales se les modifico las densidades de ruido.

A los escenarios sintéticos introducidos en la sección **??** se les acoplaron diferentes cantidades de señales de ruido. Los valores elegidos para los tres escenarios fueron 10, 25 y 50 unidades de ruido por cada instante de tiempo. La posición de cada señal de ruido se generó de forma aleatoria, con la *Toolbox* de MatLab, *Random Number Generation Tool*. Como las señales se generan aleatoriamente, el experimento se repitió 10 veces desechando los dos mejores y peores resultados.

La tasa de falsas alarmas de cada escenario, figura **??**, muestra que el ruido es elevado en comparación al número de objetivos. Se generó el ruido alrededor de los objetivos en un rango mínimo de 5 unidades y máximo de 30 unidades

para forzar una alta densidad de ruido, puesto que los escenarios son grandes. La figura **??** solo nos permite visualizar que tanta cantidad de ruido existe en comparación de los objetivos de interés, sabemos que este ruido es denso por las restricciones espaciales que se le añadieron.

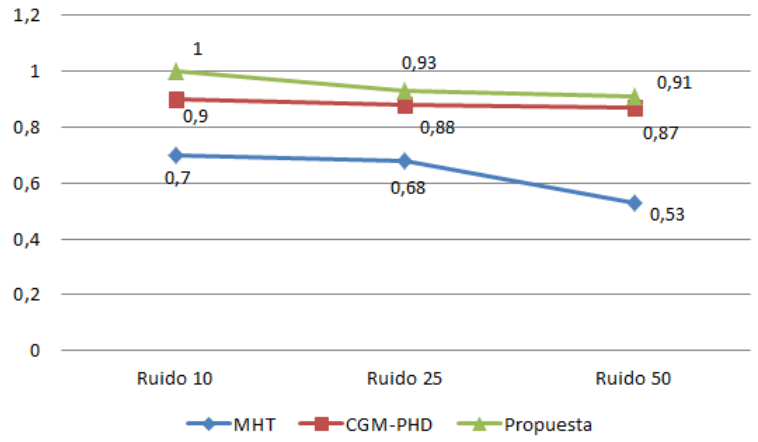
Cada método presenta diferentes grados de dificultad para manejar el ruido. En los escenarios 1 y 2 tratar con la densidad del ruido es más problemático porque existen menos objetivos de interés, esto limita al algoritmo MHT, el cual no es eficiente en escenarios con ruido. Por otro lado, los algoritmos basados en GM-PHD aunque manejan mejor el ruido comparado con MHT, sufren cuando existen oclusiones o movimientos no lineales.

En la figura **??**, se puede observar la evaluación de los métodos utilizados con la métrica MOTA, para los tres escenarios sintéticos con diferentes densidades de ruido. El método propuesto supera a los otros dos métodos, porque logra estimar con mayor exactitud a los objetivos. Comparado con MHT podemos observar una gran ventaja y significativa, vea el apendice **??**. Sin embargo con CGM-PHD la ventaja disminuye, debido a que están basados en el mismo algoritmo.

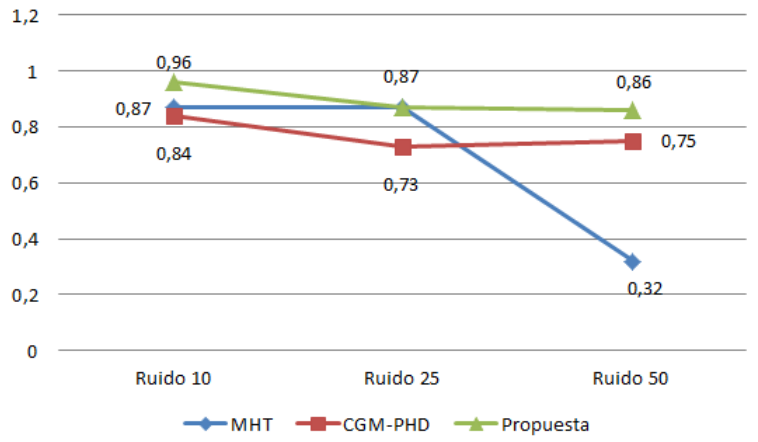
El método propuesto vence a CGM-PHD porque logra identificar a todos los objetivos en estado de oclusión, comparado a CGM-PHD que perdió el 54 %. El método propuesto puede seguir objetos en escenarios con altas densidades, con menos perdidas en la secuencia, menos falsos positivos y un número menor de falsas correspondencias a comparación de MHT y CGM-PHD.

Oclusión en escenarios sintéticos

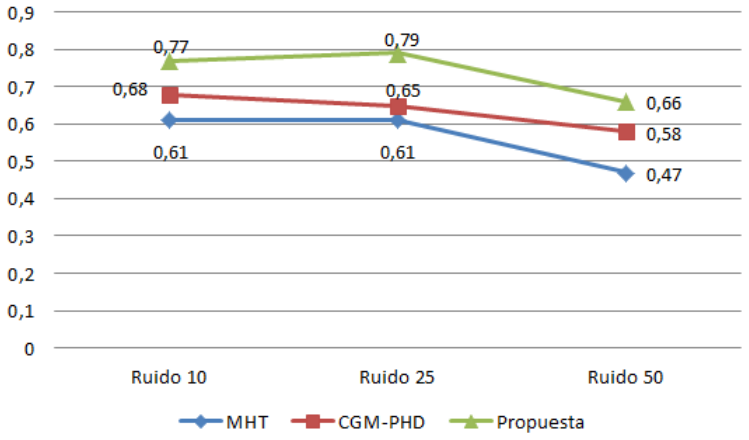
Para este experimento se utilizaron los escenarios sintéticos con dos adaptaciones, para la evaluación de los métodos en los siguientes problemas: oclusión entre objetivos y oclusión con ruido. Para evaluar el desempeño de los métodos elegidos frente al problema de oclusión se utiliza la métrica (MOR). La cual representa la tasa de oclusiones perdidas y mide la eficiencia en el manejo de la oclusión.



(a) MOTA para escenario 1.



(b) MOTA para escenario 2.



(c) MOTA para escenario 3.

Figura 5.10: Métrica MOTA para 3 escenarios con diferentes densidades de ruido.

Tabla 5.21: Tasa de oclusiones pérdidas para base de datos sintéticas.

	Tasa de oclusiones pérdidas (MOR)		
	MHT	**CGM-PHD**	**Propuesta**
Escenario 1	1	0	0
Escenario 2	1	.33	0
Escenario 3	.43	0	0

En la tabla **??** se muestra el experimento base, con los escenarios sin modificar, tal y como fueron presentados en la sección **??**. En esta tabla se puede observar como MHT no logra manejar la oclusión en escenarios con ruido. Por otra parte, el resultado en el escenario 2 para CGM-PHD nos muestra que perdió una de tres oclusiones. CGM-PHD también fue afectada por el ruido aleatorio y denso, puesto que ese escenario se eligió a partir de su trabajo de referencia [**?**].

Para medir la respuesta de los tres métodos se crearon escenarios a partir de los definidos en la sección **??**. En la tabla **??** se muestran las características del primer conjunto de modificaciones, realizadas con el objetivo de evaluar el desempeño de los métodos frente al problema de oclusión entre objetivos de interés. Cabe señalar que no se introdujo ruido alguno, con el fin de evaluar solo el manejo de oclusiones. A continuación se describen las características presentadas en la tabla **??**:

- Escenario base: indica el escenario utilizado como base para las modificaciones a realizar.

- Inicio oclusión: muestra el tiempo en que se iniciaron las oclusiones en ese escenario. La posición de cada valor en el vector funge como identificador de la oclusión, es decir en un vector [5, 12] el tiempo de inicio de la oclusión 1 sería 5 y el tiempo de inicio de la oclusión 2 sería 12.

- Fin oclusión: muestra el tiempo en que finalizaron las oclusiones, la posición de los valores en el vector identifican a la oclusión, de igual forma que en el vector de inicio de oclusión.

- Objetos ocluidos: muestra el número máximo de objetos que participaron en la oclusión, en algún intervalo de tiempo.

Por ejemplo, en el escenario base 2 se introdujo una oclusión con identificador 2 que inicio en 25 y termino en 30, en la cual participaron 3 objetos.

Tabla 5.22: Características de escenarios para evaluar la oclusión entre objetos

Escenario base	1	2	3
Inicio de oclusión	[10, 25, 45]	[10, 25, 45]	[10, 15, 15, 42, 40]
Fin de oclusión	[12, 30, 50]	[12, 30, 50]	[12, 20, 24, 45, 45]
Objetos ocluidos	[2, 3, 2]	[2, 3, 2]	[2, 3, 4, 2, 3]

En la figura **??** se puede observar la evaluación de la métrica MOR para los 3 escenarios modificados. El método propuesto supero a sus competidores, demostrando que la caracterización basada en regiones de la oclusión implementada es más efectiva. Puesto que no existe ruido en ningún escenario, se pueden examinar los algoritmos frente al problema de oclusión de una forma pura. Podemos visualizar un comportamiento similar entre MHT y CGM-PHD frente al problema de la oclusión, aunque son métodos completamente diferentes, esto se debe a que los dos métodos perdieron oclusiones donde participan más de dos objetos. CGM-PHD pierde todas las oclusiones donde más de dos objetos se ocluyen al mismo tiempo (4/28) y MHT perdió todas las oclusiones donde participan más de dos objetivo (9/28).

Por otra parte, en la figura **??** se observa el experimento realizado para evaluar las oclusiones en un ambiente con alta densidad de ruido. Los escenarios utilizados en este conjunto de experimentos fueron los mismos utilizados para la evaluación del seguimiento en escenarios con ruido, sección **??**. En esta figura, se observa que el método propuesto vence en la mayoría de los escenarios a sus competidores, gracias a que tiene un mejor manejo de la oclusión cuando existe ruido denso (tanto en escenarios sintéticos como reales).

En la figura **??**, MHT obtiene resultados malos relacionados a su capacidad de resolver oclusiones en escenarios con altas densidades de ruido, de hecho en

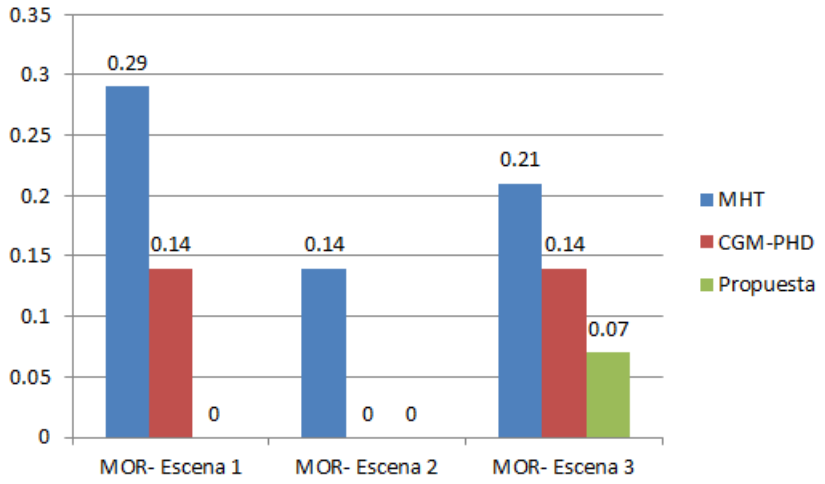


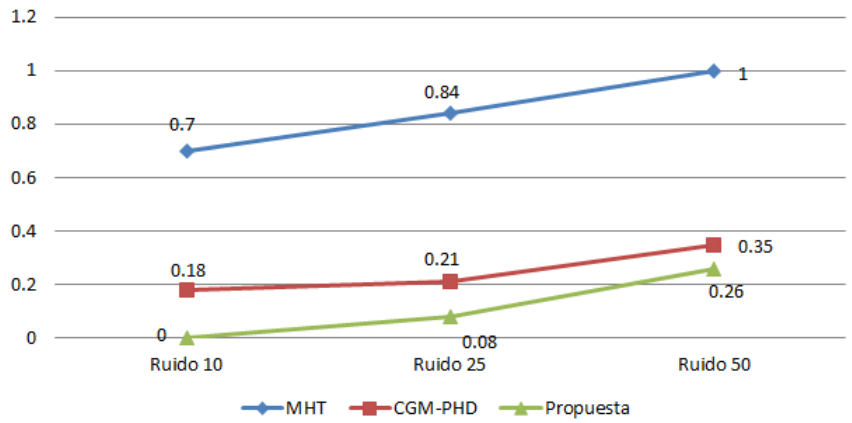
Figura 5.11: Métrica MOR para 3 escenarios sintéticos modificados.

ocasiones ni siquiera resolvía la oclusión pero identificaba ruido cercano que indujo a la métrica a clasificarla como una correcta identificación de la oclusión. También podemos notar que tanto CGM-PHD, como el método propuesto obtienen buenos resultados con base en la métrica MOR, aunque estos no son concluyentes para determinar un claro vencedor, ya que siguen la misma tendencia con una diferencia mínima. Por ejemplo en la **??**, cuando la cantidad de ruido es 25 unidades, la diferencia entre los dos métodos es de una sola oclusión pérdida del total. Por otra parte, la máxima diferencia obtenida entre los dos métodos se aprecia en **??** con ruido de 25 unidades, la cual representa 9 oclusiones.

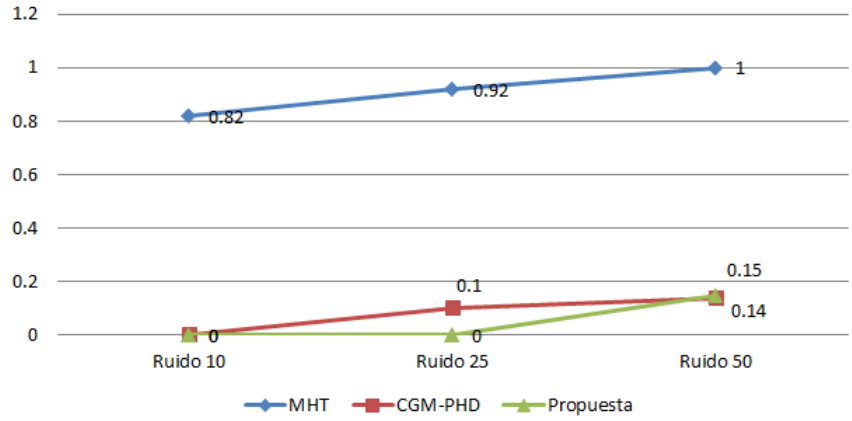
Seguimiento de objetos

En este experimento se utilizaron los escenarios sintéticos, con el propósito de comparar el método propuesto en un mismo marco de evaluación que utilizaron sus principales competidores. Se puede comparar directamente nuestro método con los resultados obtenidos por el método *Competitive Gaussian mixture probability hypothesis density filter* (CGM-PHD) en [**?**].

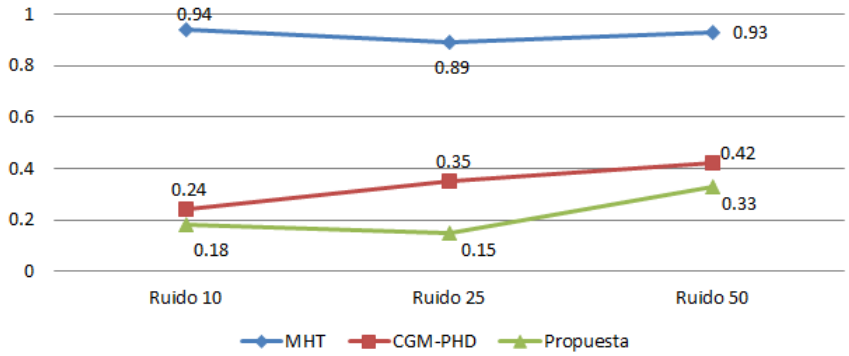
En la tabla **??** se exploran dos métricas definidas en secciones anteriores que evalúan el desempeño del método en el seguimiento de múltiples objetivos. Cabe señalar que la tabla de resultados muestran el experimento base, donde



(a) MOR para escenario 1.



(b) MOR para escenario 2.



(c) MOR para escenario 3.

Figura 5.12: Métrica MOR para 3 escenarios con diferentes densidades de ruido.

Tabla 5.23: Evaluación de los algoritmos con escenarios sintéticos

	Distancia media (OSPA)			MOTA		
	MHT	**CGM-PHD**	**Propuesta**	**MHT**	**CGM-PHD**	**Propuesta**
Escenario 1	55.1	**23.3**	**13.3**	.66	**.88**	**.88**
Escenario 2	62.2	24.6	**15.4**	.67	.72	**.79**
Escenario 3	121.8	64.9	**15.7**	.56	.72	**.83**

se generan 25 señales de ruido por cada instante de tiempo, los parámetros de la métrica OSPA son: p=2 y c=200.

El error de la distancia OSPA entre los objetivos y sus estimaciones no varía mucho en los dos primeros escenarios, ya que son muy similares. En el escenario 3 se ve una marcada diferencia entre el método propuesto y los otros métodos, provocada por el número de oclusiones, el número de objetivos y la dinámica de sus movimientos.

La métrica MOTA, muestra una variación entre el escenario 1 y 2, generada por la aparición en la oclusión de un nuevo objetivo en el tiempo 25. Se muestra la capacidad superior de nuestra propuesta para identificar nuevos objetivos, producto de la adaptación del nacimiento de objetivos. Aunque nuestra propuesta supera en la métrica MOTA y en todos los escenarios a CGM-PHD, los valores no son concluyentes para asegurar una mejora significativa.

Por último en la figura **??** se muestra la evaluación de la métrica MLR con el propósito de comparar los métodos de asignación de etiquetas de MHT contra el método propuesto. En esta figura se presentan los resultados de los métodos para 6 escenarios:

1. MHT Oclusiones y Propuesta Oclusiones: Representan los valores del método MHT y el método propuesto para los escenarios que tienen las características de la tabla **??**, los 3 escenarios sin ruido (MLR-Escena 1, MLR-Escena 2 y MLR-Escena 3).

2. MHT Ruido y Propuesta Ruido: Representan los valores obtenidos de evaluar la métrica MLR en los 3 escenarios definidos en la sección **??** con

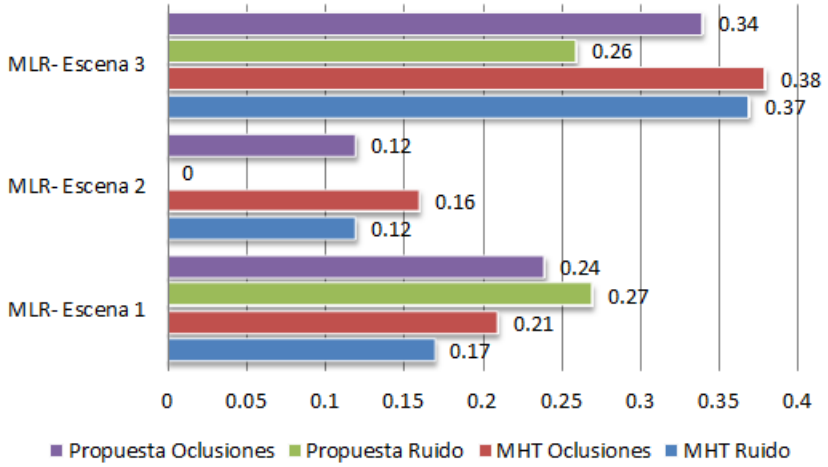


Figura 5.13: Métrica MLR para escenarios sintéticos modificados.

un ruido de 10 unidades (MLR-Escena 1, MLR-Escena 2 y MLR-Escena 3).

El algoritmo Húngaro logro mayor tasa de asignaciones correctas durante mayor tiempo en comparación de MHT, en gran parte debido a las virtudes de GM-PHD y el método propuesto para el manejo de oclusión y ruido. Tanto MHT como el método propuesto registran una menor eficiencia cuando se asocian datos en problemas con oclusiones.

Parte VI

Conclusiones y trabajo futuro

Capítulo 6

Conclusiones y trabajo futuro

6.1. Conclusiones

En este trabajo de tesis se presentó un método para el manejo de oclusión y un módulo de asignación de etiquetas para el filtro Gaussian Mixture Probability Hypothesis Density (GM-PHD). Este tipo de filtro se caracteriza por disminuir la complejidad en el proceso de seguimiento de objetivos, eliminando la asociación de datos, y por lo tanto no considerando el manejo de oclusión. Por otra parte, GM-PHD es robusto en el manejo de ruido en comparación con MHT.

Durante todo el trabajo presentado, se enfatizó la importancia del seguidor sobre el detector, construyendo una arquitectura simple pero funcional. El manejo de un método basado en el modelado del fondo, Gaussian Mixture Model [**?**], permite detectar objetos en movimiento y genera ruido el cual sirve para evaluar la propuesta realizada. La utilización de filtros morfológicos coadyuva a la creación de regiones de oclusión, evitando que un objeto de interés genere dos mediciones. Por otra parte el uso de escenarios sintéticos permite un mejor análisis de las virtudes del método propuesto, ya que sus parámetros se pueden modificar.

A lo largo de los experimentos realizados, se mostró una superioridad del método propuesto en comparación con MHT, y CGM-PHD. Hasta hace un lustro MHT era considerado uno de los métodos más usados para aplicaciones prácticas de seguimiento múltiple de objetos. Por otra parte, CGM-PHD es

una adaptación novedosa del filtro GM-PHD que permite manejar la oclusión mediante la normalización de los estados.

El método propuesto en este trabajo se evaluó en 4 áreas importantes para cualquier método de seguimiento: detección y manejo de oclusión, seguimiento de objetos, manejo de escenarios con ruido y asignación de etiquetas. Los resultados obtenidos a partir de estas evaluaciones posicionan a esta propuesta como un método competitivo y eficaz. A continuación se presentarán un conjunto de análisis sobre las áreas expuestas anteriormente.

Como se introdujo en la sección de diseño, el método propuesto caracteriza la oclusión con base en la construcción de regiones. Este tipo de caracterización de la oclusión permite trabajar con pequeños objetos, ya que su tamaño es el factor que se utiliza para construir regiones de posible oclusión. Cuando se construyen regiones de oclusión no importa cúantos objetos se encuentren en ella, el método propuesto conservara su seguidor hasta que reaparezcan en la escena o mueran con base en la disminución de su probabilidad de detección. La utilización de regiones demostró ser una herramienta eficaz para disminuir la tasa de oclusiones perdidas y elevar la efectividad del seguimiento de objetivos de interés.

Una vez caracterizada la oclusión, se adaptan las probabilidades de detección, de supervivencia y se incorpora información espacial de las regiones de oclusión en las Gaussianas de las intensidades de nacimiento y supervivencia; siempre en dependencia de las regiones descubiertas. Estos pasos son las adaptaciones realizadas a GM-PHD. De acuerdo a los resultados obtenidos se deduce que las modificaciones realizadas no repercuten en el tiempo de ejecución, debido a que el proceso más pesado es en realidad el cálculo de las regiones, que una vez obtenidas modifican las probabilidades e intensidades, mediante asignaciones a las variables utilizadas en la recursión del filtro.

Para evaluar el método se utilizaron dos tipos de escenarios, sintéticos y reales. Nuestro diseño demostró de acuerdo a estas evaluaciones trabajar bien y superar a los métodos de comparación, en los dos tipos de escenarios. Las aplicaciones de este método pueden configurarse para distintos tipos de

problemas, desde la vigilancia visual hasta el análisis de comportamiento de objetos. El propósito de este trabajo fue alcanzado al demostrar la efectividad del método propuesto en el seguimiento en videos de insectos, superando a métodos dedicados al seguimiento múltiple. Sin embargo, al priorizar al seguidor, se utilizaron escenarios con características similares y las aplicaciones practicas se acotaron a escenarios con las características mencionadas en la introducción. Por otra parte, el uso de GM-PHD como base de las adaptaciones permite acoplar el uso del método a diversos escenarios, ya que se obtuvieron tiempos de ejecución pertinentes a una amplia gama de aplicaciones.

Se puede observar que nuestra propuesta supera a sus más cercanos competidores, en los escenarios sintéticos, demostrando que la propuesta desarrollada permite el manejo de la oclusión y el seguimiento de objetivos en escenarios con altas densidades de ruido. Tenemos que recordar que los objetivos en las bases de datos utilizadas siguen una dinámica del movimiento lineal, lo que exalta las virtudes del algoritmo GM-PHD.

Por último, el método de asignación de etiquetas permite a nuestra propuesta identificar objetos en la escena. Se utilizó el algoritmo Húngaro de asignación, al cual se le incorporo información de la velocidad y dirección de los objetivos, con el fin de poder recuperar su etiqueta una vez finalizada una oclusión. El nuevo método Húngaro con impulso elevo el porcentaje de asignaciones correctas en estado de oclusión; sin embargo, cuando no se toma en cuenta la oclusión, en algunas bases de datos por el movimiento no lineal de los objetos perdió efectividad. Comparado con MHT, el algoritmo Húngaro con impulso realiza las asignaciones de etiquetas en un tiempo menor, con mayor eficiencia; sin embargo, los dos métodos tienen problemas cuando existe oclusión entre objetivos.

6.2. Trabajo futuro

Las principales direcciones para explotar las capacidades del método propuesto y las pautas para continuar el trabajo, se presentan en la siguiente lista:

- Examinar las virtudes y evaluar las capacidades del método propuesto con detectores que sean capaces de identificar objetivos ocluidos parcialmente (modelos de apariencia, patrones, etc.). Se puede plantear una arquitectura similar a [**?**], donde se utilice la información del detector para coadyuvar a la eliminación de ruido y mejorar la asociación de etiquetas. Esto permitirá utilizar el método propuesto en escenarios menos controlados, por ejemplo en multitudes.

- La utilización de múltiples sensores (cámaras), permitirá al método propuesto trabajar en aplicaciones donde se necesite realizar el seguimiento en 3 dimensiones. También ayudaría a combatir el problema de la oclusión, con la incorporación de diferentes vistas del proceso. El manejo de múltiples sensores en los filtros PHD es una de las principales áreas de investigación de los últimos años [**?**] [**?**].

- Mejorar la efectividad del etiquetado, mediante la incorporación de un método de asociación de datos especializado para el marco de seguimiento múltiple Bayesiano, en concreto para el algoritmo GM-PHD. Generalmente se puede asignar un identificador único a cada Gaussiana, pero es necesario construir mecanismos de poda para disminuir la complejidad del proceso [**?**].

- El filtro GM-PHD teóricamente no puede procesar modelos dinámicos con ruido no Gaussiano, por lo tanto se busca incorporar a GM-PHD un modelado de la dinámica de los objetivos y observaciones no lineales para aumentar la efectividad del método propuesto al seguir objetivos que tienen movimientos no lineales. Esto se puede realizar a través de métodos de integración numérica (métodos de cuadratura) o aproximaciones de linealización del sistema (expansiones de Taylor). Se han propuesto utilizar extensiones para el filtro de GM-PHD con el filtro extendido de Kalman (EK-PHD) o el filtro Unscented Kalman (UK-PHD) [**?**] [**?**].

6.3. Revisión de objetivos

- Se logró definir una arquitectura para el seguimiento de múltiples objetos en condiciones de oclusión y ruido, identificando su identidad con un algoritmo de asignación de etiquetas.
- Se implementó un algoritmo de detección y extracción de características para evaluar el método propuesto en problemas reales.
- Se diseñó un conjunto de bases de datos para evaluar el método propuesto en problemas de diferentes grados de oclusión y ruido.
- Se evaluó el desempeño de este método para objetos pequeños, con las principales métricas utilizadas.
- Se implementó el método propuesto, permitiendo analizar su desempeño en varios escenarios sintéticos y reales.
- Se comparó el método propuesto contra sus principales competidores CGM-PHD y MHT, logrando mejores resultados.
- El método propuesto se adaptó a problemas reales, y no requiere información del estado de los objetivos para su inicialización.
- El método de seguimiento propuesto independientemente del detector y el algoritmo de asignación de etiquetas, puede adaptarse a aplicaciones reales con características similares a las mencionadas en el planteamiento del problema, la tasa de frames por segundo alcanzada por el método propuesto depende del número de objetivos.

Apéndice A

Pruebas de significancia

Para comprobar los resultados obtenidos por el método propuesto son estadísticamente significativos, se realizan un conjunto de pruebas de significancia para la mayoría de las tablas mostradas en la sección de resultados.

Se utiliza para realizar las pruebas el método de Wilcoxon. Donde se verifica la hipótesis, que postula que las muestras proceden de poblaciones con la misma distribución de probabilidad; la hipótesis alternativa establece que hay diferencias.

Se verifica el comportamiento de las diferencias entre los valores de los elementos de cada par asociado, teniendo en cuenta el signo y la magnitud de la diferencia. Por último, la decisión de validar o rechazar la hipótesis se toma utilizando el valor p: si el valor p es inferior al nivel de significación, entonces la hipótesis nula es rechazada. Cuanto menor sea el valor p, más significativo será el resultado.

Las pruebas se realizaron con la implementación en MatLab de la suma de rangos de Wilcoxon, que se encuentra en la Toolbox de estadística. Esta implementación recibe como parámetros dos conjuntos de datos, regresa el valor de p y el resultado de la verificación de la hipótesis. A continuación se muestran los resultados de las pruebas aplicados a las tablas.

En las tablas **??** y **??** se utiliza la columna ”Nacimiento de objetos”para evaluar el desempeño del método propuesto en el nacimiento adaptable de objetivos de interes en regiones de oclusión con objetivos no identificados. Los

resultados obtenidos de la prueba de significancia son los siguiente: la hipótesis nula fue rechazada para los niveles de significancia por arriba de p=0.0286.

Por otra parte, la tabla **??** nos muestra la evaluación de la métrica ReR, para los diferentes métodos seleccionados. La prueba de significancia para esta tabla obtuvo los siguientes resultados: la hipótesis nula fue rechazada en p=0.0745 para el método propuesto frente a MHT; por otra parte en la prueba entre el método propuesto y CGM-PHD se rechazó la hipótesis nula en p=0.0952.

En la tabla **??** se muestra la evaluación de la métrica MOR para MHT, CGM-PHD y el método propuesto. Los resultados de las pruebas de significancia para esta tabla son: la hipótesis nula se rechazó con p=0.0397 entre MHT y la propuesta; entre la propuesta y CGM-PHD la hipótesis nula se rechazó en p=0.1429 y no alcanzo el nivel de significancia de 0.05.

En la tabla **??**, se muestra la comparación entre los diferentes métodos con la métrica OSPA. Cuando el punto de corte en OSPA es 10, no existen diferencias significativas entre las evaluaciones realizadas. Mientras que cuando el punto de corte es 5, el nivel de significancia alcanzado es 0.0159 y 0.0079, para MHT y CGM-PHD respectivamente.

Por otra parte, significativamente el método propuesto no supera en tiempo de procesamiento a GM-PHD y CGM-PHD; sin embargo, MHT es superado con un nivel de significancia de 0.051, tabla **??**.

En la figura **??** se presentan las evaluaciones de la métrica OSPA en 3 escenarios con diferentes densidades de ruido. En estas evaluaciones, el metodo propuesto logro superar a MHT con un nivel de significancia de 0.0073. Por otra parte, los mismos escenarios utilizados en la figura **??**, se ocupan para evaluar la métrica MOR. En la figura **??** se pueden observar los resultados de este experimento, donde el metodo propuesto supero a MHT con un nivel de significancia del 0.00004.

En la tabla **??** se evalúan diferentes escenarios sintéticos en base a dos métricas para el seguimiento múltiple, OSPA y MOTA. Con respecto a la métrica

OSPA no existe una diferencia significativa, aunque el método propuesto obtuvo mejores resultados. En el caso de la métrica MOTA, solo alcanzo un nivel de significancia robusto la comparación entre MHT y el metodo propuesto.

Bibliografía

[a.G.a. Perera et al., 2006] a.G.a. Perera, Srinivas, C., Hoogs, a., and Brooksby, G. (2006). Multi-Object Tracking Through Simultaneous Long Occlusions and Split-Merge Conditions. *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1 (CVPR'06)*, 1:666–673.

[Bar-Shalom et al., 2009] Bar-Shalom, Y., Daum, F., and Huang, J. (2009). The probabilistic data association filter. *Control Systems, IEEE*, 29(6):82–100.

[Baraldi and Parmiggiani, 1995] Baraldi, A. and Parmiggiani, F. (1995). An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters. *Geoscience and Remote Sensing, IEEE Transactions on*, 33(2):293–304.

[Betke et al., 2007] Betke, M., Hirsh, D., Bagchi, a., Hristov, N., Makris, N., and Kunz, T. (2007). Tracking Large Variable Numbers of Objects in Clutter. *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8.

[Bouwmans et al., 2008a] Bouwmans, T., El Baf, F., and Vachon, B. (2008a). Background modeling using mixture of gaussians for foreground detection-a survey. *Recent Patents on Computer Science*, 1(3):219–237.

[Bouwmans et al., 2008b] Bouwmans, T., El Baf, F., and Vachon, B. (2008b). Background modeling using mixture of gaussians for foreground detection-a survey. *Recent Patents on Computer Science*, 1(3):219–237.

[Bredereck et al., 2012] Bredereck, M., Jiang, X., Marco, K., and Denzler, J. (2012). Data Association for Multi-Object Tracking-by-Detection in Multi-Camera Networks.

[Cao, 2011] Cao, Y. (2011). Munkres' assignment algorithm-modified for rectangular matrices.

[Chen et al., 2011] Chen, L., Chen, Z., and Yin, F. (2011). A novel merging algorithm in gaussian mixture probability hypothesis density filter for close proximity targets tracking⋆.

[Cheng et al., 2011] Cheng, Y., Demars, K. J., and Fr, C. (2011).

[Clark and Bell, 2006] Clark, D. and Bell, J. (2006). Convergence results for the particle phd filter. *Signal Processing, IEEE Transactions on*, 54(7):2652–2661.

[Clark et al., 2006] Clark, D., Panta, K., and Vo, B.-N. (2006). The gm-phd filter multiple target tracker. In *Information Fusion, 2006 9th International Conference on*, pages 1–8. IEEE.

[Clark and Vo, 2007] Clark, D. and Vo, B.-N. (2007). Convergence analysis of the gaussian mixture phd filter. *Signal Processing, IEEE Transactions on*, 55(4):1204–1212.

[Clark et al., 2013] Clark, D., Vo, B.-t., and Vo, B.-n. (2013). Gaussian Mixture Implementations of Probability Hypothesis Density Filters for Non-linear Dynamical Models.

[Clark and Bell, 2005a] Clark, D. E. and Bell, J. (2005a). Bayesian multiple target tracking in forward scan sonar images using the phd filter. *IEE Proceedings-Radar, Sonar and Navigation*, 152(5):327–334.

[Clark and Bell, 2005b] Clark, D. E. and Bell, J. (2005b). Data association for the phd filter. In *Intelligent Sensors, Sensor Networks and Information Processing Conference, 2005. Proceedings of the 2005 International Conference on*, pages 217–222. IEEE.

[Colegrove et al., 1996] Colegrove, S. B., Davis, L., and Davey, S. J. (1996). Performance assessment of tracking systems. In *Signal Processing and Its Applications, 1996. ISSPA 96., Fourth International Symposium on*, volume 1, pages 188–191. IEEE.

[Cox et al., 1993] Cox, I., Rehg, J., and Hingorani, S. (1993). A bayesian multiple-hypothesis approach to edge grouping and contour segmentation. *International Journal of Computer Vision*, 11(1):5–24.

[Daum, 2005] Daum, F. (2005). Nonlinear filters: beyond the kalman filter. *Aerospace and Electronic Systems Magazine, IEEE*, 20(8):57–69.

[de Poortere et al., 2002] de Poortere, V., Cant, J., Van den Bosch, B., De Prins, J., Fransens, F., and Van Gool, L. (2002). Efficient pedestrian detection: a test case for svm based categorization. In *Workshop on Cognitive Vision*.

[Fry et al., 2000] Fry, S. N., Bichsel, M., Müller, P., and Robert, D. (2000). Tracking of flying insects using pan-tilt cameras. *Journal of Neuroscience Methods*, 101:59–67.

[Fry et al., 2008] Fry, S. N., Rohrseitz, N., Straw, A. D., and Dickinson, M. H. (2008). TrackFly: Virtual reality for a behavioral system analysis in free-flying fruit flies. *Journal of Neuroscience Methods*, 171:110–117.

[Fuchs and Padgett, 2012] Fuchs, T. J. and Padgett, C. (2012). Recognizing Humans in Motion : Trajectory-based Aerial Video Analysis. pages 1–11.

[Hussein et al., 2012] Hussein, I. I., DeMars, K. J., Fruh, C., Erwin, R. S., and Jah, M. K. (2012). An aegis-fisst integrated detection and tracking approach to space situational awareness. In *Information Fusion (FUSION), 2012 15th International Conference on*, pages 2065–2072. IEEE.

[JingJing and ShiQiang, 2010] JingJing, W. and ShiQiang, H. (2010). The probability hypothesis density filter based multi-target visual tracking.

[Jonker and Volgenant, 1986] Jonker, R. and Volgenant, T. (1986). Improving the hungarian assignment algorithm. *Operations Research Letters*, 5(4):171–175.

[Juan and Gwun, 2009] Juan, L. and Gwun, O. (2009). A comparison of sift, pca-sift and surf. *International Journal of Image Processing (IJIP)*, 3(4):143–152.

[Keni and Rainer, 2008] Keni, B. and Rainer, S. (2008). Evaluating multiple object tracking performance: the clear mot metrics. *EURASIP Journal on Image and Video Processing*, 2008.

[Lee, 2005a] Lee, D.-S. (2005a). Effective gaussian mixture learning for video background subtraction. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(5):827–832.

[Lee, 2005b] Lee, D.-S. (2005b). Effective gaussian mixture learning for video background subtraction. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(5):827–832.

[Li et al., 2014] Li, R., Huang, X., Lu, R., and Shen, L. (2014). Infrared small target tracking algorithm based on fusion feature matching and mean shift correction. In Patnaik, S. and Li, X., editors, *Proceedings of International Conference on Computer Science and Information Technology*, volume 255 of *Advances in Intelligent Systems and Computing*, pages 595–603. Springer India.

[Li et al., 2013] Li, X., Hu, W., Shen, C., Zhang, Z., Dick, A., and Hengel, A. V. D. (2013). A survey of appearance models in visual object tracking. *ACM Trans. Intell. Syst. Technol.*, 4(4):58:1–58:48.

[Lin et al., 2004] Lin, L., Bar-Shalom, Y., and Kirubarajan, T. (2004). Data association combined with the probability hypothesis density filter for multitarget tracking. In *Defense and Security*, pages 464–475. International Society for Optics and Photonics.

[Lowe, 2004] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110.

[Maggio and Cavallaro, 2011] Maggio, E. and Cavallaro, A. (2011). *Video Tracking: Theory and Practice*. Wiley.

[Maggio et al., 2008] Maggio, E., Taj, M., and Cavallaro, A. (2008). Efficient multitarget visual tracking using random finite sets. *Circuits and Systems for Video Technology, IEEE Transactions on*, 18(8):1016–1027.

[Mahler, 2007a] Mahler, R. (2007a). Phd filters of higher order in target number. *Aerospace and Electronic Systems, IEEE Transactions on*, 43(4):1523–1543.

[Mahler, 2007b] Mahler, R. P. (2007b). *Statistical multisource-multitarget information fusion*, volume 685. Artech House ˆ eBoston Boston.

[Mahler and Zajic, 2001] Mahler, R. P. and Zajic, T. (2001). Multitarget filtering using a multitarget first-order moment statistic. In *Aerospace/Defense Sensing, Simulation, and Controls*, pages 184–195. International Society for Optics and Photonics.

[Mahler, 1994] Mahler, R. P. S. (1994). Random-set approach to data fusion. volume 2234, pages 287–295.

[Mahler and Zajic, 2002] Mahler, R. P. S. and Zajic, T. (2002). Bulk multitarget tracking using a first-order multitarget moment filter. volume 4729, pages 175–186.

[Mori et al., 1986] Mori, S., Chong, C.-Y., Tse, E., and Wishner, R. (1986). Tracking and classifying multiple targets without a priori identification. *Automatic Control, IEEE Transactions on*, 31(5):401–409.

[P. Fiala, 2009] P. Fiala, T. Jirku, R. K. (2009). Change detection in the video sequences with small density of information. *PIERS Proceedings, Beijing, China*, pages 23–27.

[Panta et al., 2009] Panta, K., Clark, D. E., and Vo, B.-N. (2009). Data association and track management for the gaussian mixture probability hypothesis density filter. *Aerospace and Electronic Systems, IEEE Transactions on*, 45(3):1003–1016.

[Panta et al., 2007] Panta, K., Vo, B.-N., and Singh, S. (2007). Novel data association schemes for the probability hypothesis density filter. *Aerospace and Electronic Systems, IEEE Transactions on*, 43(2):556–570.

[Piccardi, 2004] Piccardi, M. (2004). Background subtraction techniques: a review. 4:3099–3104 vol.4.

[Pinto, 2008] Pinto, R. F. (2008). Biometric system for people identification based on the veins'image on the back of the palm of the hand.

[Plataniotis and Venetsanopoulos, 2000] Plataniotis, K. and Venetsanopoulos, A. (2000). *Color Image Processing and Applications*. Digital Signal Processing - Springer-Verlag. Springer.

[Pollard et al., 2009] Pollard, E., Plyer, A., Pannetier, B., Champagnat, F., and Besnerais, G. L. (2009). GM-PHD filters for multi-object tracking in uncalibrated aerial videos.

[Prokaj et al., 2011] Prokaj, J., Duchaineau, M., and Medioni, G. (2011). Inferring tracklets for multi-object tracking. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*, pages 37–44. IEEE.

[Pulford, 2005] Pulford, G. (2005). Taxonomy of multiple target tracking methods. In *Radar, Sonar and Navigation, IEE Proceedings-*, volume 152, pages 291–304. IET.

[Radke et al., 2005] Radke, R. J., Andra, S., Al-Kofahi, O., and Roysam, B. (2005). Image change detection algorithms: a systematic survey. *Image Processing, IEEE Transactions on*, 14(3):294–307.

[Rago et al., 1995] Rago, C., Willett, P., and Streit, R. (1995). A comparison of the jpdaf and pmht tracking algorithms. In *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on*, volume 5, pages 3571–3574. IEEE.

[Ramírez, 2003] Ramírez, Á. S. (2003). El filtro de kalman. *Documento de trabajo del Banco Central de Costa Rica, elaborado en la División Económica, Departamento de Investigaciones Económicas*.

[Reuter et al., 2012] Reuter, S., Meissner, D., and Dietmayer, K. (2012). Multi-object tracking at intersections using the cardinalized probability hypothesis density filter. In *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on*, pages 1172–1177. IEEE.

[Ristic et al., 2010] Ristic, B., Vo, B.-N., and Clark, D. (2010). Performance evaluation of multi-target tracking using the ospa metric. In *Information Fusion (FUSION), 2010 13th Conference on*, pages 1–7. IEEE.

[Ristic et al., 2011] Ristic, B., Vo, B.-N., Clark, D., and Vo, B.-T. (2011). A metric for performance evaluation of multi-target tracking algorithms. *Signal Processing, IEEE Transactions on*, 59(7):3452–3457.

[Rongrong, 2008] Rongrong, C. (2008). Birth density modeling in multi-target tracking using the gaussian mixture phd filter.

[Schikora et al., 2012] Schikora, M., Gning, A., Mihaylova, L., Cremers, D., and Koch, W. (2012). Box-particle phd filter for multi-target tracking. In *Information Fusion (FUSION), 2012 15th International Conference on*, pages 106–113.

[Sezgin et al., 2004] Sezgin, M. et al. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic imaging*, 13(1):146–168.

[Shitrit, 2013] Shitrit, H. B. (2013). Tracking Multiple People under Global Appearance Constraints $*$. (247022).

[Sidenbladh, 2003a] Sidenbladh, H. (2003a). Multi-target particle filtering for the probability hypothesis density. *arXiv preprint cs/0303018*.

[Sidenbladh, 2003b] Sidenbladh, H. (2003b). Multi-target particle filtering for the probability hypothesis density. *arXiv preprint cs/0303018*.

[Sidenbladh and Wirkander, 2004] Sidenbladh, H. and Wirkander, S.-L. (2004). Particle filtering for random sets. *IEEE Trans. Aerosp. Electron. Syst*.

[Sobral, 2013] Sobral, A. (2013). BGSLibrary: An opencv c++ background subtraction library. In *IX Workshop de Visão Computacional (WVC'2013)*, Rio de Janeiro, Brazil.

[Sonka et al., 2008] Sonka, M., Hlavac, V., Boyle, R., et al. (2008). *Image processing, analysis, and machine vision*, volume 3. Thomson Toronto.

[Strigel et al., 2013] Strigel, E., Meissner, D., and Dietmayer, K. (2013). Vehicle detection and tracking at intersections by fusing multiple camera views. *2013 IEEE Intelligent Vehicles Symposium (IV)*, (Iv):882–887.

[Svensson, 2010] Svensson, D. (2010). *Target tracking in complex scenarios*. Chalmers University of Technology.

[Tobias and Lanterman, 2008] Tobias, M. and Lanterman, A. (2008). Techniques for birth-particle placement in the probability hypothesis density particle filter applied to passive radar. *Radar, Sonar & Navigation, IET*, 2(5):351–365.

[Vihola, 2005] Vihola, M. (2005). Random set particle filter for bearings-only multitarget tracking. In *Defense and Security*, pages 301–312. International Society for Optics and Photonics.

[Viola and Jones, 2001] Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. 1:I–511.

[Vo and Ma, 2006] Vo, B.-N. and Ma, W.-K. (2006). The gaussian mixture probability hypothesis density filter. *Signal Processing, IEEE Transactions on*, 54(11):4091–4104.

[Vo and Singh, 2004] Vo, B.-N. and Singh, S. (2004). On the bayes filtering equations of finite set statistics. In *Control Conference, 2004. 5th Asian*, volume 2, pages 1264–1269. IEEE.

[Vo et al., 2003] Vo, B.-N., Singh, S., and Doucet, A. (2003). Sequential monte carlo implementation of the phd filter for multi-target tracking. In *Proc. Int'l Conf. on Information Fusion*, pages 792–799.

[Vo et al., 2005] Vo, B.-N., Singh, S., and Doucet, A. (2005). Sequential monte carlo methods for multitarget filtering with random finite sets. *Aerospace and Electronic Systems, IEEE Transactions on*, 41(4):1224–1245.

[Wang et al., 2008a] Wang, Y., Jing, Z., and Hu, S. (2008a). Data association for phd filter based on mht. In *Information Fusion, 2008 11th International Conference on*, pages 1–8. IEEE.

[Wang et al., 2008b] Wang, Y.-D., Wu, J.-K., Kassim, A. A., and Huang, W. (2008b). Data-driven probability hypothesis density filter for visual tracking. *Circuits and Systems for Video Technology, IEEE Transactions on*, 18(8):1085–1095.

[Wright et al., 2009] Wright, J., Yang, A., Ganesh, A., Sastry, S., and Ma, Y. (2009). Robust face recognition via sparse representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(2):210–227.

[Wu, 2013] Wu, J. (2013). PHD Filter for Multi-target Visual Tracking with Trajectory Recognition.

[Xu et al., 2004] Xu, Y. D., Allen, J., and Jin, J. S. (2004). Robust mean-shift tracking with extended fast colour thresholding.

[Yazdian-Dehkordi et al., 2012a] Yazdian-Dehkordi, M., Azimifar, Z., and Masnadi-shirazi, M. (2012a). Competitive gaussian mixture probability hypothesis density filter for multiple target tracking in the presence of ambiguity and occlusion. *Radar, Sonar Navigation, IET*, 6(4):251–262.

[Yazdian-Dehkordi et al., 2012b] Yazdian-Dehkordi, M., Rojhani, O., and Azimifar, Z. (2012b). Visual target tracking in occlusion condition: A gm-phd-based approach. In *Artificial Intelligence and Signal Processing (AISP), 2012 16th CSI International Symposium on*, pages 538–541.

[Yoon et al., 2013] Yoon, J. H., Yoon, K.-J., and Kim, D. Y. (2013). Multi-object tracking using hybrid observation in phd filter. In *ICIP*, pages 3890–3894.

[Yu and Morel, 2011] Yu, G. and Morel, J.-M. (2011). Asift: an algorithm for fully affine invariant comparison. *Image Processing On Line*, 2011.

[Zajic and Mahler, 2003] Zajic, T. and Mahler, R. P. (2003). Particle-systems implementation of the phd multitarget-tracking filter. In *AeroSense 2003*, pages 291–299. International Society for Optics and Photonics.

[Zhou et al., 2013] Zhou, X., Li, Y., and He, B. (2013). Game-theoretical occlusion handling for multi-target visual tracking. *Pattern Recognition*, 46(10):2670–2684.

[Zhu et al., 2011] Zhu, H., Han, C., and Lin, Y. (2011). Particle labeling phd filter for multi-target track-valued estimates. In *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on*, pages 1–8.

[Zivkovic, 2004] Zivkovic, Z. (2004). Improved adaptive gaussian mixture model for background subtraction. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 2, pages 28–31. IEEE.

[Zou et al., 2011] Zou, S., Liedo, P., Altamirano-Robles, L., Cruz-Enriquez, J., Morice, A., Ingram, D. K., Kaub, K., Papadopoulos, N., Carey, J. R., and Melov, S. (2011). Recording Lifetime Behavior and Movement in an Invertebrate Model. *PLOS One*, 6.