



**I
N
A
O
E**

Fusión de múltiples cámaras móviles para seguimiento de objetos

por

Lic. Jorge Eduardo Xalteno Altamirano

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

Supervisada por:

Dr. Leopoldo Altamirano Robles

Dr. Carlos Alberto Reyes

Marzo 2010, Tonanzintla, Puebla

© INAOE 2009

Derechos reservados

El autor Otorga al INAOE el permiso de reproducir y distribuir copias de su totalidad o en partes de esta tesis

RESUMEN

El siguiente trabajo se enfoca en hacer seguimiento de objetos utilizando sensores en movimiento y fusión basada en la teoría Dezert-Smarandache para fusión de información; el movimiento de los sensores es aprovechado para lograr observar objetos parcialmente ocultos.

El movimiento de las cámaras se realiza de dos formas distintas, movimiento lineal y movimiento circular. En el movimiento lineal los sensores se mueven en línea recta y se sitúan perpendicularmente uno del otro; en el movimiento circular los sensores están dispuestos alrededor del centro de la escena y giran sobre este centro, describiendo el movimiento circular.

El movimiento de los sensores es una ventaja sobre los sistemas que utilizan sensores estáticos, en este trabajo se utiliza el movimiento de los sensores para observar objetos parcial o totalmente ocultos. A diferencia de los sistemas con sensores estáticos donde es necesario utilizar algún método de predicción para resolver la oclusión, con los sensores en movimiento esta solución se logra sin hacer uso de otro método más que el mismo movimiento de los sensores.

Durante el movimiento de los sensores se realiza una acumulación de evidencia, esta acumulación permite tener una mayor certeza sobre la observación del objeto y así tomar una decisión más precisa.

Finalmente, se presenta una comparación entre los resultados de la fusión utilizando la teoría Dezert-Smarandache y la fusión Bayesiana, se realizaron experimentos para determinar que enfoque daba mejores resultados, para la evaluación de estos experimentos se utilizaron tasas de falsos y verdaderos positivos, dando como resultado que la teoría Dezert-Smarandache da mejores resultados.

ABSTRACT

In this work, a solution for the object tracking with moving cameras problem is presented; data fusion based on the Dezert-Smarandache Theory is used; the sensor movement is used to observe partially occluded objects.

The movement of the cameras is made in two different ways, named, linear movement and circular movement. In the linear movement, the cameras are set perpendicularly from each other; in the circular movement, the cameras are set around the scene and they turn around the center of the scene.

The movement of the cameras represents an advantage over the systems that use fixed cameras, because the movement of the cameras is used to observe partially or totally occluded objects, as in this work. As a difference from the systems with fixed cameras in which is necessary to use any method of prediction in order to solve the occlusion, with moving cameras this solution is reached without using any other method but the movement of the cameras itself.

During the movement of the cameras, an accumulation of evidence is done, this accumulation aloud to have a larger certainty of the object in order to take a more accurate decision.

Finally, a comparison between the results of the fusion based on Dezert-Smarandache theory and the Bayesian fusion is presented; there were made experiments in order to determine which approach was giving better results; for the evaluation of the experiments there were used rates of false and true positives, giving as a result that the Dezert-Smarandache Theory achieves better results.

AGRADECIMIENTOS

Agradezco a mis asesores Dr. Leopoldo Altamirano Robles y Dr. Carlos Alberto Reyes García por el apoyo, dedicación, observaciones y aportaciones a este trabajo de tesis.

Al Dr. Luis Carlos Altamirano Robles quien fue pieza clave para la realización de este trabajo, por sus consejos, ideas y apoyo personal.

A mi madre y mis hermanos por estar siempre junto a mí y compartir momentos en familia, a mi abuelito por sus consejos y por su apoyo.

A mi familia por su apoyo incondicional y por los consejos y jalones de orejas que me dieron.

A María Antonia Ruíz Díaz por su ayuda prestada para la realización de la tesis y por aguantar mi desesperación.

A todos mis compañeros de la maestría, por compartir dos años de buenos y malos momentos.

Finalmente al CONACyT por el apoyo económico que hizo posible el final de este trabajo.

INDICE GENERAL

	PAGINAS
CAPITULO 1	
Introducción -----	1
1.1 Problemática	4
1.2 Objetivos del trabajo de tesis	4
1.3 Contribuciones	5
1.4 Método propuesto y principales resultados	5
1.5 Contenido del documento	7
CAPITULO 2	
Estado del arte -----	9
2.1 Seguimiento de objetos -----	10
2.2 Múltiples cámaras -----	10
2.3 Niveles y arquitecturas de fusión -----	12
2.4 Conclusión.....	20
CAPITULO 3	
Marco teórico -----	23
3.1 Visión dinámica	23
3.1.1 Detección de cambios	25
3.1.1.1 Diferencia de imágenes	26
3.1.1.2 Umbralización	26
3.1.1.3 Segmentación	27

3.2 Filtro de tamaño	28
3.3 Transformaciones proyectivas.....	28
3.4 Momentos estadísticos	31
3.5 Fusión Bayesiana	32
3.6 Modelo Dezert-Smarandache	34
3.6.1 Teoría Dempster-Shafer	35
3.6.1.1 Regla de Dempster	36
3.6.2 Teoría Dezert-Smarandache	38
3.6.2.1 Conjunto Hiperpotencia	39
3.6.2.2 Funciones de Creencia Generalizadas	40
3.6.2.3 Regla de combinación clásica DS _m	41
3.6.2.4 Regla de combinación Híbrida DS _m	41
3.7 Fusión Multicámara para Seguimiento de Objetos	
Basada en Teoría Evidencial	43

CAPITULO 4

Propuesta de solución	45
4.1 Esquema del método propuesto.....	45
4.2 Arreglo experimental	47
4.3 Experimentos realizados	47
4.3.1 Movimiento de las cámaras	47
4.3.1.1 Movimiento Horizontal	47
4.3.1.2 Movimiento Circular	48
4.3.1.3 Detección de movimiento con cámaras	

en movimiento horizontal	49
4.3.1.4 Detección de movimiento con cámaras	
en movimiento circular	50
4.4 Eliminación de ruido	51
4.5 Correspondencia de datos	52
4.6 Creación del marco de discernimiento	53
4.7 Creación de masas	55
4.8 Creación del conjunto hiperpotencia	56
4.9 Fusión de información	58
CAPITULO 5	
Resultados experimentales	61
5.1 Pruebas con las cámaras en movimiento horizontal	64
5.2 Pruebas con las cámaras en movimiento circular	67
5.3 Pruebas con las cámaras en movimiento para evitar	
la oclusión del objeto	69
5.4 Acumulación de evidencia a través de varios cuadros	71
5.5 Comparación entre fusión evidencial y fusión bayesiana	75
5.6 Métricas de evaluación	79
CAPITULO 6	
Conclusiones y trabajo futuro	83
6.1 Aportaciones	84
6.2 Trabajo futuro	85

ANEXO A

Publicaciones obtenidas 87

Referencias ----- 89

INDICE DE FIGURAS

	PAGINAS
Figura 2.1 Arquitecturas generales de los modelos de fusión	13
Figura 3.1 Homografía	29
Figura 4.1 Movimiento horizontal de las cámaras	48
Figura 4.2 Movimiento circular de las cámaras	49
Figura 4.3 Plantilla.....	50
Figura 4.4 Imagen de la cámara 2	50
Figura 4.5 Imagen resultante de la resta de las figuras 4.3 y 4.4	50
Figura 4.6 Detección por color.....	51
Figura 4.7 Rectángulo y eje vertical del objeto.....	53
Figura 4.8 Creación del marco de discernimiento.	
Cada celda es un elemento del maro.....	54
Figura 5.1 Eje vertical del objeto, secuencia con movimiento horizontal	63
Figura 5.2 Eje vertical del objeto, secuencia con movimiento circular.	63
Figura 5.3 Ejemplo de la acumulación de evidencias.....	72
Figura 5.4 Comparación entre las evidencias mostradas por las cámaras y las decisiones tomadas como las técnicas de fusión utilizadas en este trabajo	77
Figura 5.5 Incremento de la evidencia con el movimiento de las cámaras	79

INDICE DE TABLAS

	PAGINAS
Tabla 4.1 Áreas utilizadas para el filtro de tamaño.....	52
Tabla 5.1 Secuencia 1y resultado de la fusión sin acumulación en el tiempo	65
Tabla 5.2 Secuencia 2 y resultado de la fusión sin acumulación en el tiempo....	67
Tabla 5.3 Secuencia 3 y resultado de la fusión sin acumulación en el tiempo....	70
Tabla 5.4 Resultados de la fusión de la primera terna de la tercera secuencia de imágenes.....	73
Tabla 5.5 Tiempos de ejecución.....	74
Tabla 5.6 Fragmento de la secuencia 1 y resultado de la fusión Bayesiana.....	75
Tabla 5.7 Comparación entre los distintos métodos utilizados en este trabajo....	80

Capítulo 1

Resumen

En este capítulo se da una introducción al trabajo que se realizó en esta tesis, se mencionan algunas áreas de la computación que se tocan con en este trabajo tales como visión artificial y sistemas de vigilancia, se mencionan también algunos problemas que se tienen al utilizar sensores estáticos y como pueden solucionarse con el movimiento de estos sensores. Se menciona la problemática a resolver con este trabajo de tesis, así como los objetivos perseguidos y las contribuciones del trabajo. Finalmente se muestra la distribución general del documento.

INTRODUCCIÓN

La visión artificial ha sido empleada en múltiples campos de la ciencia y la tecnología; sus aplicaciones son variadas, e incluyen: detección temprana de cáncer, entre otras aplicaciones médicas; las aplicaciones militares también son múltiples; el seguimiento de objetos y la inspección de objetos manufacturados por el hombre también se incluyen. La vigilancia de ambientes es un área de gran aplicación que ha recibido una especial atención dentro del área de la visión artificial.

La vigilancia de ambientes es el proceso de detectar objetos que necesitan ser observado, normalmente esto se encuentran en movimiento. En este caso la vigilancia se transforma en el proceso de seguir a los objetos, definir sus trayectorias (a menudo

personas o automóviles), y establecer patrones de comportamiento o tendencias inferidas de sus movimientos.

Usualmente, la vigilancia de ambientes es realizada mediante un sensor situado estratégicamente, de tal manera que la escena de interés sea cubierta apropiadamente por el sensor. Sin embargo, en muchas escenas es imposible vigilar el entorno completo con un solo sensor, por lo que se hace necesario el empleo de múltiples sensores.

El empleo de múltiples sensores involucra resolver problemáticas diferentes a los que surgen al tener sistemas con tan solo un sensor. Así por ejemplo, tenemos la problemática de cómo fusionar la información proveniente de los sensores, realizar la correspondencia entre los puntos que componen la escena para obtener información de profundidad o la determinación de la unicidad de un mismo objeto en las múltiples imágenes obtenidas, por nombrar algunas.

A pesar de la problemática involucrada al utilizar varias cámaras para vigilar una escena, existen puntos positivos. Una de sus ventajas es tener mayor campo de visión cubierto por las cámaras, así como una mejor fuente de información para analizar la situación que se presenta en la escena. Es por ello que muchos sistemas biológicos emplean más de un sensor para percibir el mundo en el que viven, y de la misma forma, se intenta dotar a la computadora de algoritmos capaces de imitar la forma en la que el humano percibe e interactúa con el mundo.

Muchos sistemas de vigilancia actuales utilizan personas para revisar constantemente las escenas tomadas con cámaras; sin embargo este trabajo es tedioso, repetitivo y requiere de una gran atención durante mucho tiempo, lo cual origina errores en la interpretación. Por ello es deseable crear sistemas que realicen este trabajo de forma automática y eficazmente.

Los sistemas de vigilancia trabajan bajo el principio de la detección de cambios en una escena, es decir, encontrar que objetos cambian ó se mueven dentro de la escena que están sensando; esto es posible gracias a técnicas como por ejemplo la estimación de fondo o el flujo óptico; sin embargo en ocasiones estos algoritmos no bastan para hacer una buena

detección de los objetos en movimiento, por tanto, se deben emplear otras técnicas que ayuden a resolver este problema.

Los sensores, como se comentó anteriormente, tienen un campo de visión limitado, y aunque es posible tener un mayor campo de visión utilizando lentes que permiten incrementar el tamaño del campo de visión, esto induce aberraciones en la imagen, como en el caso de las cámaras panorámicas. Para tener un mayor campo de visión, otra solución es hacer que los sensores se muevan, de esta forma con una lente convencional se logra cubrir un mayor campo de visión.

Por otro lado, en la mayoría de las secuencias y escenas que obtienen los sistemas de visión se presenta el problema, cuando los objetos están en movimiento, de la posible oclusión parcial de los objetos monitoreados, ocasionando que durante algún tiempo no se sabe dónde está el objeto de interés. Existen formas de solucionar el problema de la oclusión como los enfoques Markovianos [32], que predicen la posición futura del objeto, dada la última posición conocida de la trayectoria. Si el objeto se queda oculto durante mucho tiempo estos enfoques no funcionan apropiadamente. Una solución a esta problemática de la oclusión es mover las cámaras para que el objeto que se está ocultando entre en el campo de visión de la cámara. Este es precisamente el enfoque que en este trabajo se utiliza: dos o más sensores en movimiento para tener un mayor campo de visión y solucionar el problema de la oclusión parcial.

El permitir que múltiples cámaras sensen una misma escena introduce también nuevas problemáticas para un sistema de vigilancia; éstas incluyen: determinar la correspondencia entre las imágenes obtenidas por los sensores, sobre todo si los sensores se encuentran en movimiento y lograr que las cámaras cooperen entre ellas para que la información global sea más útil que cualquiera obtenida individualmente. Para la segunda problemática se propone la fusión de información con la cual se logra complementar la información dada por uno de los sensores con la información proporcionada por los demás.

1.1 PROBLEMÁTICA

Después de la revisión del estado del arte, el cual se presenta en el siguiente capítulo, se observó que la fusión de sensores en general no aborda el problema del ocultamiento parcial de objetos de interés en la escena, en especial la evaluación apropiada de la porción del objeto que está a la vista de la cámara, lo cual limita seriamente las capacidades de detección de los sistemas. Con base en esto, se observó que la posible solución era permitir el movimiento relativo de las cámaras para mejorar la evidencia de los objetos en la escena, tal y como lo hacen los sistemas biológicos en el mundo real, cuando se mueven de lugar a otro para obtener una mejor panorámica de la escena de interés, así mismo proponer un método para evaluar la porción visible de los objetos que se están siguiendo e incorporarla al esquema de fusión utilizado. Se propone también la acumulación de evidencias a lo largo de la secuencia de imágenes a procesar.

1.2 OBJETIVOS DEL TRABAJO DE TESIS

El objetivo general de este trabajo de tesis consiste en proponer un método para hacer seguimiento confiable de objetos bajo las condiciones de oclusión parcial o total usando sensores móviles.

Como objetivos particulares se tienen los siguientes:

- Implementar el seguimiento de objetos para sensores en movimiento.
- Implementar un enfoque de fusión basado en DSMT que permita acumular la evidencia de las cámaras en movimiento para tener mayor certeza sobre el seguimiento.
- Evaluar la porción visible de los objetos que se están siguiendo e incorporar el resultado de esta evaluación al esquema de fusión.
- Utilizar el método de evidencias acumuladas para mejorar la certeza en el seguimiento de objetos.
- Comparar el método de fusión propuesto contra fusión Bayesiana.

- Comparar el método de fusión con evidencias simples contra el método de fusión con evidencias acumuladas.

1.3 CONTRIBUCIONES

La contribución de este trabajo de tesis es mostrar que utilizando cámaras en movimiento y fusionando la información de estas cámaras es posible tener una mejor evidencia y entendimiento de la escena que se está vigilando; también mostrar que el movimiento de las cámaras permite observar objetos ocluidos, dando así una solución a éste problema. La lista siguiente muestra las contribuciones:

- La fusión de la información acumulada da mejores resultados que la fusión de información en un solo cuadro.
- El movimiento de los sensores permite observar objetos ocultos y de esta forma se tiene mayor evidencia del objeto

1.4 MÉTODO PROPUESTO Y PRINCIPALES RESULTADOS

Para alcanzar el objetivo propuesto se presenta el siguiente método, con el cual se pretende alcanzar el objetivo del seguimiento confiable de objetos.

1. Adquisición de imágenes
 - Por medio de cámaras en movimiento dispuestas alrededor de la escena se toman secuencias de imágenes, el movimiento de las cámaras es en línea recta y en forma circular.
2. Identificación del objeto a seguir

- Utilizando plantillas y detección por color se segmenta el objeto que se quiere seguir.
3. Eliminación de ruido
 - La eliminación de ruido se realiza por medio de un filtro de tamaño ya que el ruido resultante del paso anterior es fácil de remover por medio de este método.
 4. Cálculo de la homografía entre cámaras
 - Para el cálculo de la homografía entre las cámaras se utilizaron una matriz de homografía y una matriz de transformaciones (rotación y traslación)
 5. Evaluación de la porción visible del objeto
 - El eje horizontal del objeto se puede utilizar para saber si la cámara está observando una mayor parte del objeto y utilizar esta información para mejorar la evidencia.
 6. Calculo de su nivel de creencia
 - Utilizando el eje vertical del objeto se crea un marco de discernimiento, a este marco se le asignan evidencias que serán utilizadas en la fusión.
 7. Fusión de creencias
 - Finalmente se fusionan las evidencias utilizando la teoría Dezert-Smarandache.

Los principales resultados obtenidos con este trabajo son:

- Se muestra que la utilización de DSMT tiene mejores resultados que la fusión Bayesiana.
- La acumulación de evidencias mejora la certeza de las observaciones.
- El movimiento de las cámaras permite tener mayor certeza de las observaciones de las cámaras.

1.5 CONTENIDO DEL DOCUMENTO

La distribución del resto del documento es la siguiente:

- **Capítulo 2.** Estado del arte, se da una visión de los trabajos hasta el momento sobre fusión de información enfocada al seguimiento de objetos y la estructura general de estos trabajos de acuerdo al modelo JDL propuesto por el “American Joint Directors Laboratories Data Fusion Subpanel”.
- **Capítulo 3.** Marco teórico, en esta parte del documento se describen las técnicas utilizadas en el trabajo de tesis propuesto como: detección de movimiento, filtro de tamaño, fusión Dezert-Smarandache, entre otros.
- **Capítulo 4.** Propuesta de solución, se describe la forma en la que fue solucionado el problema propuesto.
- **Capítulo 5.** Pruebas experimentales, en esta parte del documento se explican los resultados obtenidos al implementar la solución propuesta.
- **Capítulo 6.** Trabajo futuro, en este capítulo se indican posibles caminos a seguir para continuar con el trabajo actual.

Capítulo 2

Resumen

En este capítulo se presenta el estado del arte acerca de sistemas de vigilancia, seguimiento de objetos, uso de múltiples cámaras, y niveles de fusión, con el fin de tener un marco de referencia sobre estos temas.

ESTADO DEL ARTE

La visión artificial es un área de la computación y especialmente de la inteligencia artificial, cuyas técnicas permiten la obtención de información por métodos ópticos, tanto de manera automática como asistida, con ella, es posible hacer que la computadora reconozca objetos y pueda posteriormente manipularlos o dar información sobre el objeto que está percibiendo. Por medio de la visión artificial, es posible que la computadora reconozca automóviles, personas, bicicletas, botellas, rostros, etcétera.

Para vigilancia se han creado sistemas especializados, la función de estos sistemas es observar una escena con el fin de encontrar y seguir objetos o personas que pasen por el área vigilada, encontrar actividades anómalas que realicen estos objetos o para advertir sobre elementos fuera de contexto que ocurran en la escena.

2.1 SEGUIMIENTO DE OBJETOS

El seguimiento de objetos es la estimación de la posición y estado de un objeto en movimiento [33]. Existen variadas técnicas para hacer este proceso. Se puede hacer el seguimiento utilizando filtros sobre las imágenes o haciendo correlación del objeto que se está buscando sobre la imagen. También es posible distinguir los objetos en movimiento por medio de restas de imágenes consecutivas o por medio de una diferencia entre la escena sin objetos (plantilla) y la imagen actual vista por la cámara. Este es un método sencillo y rápido de realizar, además de ser exacto si se tiene conocimiento de la escena *a priori*. [24]

Es posible también crear un modelo de la escena que mejore el rendimiento de la detección de objetos en movimiento por medio de un enfoque de fondo adaptable [13]. Sin embargo, no todos los métodos mencionados anteriormente funcionan cuando se tienen cámaras en movimiento, para este caso, es necesario utilizar otros métodos, como flujo óptico [24] o utilizar una variante de los métodos mencionados.

2.2 MÚLTIPLES CÁMARAS

Cuando se tiene un sistema observando una escena, éste puede tener problemas al momento de seguir objetos. Puede no distinguir el objeto si éste se encuentra en una posición extraña, puede ser que no le sea posible detectar al objeto si este pasa muy rápido por su campo de visión o puede simplemente fallar la cámara.

Una solución a esta problemática la presenta la utilización de dos o más cámaras del mismo tipo, colocándolas estratégicamente en el espacio que se quiere monitorear, lo que permite tener un mayor campo de visión al distribuir las cámaras a lo largo del espacio que se quiere vigilar. Es posible también utilizar la información aportada por cada cámara si se tienen cámaras que apuntan hacia la misma escena pero de distinto punto de vista, en consecuencia se puede distinguir mejor que objeto es el que se está siguiendo o determinar en mejor manera la trayectoria que éste está formando.

En [11] se ha utilizado este método, para reconocer cuando una persona está sobre el riel de un ferrocarril. Se dispusieron dos cámaras viendo el cruce peatonal sobre las vías del tren, las cámaras están una opuesta a la otra, esto les da a cada una un ángulo distinto sobre la escena, las cámaras se ayudan proporcionando información sobre la escena a una computadora que controla a ambas cámaras.

En [12] se presenta un enfoque para el seguimiento de múltiples objetos, este enfoque considera la detección y estimación de la trayectoria del objeto en el espacio y tiempo como un problema de optimización. Este enfoque se formula en un marco de selección basado en una Longitud Mínima de Descripción, la cual permite que el sistema se recupere de pérdidas temporales del seguimiento. Este trabajo efectúa el reconocimiento de objetos en múltiples vistas y múltiples categorías para la detección de automóviles y peatones en las imágenes de entrada. Los objetos en 2D son verificados para su consistencia con la geometría de la escena y son convertidos en observaciones de 3D, las cuales son acumuladas en un marco coordinado del mundo. Posteriormente, el seguimiento es llevado a cabo al realizar un modelo de selección después de cada cuadro. Las pruebas de este trabajo fueron realizadas sobre secuencias de imágenes reales, tomadas en un vehículo en movimiento y en un área transitada de la ciudad.

En [13] se propone un modelo que no requiere los parámetros del movimiento de la cámara para hacer la detección, el fondo es aprendido al vuelo y la solución del problema es integrada con el seguimiento de objetos. El movimiento de la cámara es estimado como el clúster dominante en una mezcla de Gaussianas aprendidas sobre los puntos de correlación entre cuadros consecutivos. Las regiones objetivo son detectadas como cambios sobre el fondo ya aprendido. En las escenas con varios objetos en movimiento, se escoge solamente uno y es el que se sigue y se mantiene en el centro de la imagen mientras se mueva y la cámara pueda mantenerlo en el campo de visión. El movimiento de la cámara es controlado utilizando un error dinámico.

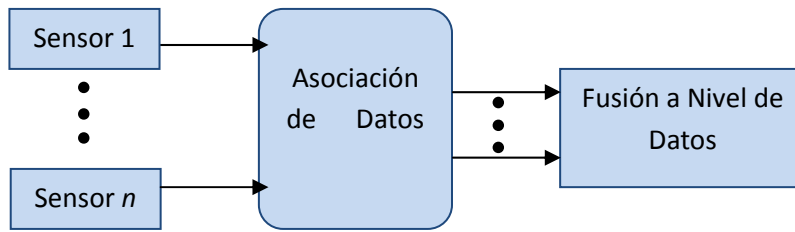
En [28] se muestra un método para realizar fusión de imágenes producidas por múltiples sensores, mediante un enfoque probabilístico. Las imágenes obtenidas de los sensores contienen ruido, lo que dificulta la identificación de objetos de interés en las

imágenes. Las imágenes obtenidas por los sensores son utilizadas para crear un modelo, posteriormente se utiliza un marco de trabajo Bayesiano, el cual provee probabilidades o estimaciones máximas de la escena observada por los sensores. En este trabajo se muestra que al utilizar una fusión de información proveniente de múltiples sensores es posible mejorar la calidad de la imagen y con ello obtener una mejor apreciación o estimación de la escena real.

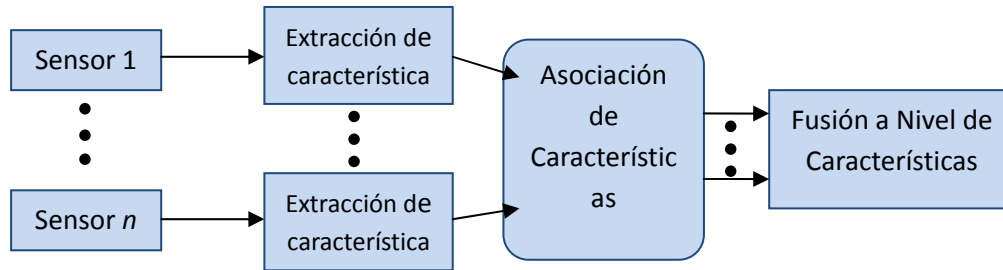
Existen trabajos que hacen vigilancia automática de escenas tomando en cuenta la información de varios sensores y fusionando dicha información tal es el caso de [29], donde toman imágenes de varias cámaras digitales (cámaras de video normales), y segmentan los objetos en movimiento. Posteriormente toman estos objetos y les extraen los momentos de Hu como sus características. Con ellas hacen la fusión de información por medio de un enfoque basado en el filtro de Kalman, el procedimiento que siguen es el siguiente: 1) recolectar la información disponible de cada sensor; 2) agrupar y asignar la información a los objetos en movimiento segmentados en el cuadro anterior; 3) actualizar el estado de cada objeto en movimiento con la información obtenida del filtro de Kalman. Este trabajo utiliza múltiples sensores y fusión de información, sin embargo los sensores son fijos.

2.3 NIVELES Y ARQUITECTURAS DE FUSIÓN

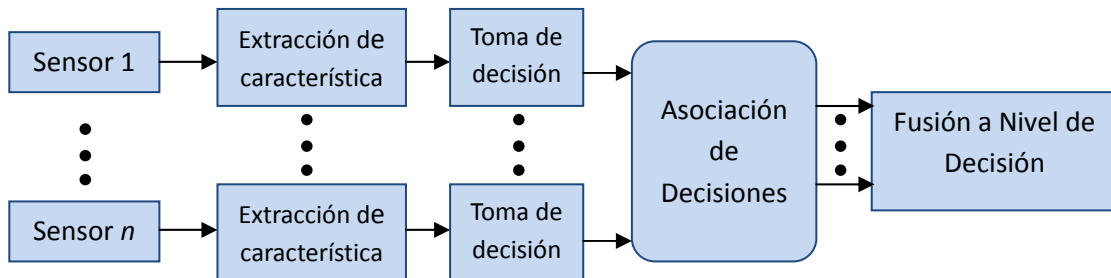
Al hacer fusión de información se pueden tomar los datos tal y como se nos presentan y fusionarlos o procesar estos datos para sacar más información de ellos y finalmente fusionarlos, cuando se toman los datos tal y como el sensor los arroja y se fusionan, se tiene que este es una arquitectura de fusión a nivel de datos, cuando a estos datos se les obtienen características y se fusionan, se tiene una arquitectura de fusión a nivel de características y finalmente, si con estas características se obtienen decisiones sobre lo que se está observando se tiene una arquitectura de fusión a nivel de decisión. En la figura 2.1 se pueden ver esquemas de las arquitecturas mencionadas.



a) Arquitectura de Fusión a nivel de datos



a) Arquitectura de Fusión a nivel de características



a) Arquitectura de Fusión a nivel de decisión

Figura 2.1 Arquitecturas generales de los niveles de fusión

En [14] se da un panorama general de Fusión de múltiples sensores aplicados al seguimiento de objetivos; de acuerdo con este trabajo, existen 4 niveles en los cuales se puede dividir la fusión de información, esto es de acuerdo al modelo JDL (Joint Directors Laboratories) propuesto por el “American Joint Directors Laboratories Data Fusion Subpanel”, estos niveles son:

1. **Nivel 1, Refinamiento del objeto:** compuesto de registro de datos, asociación de datos, estimación de los atributos de la posición e identificación.
2. **Nivel 2, Evaluación de la situación:** fusiona la cinemática (movimiento del objeto) y características temporales de los datos para crear una descripción de la situación en términos de indicaciones de cuidado, planes de acción e inferencias acerca de la distribución de las fuerzas y la información. Un algoritmo de Evaluación de la Situación decidirá si un objeto actúa en forma hostil y en qué forma reaccionar.
3. **Nivel 3, Evaluación de peligro o acción:** este nivel describe qué acción está llevando a cabo y si ésta es una acción indebida hecha por el objeto que se está siguiendo.
4. **Nivel 4, Evaluación de proceso:** es una evaluación de los procesos anteriores, con el fin de verificar que la adquisición de datos y la fusión se están llevando a cabo de forma correcta.

El primer nivel de esta jerarquía se divide a su vez en los siguientes niveles:

- I. **Registro de datos:** las funciones de registro de datos alinean los datos en un marco de referencia común. Normalmente esto se hace cambiando el sistema de coordenadas Cartesianas a latitud, longitud y altura sobre el mar, por ejemplo.
- II. **Asociación de datos:** El paso de asociación compara e intenta coleccionar medidas del objeto con respecto al mundo real y llevarlas a un solo seguimiento, la dificultad es al distinguir de que objeto se origina cada medida.

Algunos algoritmos utilizados para este paso son:

- Vecinos más cercanos: es la forma más simple de asociar datos, en este algoritmo, la medida más cercana para establecer el seguimiento es el objeto más cercano.
- Joint Probabilistics Data Association (JPDA): Bar-Shalom et al. crearon dos filtros relacionados. El primero es un filtro de asociación de

datos probabilístico. Este trabaja en casos de un solo objeto. Todas las medidas en la región alrededor del objeto se asumen como posibles para actualizar el seguimiento. Una probabilidad posterior de asociación es calculada para cada una de las medidas validadas. Estas probabilidades son utilizadas como pesos para calcular un promedio pesado, el cual es añadido al seguimiento.

El segundo filtro extiende el seguimiento a múltiples objetivos.

- **Relajación Lagrangiana:** Puede mostrarse que la asociación de datos de múltiples sensores es un problema NP-Duro. En [30] se da evidencia de que, un algoritmo NP-Duro no puede ser resuelto computacionalmente en una forma eficiente, sin embargo existen algoritmos de aproximación que proporcionan soluciones óptimas; la Relajación Lagrangiana es una de tales soluciones. Pattipatti et al. [15] aplicaron esta solución, la cual garantiza la solución en un tiempo polinomial y en un espacio $O(n)^3$.
- **Redes Neuronales Artificiales (ANN):** El seguimiento de varios objetivos puede verse como una asociación o agrupamiento de objetivos para hacer corresponder el mismo objetivo. Winter y Favier [16] proponen una forma de utilizar ANN para resolver este problema, la solución se basa en una red de Hopfield.
- **Lógica Difusa:** La desventaja de los filtros PDA y JPDA es que mientras el número de objetivos crece, el tiempo de cómputo crece exponencialmente, la lógica difusa hace uso del “sentido común” en lugar de encontrar soluciones matemáticas. Hong et al. [17] crearon un número de reglas difusas para la fusión de datos; convirtieron los datos obtenidos por los sensores en conjuntos difusos con los valores {NL, NS, ZO, PS, PL} ({Negativo Grande, Negativo Pequeño, Zero, Positivo Pequeño y Positivo Grande} por sus siglas en inglés). Los resultados mostraron que al utilizar lógica difusa, disminuye el tiempo de cómputo y si se utilizan más conjuntos difusos, se incrementa la exactitud.

III. **Estimación de posición y atributos:** es el proceso en el cual se toman las medidas asociadas y se calcula el estado del objetivo. Para este paso se utiliza mayormente el filtro de Kalman para predecir la posición siguiente del objetivo. También es utilizada la técnica de filtro de partículas.

IV. **Identificación:** El paso de identificación clasifica el objeto del cual se obtuvieron las medidas, una vez identificado localmente, debe ser fusionado con las estimaciones remotas para obtener una solución global.

Según esta clasificación, el trabajo propuesto en esta tesis se coloca en este punto, existen varias formas de hacer esto:

- Inferencia Bayesiana: esta es una técnica que utiliza probabilidades para representar grados de pertenencia. El teorema de Bayes es utilizado para hacer estimaciones subjetivas de creencia, Hall en [18] propuso una lista de problemas con la inferencia Bayesiana que incluye:
 - Problemas al definir probabilidades *a priori*
 - Complejidad cuando hay muchas hipótesis potenciales y muchas condiciones dependientes de eventos
 - Las hipótesis deben ser mutuamente exclusivas
 - La limitante para describir incertidumbre en las decisiones.
- Regla de combinación Dempster-Shafer: Dempster y posteriormente Shafer generalizaron el modelo de creencia Bayesiana para permitir la representación de incertidumbre. Esto es requerido para modelar la situación en la cual un algoritmo de clasificación no puede clasificar un objetivo o no puede listar exhaustivamente todas las clases a las cuales puede pertenecer. Sin embargo, el precio a pagar por esta mejora es el aumento de la complejidad en tiempo y espacio, la cual es de 2^{2^n} .

- Redes neuronales artificiales: es un sistema de elementos de procesamiento paralelo-distribuido, conectados topográficamente en un grafo, es decir, los elementos del grafo se conectan de forma definida. Los datos no son guardados separadamente, sino que están ligados intrínsecamente. Uno de los problemas más difíciles con las ANN es escoger la topología de red más adecuada al problema.
- Sistemas expertos: Aunque las redes neuronales son buenas para resolver problemas generales, continuamente son muy lentas para el procesamiento (aprendizaje) y uno de los problemas difíciles con estas es encontrar el volumen correcto de datos, tanto si los datos son demasiados como si son muy pocos el sistema tomará decisiones malas. Kilter [19] sugiere que los sistemas expertos deberían ser usados para hacer la identificación en cada plataforma y un promedio pesado es tomado de los resultados.
- Enfoques basados en votaciones y sumas: La fusión por votación y suma son dos enfoques simples y populares. En ambos, un banco de algoritmos de clasificación es usado; estos pueden estar localizados en el sensor o en el algoritmo de fusión. En la fusión con sumas, la confianza de cada clasificador en cada hipótesis es sumada y la hipótesis con el mayor valor es utilizada. En fusión por votación, la hipótesis que ha sido más votada es la que se escoge; la votación consiste en que uno de los sensores tenga varias opciones por las cuales votar, dependiendo de la observación del sensor, éste decide cuál de sus opciones es la más acertada y a esa opción le da su voto.

En [20] se propone un nuevo algoritmo de fusión de imágenes basado en la transformada (NSCT, nonsampled contourlet transform), con el propósito de resolver los problemas presentados en la fusión de imágenes multifocales. Basado en el vector direccional normal, un esquema de selección combinado con el esquema de promedio es presentado para los coeficientes de la subbanda pasabajas. Basado en un contraste direccional y el vector de desviación estándar un principio de selección es utilizado para los coeficientes de la subbanda direccional pasabandas. Los resultados experimentales muestran que el algoritmo propuesto no solamente puede extraer la información más

importante de las imágenes fuente, sino que evita efectivamente la introducción de información artificial. Esto mejora significativamente la transformación discreta basada en wavelets y el método de fusión basado en la transformada wavelet discreta, en términos de la calidad visual y la evaluación objetiva, especialmente cuando las fuentes de las imágenes no están registradas perfectamente.

En el desarrollo de vehículos inteligentes, el seguimiento del camino en vehículos sin conductor está basado en el manejo autónomo y navegación automática. Es muy importante encontrar la posición exacta de un vehículo para el seguimiento del camino, y es posible encontrar la información de la posición mediante GPS. Sin embargo la información del GPS no es la posición actual sino, la posición anterior ya que el vehículo está en movimiento y el GPS tiene un retraso en el tiempo. Se estima la distancia utilizando un sensor de dirección y un sensor de velocidad para compensar el error del GPS, utilizando una fusión de datos con la información de estos dos sensores. En el control de dirección, reglas fuzzy óptimas para el seguimiento del camino pueden ser encontradas a través de la simulación en Simulink. Experimentos reales de conducción muestran que las reglas fuzzy son buenas para el control de la dirección y el error en la posición del GPS se compensa bien mediante el método de estimación propuesto. [21]

En [22] se presenta un sistema de detección en un cruce de peatones, el cual utiliza sensores de video pre-existentes orientados con el tráfico, estos sensores proveen medidas espaciales en áreas a lo largo del cruce de peatones. La detección del cruce de peatones está basada en el reconocimiento de patrones de ocupación presentados por los peatones cuando se mueven a través del cruce peatonal. Para mejorar la habilidad para detectar peatones de los sensores no dedicados, se presenta un proceso de fusión basado en evidencia que explota la información redundante proveniente de uno o dos sensores: la fusión intra-sensor utiliza características espaciotemporales de las medidas y la fusión inter-sensor, utiliza la redundancia entre los dos sensores. Como parte del proyecto TRACKSS en sensores cooperativos avanzados para las aplicaciones hacia el tráfico, se recolectaron datos reales en una intersección equipada con dos cámaras. Los resultados obtenidos muestran que los datos del proceso de fusión mejoran la calidad de los patrones de ocupación obtenidos y

llevan a una tasa de detección alta de los cruces de peatones con sensores multipropósito en condiciones operacionales, especialmente cuando un sensor secundario está disponible.

Los avances recientes en el procesamiento de señales y tecnologías de información tienen como resultado la utilización de múltiples sensores para el monitoreo efectivo de las condiciones de las herramientas, la cual es una información de crucial importancia para el controlador del proceso. Interesantemente, la abundancia de datos recolectados de múltiples sensores permite emplear varias técnicas tales como la extracción de características y métodos de selección y clasificación para generar tal información. Mientras el uso de múltiples sensores ha mejorado la exactitud en la clasificación de las condiciones de las herramientas, el diseño del sistema de monitoreo de las condiciones de las herramientas (TCM por sus siglas en inglés) para reducir su complejidad e incrementar la robustez ha sido raramente estudiado.

En suma, dos métodos de selección de características y tres clasificadores con una técnica de máquinas de ensamblaje son considerados como componentes del diseño. Diferentes métodos de fusión son evaluados: (1) fusión a nivel de decisión y (2) fusión a nivel de características, estos métodos de fusión pertenecen al nivel IV en la jerarquía del JDL. Los resultados experimentales muestran que el diseño de TCM basado en la fusión a nivel de características puede mejorar significativamente la exactitud de la clasificación de las condiciones de las herramientas. También muestra que una mayor exactitud puede lograrse al utilizar sensores de fuerza, vibración y acústicos junto con un método de selección basado en correlación. [23]

La fusión de información es utilizada también para robótica como se muestra en [34], donde presentan un enfoque de fusión basado en la teoría Dempster-Shafer, en este trabajo se presenta un sistema de manejo de incertidumbre utilizado para la ejecución de la arquitectura SFX (Sensor Fusión Effects), esta arquitectura es un sistema genérico de fusión de información para robots móviles autónomos. La ejecución utiliza creencias generadas por un perceptor para proseguir con la tarea, terminar la tarea, o investigar la situación esperando obtener creencia suficiente. La teoría Dempster-Shafer se utiliza como fundamento para el manejo de incertidumbre, la implementación de la arquitectura

incorpora evidencia de las observaciones de los sensores y el conocimiento del dominio en tres niveles de abstracción. En el Carnegie-Mellon University Mobile Robot Laboratory (MRL) se creó una representación de incertidumbre y conocimiento incompleto de sensores llamada rejilla de certidumbre (certainty grids) [35] la cual fue utilizada en varios programas de control de robots móviles. Las rejillas de certidumbre probaron ser una poderosa y eficiente solución para la fusión de sensores y en conjunto con un fundamento estadístico Bayesiano mejora la fusión.

En [36] se propone un enfoque novedoso para la detección y seguimiento de personas en secuencias utilizando filtros de partículas. La detección y el seguimiento son realizados mediante un mapa de ocupación con una representación de color novedosa. Con estos mapas se propone un algoritmo de filtro de partículas múltiple para la detección y seguimiento de personas. Los experimentos muestran que este enfoque es capaz de detectar y mantener el seguimiento de cada persona con un error bajo y tratar con ocultamiento parcial o total. En [37] utilizando una extensión de un filtro de partículas Bayesiano y la teoría Dempster-Shafer, se resuelve el problema de multiseguimiento (multi-tracking) al combinar evidencias de múltiples sensores poco fidedignos y heterogéneos. El modelado de incertidumbre y de la ausencia de conocimiento en este enfoque es especialmente atractivo ya que no se requiere de condiciones *a priori* que podrían ser difíciles de obtener en problemas complejos. Los sensores proveen un grado de evidencia conforme a su confiabilidad, la confiabilidad es calculada con base en la porción visible de los objetos y sus oclusiones, las evidencias colectadas del conjunto de cámaras son fusionadas considerando la confiabilidad para calcular la mejor hipótesis.

2.4 CONCLUSIÓN

De acuerdo a la revisión del estado del arte, se tiene que existen trabajos que hacen seguimiento de objetos con un sensor, y aquellos que hacen el seguimiento con dos o más sensores, dentro de estos sistemas que utilizan dos cámaras existen aquellos que utilizan algún método para mejorar la observación de la escena (fusión de información) está fusión puede ser a varios niveles: a nivel de datos, a nivel de características y a nivel de decisión.

Los enfoques utilizados para la fusión de información son distintos, tal es el caso de la fusión Bayesiana, la lógica difusa o la teoría Dempster-Shafer, estos enfoques son utilizados en cualquier nivel de fusión. En muchos casos el seguimiento de los objetos no es una tarea fácil, ya que esto pueden ocultarse parcialmente a los sensores e impidiendo su observación. Existen trabajos que tratan con el problema de oclusión, esto lo hacen utilizando filtros de Kalman u otros enfoque de predicción, también se utilizan enfoques que utilizan filtros de partículas modificados para lograr este objetivo.

En este caso, tenemos que el trabajo actual se sitúa dentro de los enfoques que utilizan múltiples sensores móviles y fusión de información para el seguimiento de objetos móviles, también se trata con el problema de oclusión parcial de los objetos utilizando el movimiento de las cámaras.

Capítulo 3

Resumen

En este capítulo se describen las teorías y técnicas utilizadas en este trabajo: las técnicas para detectar objetos en movimiento y para calcular momentos estadísticos así como la teoría Dezert-Smarandache y la Fusión Bayesiana utilizadas para realizar fusión de información.

MARCO TEÓRICO

3.1 VISIÓN DINÁMICA

La mayoría de los sistemas biológicos han evolucionado para tratar con medios ambientes en movimiento, de la misma forma, la visión por computadora ha evolucionado en este sentido. Para la visión por computadora el tratar con cambios de iluminación y cambios de las direcciones de observación es indispensable para poder interactuar con el mundo real. A pesar de que inicialmente la visión computacional trabajaba con imágenes estáticas, ha evolucionado para poder tratar con imágenes dinámicas, este enfoque es utilizado en múltiples tareas.

La entrada para un sistema de visión dinámica es un conjunto de imágenes secuenciales tomadas de un mundo en constante cambio, las cámaras que toman estas secuencias pueden estar en movimiento también. Cada cuadro representa una imagen en un instante particular de tiempo. Los cambios en una escena pueden deberse a cambios en la estructura, tamaño o forma de un objeto. Es asumido que los cambios en una escena son

debidos a la cámara y/o a los movimientos del objeto y que los objetos son rígidos o casi-rígidos. El sistema debe detectar los cambios, caracterizar el movimiento utilizando abstracción de alto nivel, recuperar la estructura del objeto y reconocer los objetos en movimiento.

Una escena usualmente contiene varios objetos. Una imagen de una escena en un tiempo dado representa una proyección de la escena, la cual, depende de la posición de la cámara. Existen cuatro posibles combinaciones dada la naturaleza dinámica de la cámara y del objeto observable: [24]

1. Cámara fija, objeto fijo (CFOF)
2. Cámara fija, objeto móvil (CFOM)
3. Cámara móvil, objeto fijo (CMOF)
4. Cámara móvil, objeto móvil (CMOM)

Para el análisis de las secuencias de imágenes se requieren distintas técnicas en cada una de las clases anteriores. El primer caso es simplemente un análisis de escena estática.

Una secuencia de imágenes proporciona mucha más información para la ayuda del entendimiento de una escena pero incrementa significativamente la cantidad de datos a ser procesados por el sistema. La aplicación de técnicas de análisis de escenas estáticas en cada cuadro, requiere una gran cantidad de cómputo. Afortunadamente, se ha comprobado que en muchos casos, el análisis de escenas dinámicas es más sencillo que en las escenas estáticas.[24]

El análisis de escenas dinámicas tiene tres fases [24]

- Fase periférica
- Fase de atención
- Fase cognitiva

La fase periférica se enfoca en la extracción de información aproximada, la cual es de mucha ayuda en fases posteriores del análisis. Esta información indica la actividad en una escena y es utilizada para decidir en qué partes de la escena es necesario un análisis más cuidadoso. La fase de atención se concentra en analizar las partes activas de la escena y extraer información del movimiento del objeto, así mismo prepara información para un historial de los eventos que toman lugar en la escena. La fase cognitiva aplica conocimiento acerca de los objetos, movimiento y otros conceptos dependientes de la aplicación para analizar la escena en términos de los objetos presentados y los eventos que están tomando lugar.

La entrada a un sistema de análisis de escenas dinámicas es una secuencia de cuadros, representados por $F(x,y,t)$ donde x y y son coordenadas espaciales en el cuadro presentando la escena en el tiempo t . El valor de la función representa la intensidad del pixel. Se asume que la imagen es adquirida utilizando cámaras localizadas en el origen de un sistema coordinado tridimensional. La proyección utilizada en este sistema centrado en el observador puede ser perspectiva u ortogonal.

Dado que los cuadros son tomados usualmente en intervalos regulares, asumiremos que t representa el cuadro t -ésimo de la secuencia, en lugar del cuadro tomado en el tiempo absoluto t .

3.1.1 Detección de cambios

La detección de cambios en dos imágenes sucesivas es un paso importante en muchas aplicaciones. Cualquier cambio perceptible en una escena resulta en un cambio en los cuadros consecuentes de esa escena. Las características del movimiento pueden ser analizadas si tal cambio es detectado.

La mayoría de las técnicas para el análisis de escenas dinámicas están basadas en la detección de cambios en una secuencia de imágenes. Comenzando con los cambios entre cuadros se puede realizar un análisis de la secuencia. Los cambios pueden ser detectados en distintos niveles: pixel, bordes o región. La detección de cambios al nivel de pixel pueden

ser sumados para obtener información útil con la cual los requerimientos de cómputo en fases posteriores pueden disminuir. [24]

3.1.1.1 Diferencia de imágenes

El método más simple para detectar cambios entre dos cuadros es comparar directamente los píxeles correspondientes entre dos cuadros para determinar si hubo algún cambio. En la forma más simple, una diferencia binaria $DP_{jk}(x,y)$ entre dos imágenes $F_j(x,y)$ y $F_k(x,y)$ es obtenida por:

$$DP_{jk}(x,y) \begin{cases} 1 & \text{si } |F_j(x,y) - F_k(x,y)| > r \\ 0 & \text{de otra forma} \end{cases} \quad 3.1$$

donde r es un umbral.

En una diferencia de imágenes, los píxeles que tienen valor 1 son considerados como el resultado del movimiento del objeto o de cambios de iluminación [24].

3.1.1.2 Umbralización

El proceso de umbralización de una imagen en tonos de gris consiste en seleccionar aquellos píxeles que cumplen con una condición, esta condición puede ser tan simple o tan compleja como se necesite, por ejemplo pueden seleccionarse todos aquellos píxeles que tengan un valor de 255, a esta selección se le conoce como umbralización.

En el caso de imágenes en RGB se trabaja con los tres canales, si se quiere umbralizar el color rojo puro, entonces se seleccionarían los píxeles cuyo valor en el canal R sea 255 y cuyos valores tanto en G y B sean 0.

Con la umbralización se obtiene una imagen binaria, la cual contiene 1's en los pixeles seleccionados y 0's en aquellos que no cumplieron la condición.

3.1.1.3 Segmentación

Uno de los problemas más importantes en un sistema de visión es identificar las subimágenes que representan objetos. Esta operación, natural para personas, es realmente difícil para las computadoras. La partición de una imagen en regiones es llamada segmentación. Idealmente una partición representa un objeto o parte del objeto. Formalmente la segmentación puede ser definida como un método de particionar una imagen, $F(i,j)$ en subimágenes llamadas regiones P_1, \dots, P_k , de tal forma que cada subimagen es un objeto candidato.

Definición 3.1: La segmentación consiste en agrupar pixeles en regiones de tal forma que:

- $\bigcup_{i=1}^k P_i = F(i,j)$ ($\{P_i\}$ es una partición exhaustiva.)
- $P_i \cap P_j = \emptyset \neq j$ ($\{P_i\}$ es una partición exclusiva.)
- Cada región P_i satisface un predicado en consecuencia todos los puntos de una partición tienen una propiedad en común
- Los pixeles que pertenecen a regiones adyacentes, no satisfacen el predicado del punto anterior.

Como se dijo en la definición 3.1, una partición satisface un predicado. Este predicado podría ser tan simple como: “*valor de los pixeles que forman la partición igual a una constante*” pero es más complejo en la mayoría de las aplicaciones.

Una imagen binaria es obtenida aplicando una segmentación a una imagen en escala de grises. Si los valores de la intensidad de un objeto están en un intervalo y los valores de los pixeles del fondo están fuera de este intervalo, una imagen puede ser obtenida utilizando una operación de umbralización que coloca a los puntos en ese intervalo en 1 y los puntos fuera de ese intervalo en 0. Para que la umbralización sea efectiva al separar el

fondo, es necesario que los niveles de intensidad del fondo tengan suficiente contraste y que conozcamos los niveles de intensidad tanto del objeto como del fondo.

3.2 FILTRO DE TAMAÑO

En general, para una imagen binaria es bien sabido que el área A esta dada por:

$$A = \sum_i^n \sum_j^m B[i, j] \quad B \in \{0,1\} \quad 3.2$$

Esta es conocida como el momento de orden cero.

Una imagen de diferencia obtenida utilizando la fórmula 3.1 tiene como resultado usualmente muchos pixeles de ruido. Un filtro de tamaño simple es efectivo para eliminar mucho del ruido en la imagen de diferencia. Los pixeles que no pertenecen a un grupo conectado de un tamaño mínimo existen normalmente debido al ruido y pueden ser filtrados. Solo los pixeles en una imagen de diferencia que pertenecen a un componente más grande que un umbral son retenidos para un análisis posterior. Para la detección de movimiento, este filtro es muy efectivo, pero desafortunadamente también filtra algunas señales deseables, tal como aquellas pertenecientes a objetos móviles pequeños.

3.3 TRANSFORMACIONES PROYECTIVAS

Estas transformaciones modelan la distorsión geométrica que se introduce sobre un plano cuando se toma una imagen del mismo con una cámara digital que responde al modelo de orificio de aguja (pinhole). Bajo una transformación de perspectiva, algunas propiedades geométricas se conservan, tales como colinealidad (una línea recta se proyecta a una recta), mientras otras propiedades no, por ejemplo paralelismo, en general las líneas paralelas no se presentan como tales en la imagen. La geometría proyectiva modela el

proceso de adquisición de la imagen al mismo tiempo que da una representación matemática apropiada para los cálculos.

Se le llama transformación a una función que hace corresponder cada punto del espacio con otro punto del mismo o de otro espacio. Una transformación homográfica es una transformación que conserva la naturaleza de los elementos transformados, es decir, un punto se transforma en otro punto, una recta en otra recta y un plano en otro plano.

Una línea en el plano se representa por una ecuación tal como $ax + by + c = 0$. Por tanto, una línea puede representarse de forma natural por un vector $(a, b, c)^t$. La correspondencia entre líneas y vectores $(a, b, c)^t$ no es uno a uno ya que las líneas $ax + by + c = 0$ y $(ka)x + (kb)y + (kc) = 0$ son la misma para cualquier constante k distinta de cero. De hecho, dos vectores relacionados por una escala global son considerados como equivalentes. Una clase de equivalencia de vectores bajo esta relación de equivalencia de vectores en $R^3 - (0,0,0)^t$ forma el espacio proyectivo P^2 .

Las imágenes de puntos 3D sobre un plano están relacionadas por una homografía por lo tanto se dice que un plano induce una homografía entre dos vistas conocidas, como en la figura 3.1. La homografía transfiere puntos de una vista a la otra.

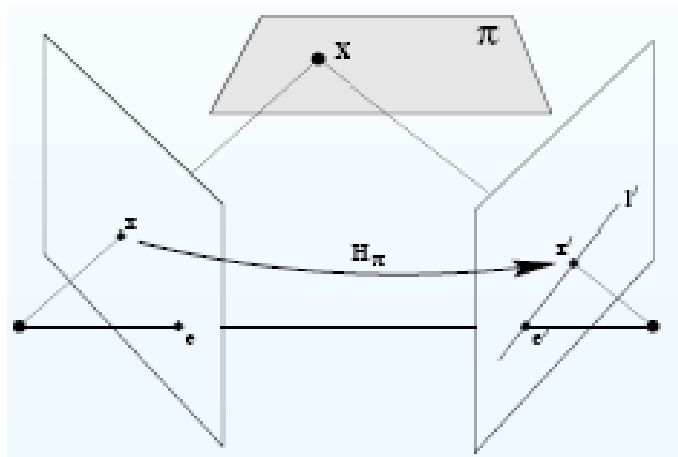


Figura 3.1 Homografía

Existen dos relaciones entre las vistas, a través de la geometría epipolar y la determinada por la homografía del punto anterior. La primera, la geometría epipolar se refiere a la geometría que relaciona una escena y dos proyecciones de ella, un punto en una vista determina su recta epipolar en la otra vista y segundo, un punto determina un punto en la otra a través de la homografía. En la figura 3.1 se muestra un diagrama que representa una geometría epipolar.

La matriz fundamental F es la representación algebraica de la geometría epipolar, es decir, es una matriz de proyección que lleva puntos de una proyección a otra, la matriz fundamental cumple que para cada par de puntos correspondientes $x \leftrightarrow x'$ en las dos imágenes se tiene:

$$x'^T F x = 0 \quad 3.3$$

Suponiendo un plano π que no contenga C, C' , se puede observar en la figura 3.1 que hay una homografía, una proyección, entre los puntos de la imagen I_1 y π ; y entre los puntos de π y I_2 . Para puntos X en π que tengan una imagen x en I_1 ($x = HX$) y una imagen x' en I_2 ($x' = H_0X$), se puede establecer una homografía entre x y x' ($x' = H_\pi x$, $H_\pi = H'H^{-1}$)

Supongamos que tenemos las imágenes de tres puntos X_i en dos imágenes y además conocemos la matriz F de las imágenes, la homografía inducida por el plano de los tres puntos en principio puede calcularse de dos formas distintas:

1. Las coordenadas de los puntos X_i son recuperadas usando algoritmos de reconstrucción, se calcula el plano que pasa por los tres puntos y como consecuencia se calcula la homografía inducida por un plano.
2. Se ataca directamente el problema de estimar H , donde H es la matriz de homografía para los puntos X_i , para ello necesitamos 4 puntos en correspondencia. Dado que tenemos tres puntos en correspondencia, necesitamos un cuarto que será el epípolo de cada imagen que sabemos que por una homografía se corresponden.

Se puede ver la ecuación para determinar H como:

$$F = [e']_x H = H^{-T} [e']_x \quad 3.4$$

Por tanto, el problema se reduce a encontrar la matriz H que haga la transformación.[26]

3.4 MOMENTOS ESTADÍSTICOS

Como se menciona en [25] el momento de orden $(p + q)$ de una imagen digital $f(x,y)$ está definido por:

$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y) \quad 3.5$$

Para $p, q = 0,1,2,\dots$ donde las sumatorias están definidas sobre los valores de las coordenadas espaciales x y y . Los momentos centrales correspondientes están definidos como:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad 3.6$$

donde:

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad 3.7$$

y

$$\bar{y} = \frac{m_{01}}{m_{00}} \quad 3.8$$

El *momento central normalizado* de orden $(p + q)$ está definido por:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\frac{p+q}{2}}} \quad 3.9$$

para $p, q = 0, 1, 2, \dots$, donde

$$\gamma = \frac{p+q}{2} + 1 \quad 3.10$$

para $p + q = 2, 3, \dots$

Un conjunto de siete *momentos invariantes o momentos de Hu en 2D* que son insensibles a la traslación, escala, cambio, espejo y rotación pueden ser derivados de estas ecuaciones [24]. A continuación se muestran 4 de ellos:

$$\begin{aligned} \phi_1 &= \eta_{20} + \eta_{02} \\ \phi_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\ \phi_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\ \phi_4 &= (\eta_{30} - \eta_{12})^2 + (\eta_{21} - \eta_{03})^2 \end{aligned} \quad 3.11$$

3.5 FUSIÓN BAYESIANA

Una forma popular de hacer la fusión de datos es utilizando probabilidad, como en [30] donde se fusionan varias métricas del modo de andar, utilizando fusión basada en el teorema de Bayes.

La fusión Bayesiana se basa en el teorema de Bayes, utilizando las probabilidades otorgadas por los sensores a un evento particular y tomando como probabilidad *a priori* la probabilidad que el sensor otorgó en el paso anterior al mismo evento, o tomando una distribución particular para estas probabilidades.

El teorema de Bayes es el siguiente:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad 3.12$$

Si existen varios eventos A_i que son distinguibles de B en alguna forma, entonces el denominador (B) actúa solamente como una normalización, entonces

$$P(A|B) = \frac{P(B|A)P(A)}{\sum_i P(B|A_i)P(A_i)} \quad 3.13$$

Las ecuaciones 3.12 y 3.13 son conocidas como reglas de Bayes y han dado muchos frutos desarrollando las ideas en fusión de datos.

La ecuación 3.13 denota la probabilidad de que ocurra A dado que ocurrió B , pero se pueden tomar más de un evento para que el evento A ocurra, en ese caso la ecuación de Bayes quedaría de la siguiente forma:

$$P(A|B_1, B_2, \dots, B_i) = \frac{P(A)P(B_1, B_2, \dots, B_i|A)}{P(B_1, B_2, \dots, B_i)} \quad 3.14$$

Y asumiendo independencia condicional entonces:

$$P(B_i|A, B_j) = P(B_i|A) \quad 3.15$$

De la ecuación 3.30 se asume que el denominador es una constante y es despreciado, por lo tanto se tiene que la ecuación de Bayes queda de la siguiente forma:

$$\begin{aligned} P(A|B_1, B_2, \dots, B_i) &= P(A)P(B_1|A)P(B_2|A)P(B_3|A) \dots \\ &= P(A) \prod_{i=1}^n P(B_i | A) \end{aligned} \quad 3.16$$

El valor dado por cada sensor en el tiempo anterior es tomado como probabilidad *a priori*, con esta probabilidad se calcula la probabilidad de que el evento ocurra. Posteriormente, para tomar una decisión se toma el valor más alto de las probabilidades resultantes.

Al observar la ecuación 3.16 se tiene que la complejidad en tiempo es de orden $O(nm)$ donde n es el número de eventos y m es el número de sensores.

3.6 MODELO DEZERT-SMARANDACHE

La teoría Evidencial o Teoría de Evidencia es una rama de las matemáticas del razonamiento con incertidumbre que permite considerar e integrar información para llegar a tomar decisiones. La teoría fue propuesta primero por Arthur Dempster y posteriormente modificada por Glen Shafer, por lo que también se conoce como Teoría Evidencial Dempster-Shafer (DST), en [6] se puede encontrar una descripción de esta teoría, y puede ser interpretada como una generalización de la teoría de probabilidad, donde la asignación de probabilidad se hace sobre conjuntos en lugar de elementos mutuamente exclusivos. En algunos trabajos esta teoría es comparada con la teoría bayesiana.

La aplicación de la Teoría Evidencial se ha visto reflejada en diversos campos. Por ejemplo, Guido Fioretti la compara con la teoría Shackle para poder tomar decisiones económicas, por ejemplo cuándo invertir en farmacéutica y cuándo en biotecnología. Sin embargo, la aplicación más generalizada en los días recientes es en la fusión de información. En este sentido, la parte principal de la DST es la regla de fusión de Dempster.

El modelo Dempster-Shafer demostró dar resultados útiles aplicado a fusión de información y sus principales ventajas sobre la teoría de probabilidad bayesiana son no requerir de conocimiento previo y manejar información incompleta e incierta [6]; además de que en la DST, la evidencia es representada con la función de creencia de Shafer, en lugar de una función de densidad de probabilidad. Sin embargo, cuando la evidencia proporcionada por las fuentes tiene ciertas condiciones, se producen deficiencias en los resultados de la combinación, como el conocido resultado contraintuitivo. Existe una infinidad de casos donde la regla de Dempster puede asignar certidumbre a una opinión

minoritaria. Por esto, trabajos posteriores como [39,40,41,42] propusieron reglas de combinación diferentes o incluso llegaron a generar nuevos modelos que no requieren las probabilidades *a priori* del modelo bayesiano [38], pero que fueran robustas a cualesquiera condiciones de la evidencia. Por ello en la actualidad el termino “Teoría Evidencial” no se restringe a la Teoría Dempster-Shafer.

Una alternativa a los problemas encontrados en la DST, es la Teoría Dezert-Smarandache (DSmT) [39], que toma como base a la DST, pero la extiende tanto en el modelo como en la regla de fusión para resolver problemas originados por fuentes de evidencia conflictivas, inciertas e imprecisas. Siendo un modelo matemático, éste no contempla ninguna característica relacionada con una aplicación en específico. Por ello es que para cada aplicación es necesario extender las funciones para tratar con problemas particulares.

3.6.1 Teoría Dempster-Shafer

La Teoría Dempster-Shafer se basa en el modelo de Dempster-Shafer [6], que fue una modificación del modelo original de Dempster. En el modelo Dempster-Shafer se define un marco de discernimiento $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ con n hipótesis θ_i que cumplen la exhaustividad y exclusividad. Lo que las hipótesis θ_i representan depende del contexto del problema que se está resolviendo.

Partiendo de Θ se define un conjunto 2^Θ llamado *conjunto potencia de Θ* .

El modelo de Shafer considera tres funciones:

La *función de creencia básica*, definida como:

$$m: 2^\Theta \rightarrow [0,1]$$

3.17

La ecuación 3.11 es asociada con un cuerpo de evidencia dado. Esto quiere decir que la evidencia nueva es representada con esta función, la cual es matemáticamente similar a la función de probabilidad.

$$\begin{aligned} m(\emptyset) &= 0 \\ \sum_{A \in 2^\Theta} m(A) &= 1 \end{aligned} \tag{3.18}$$

Shafer definió las funciones de creencias y plausibilidad en $A \subseteq \Theta$ como:

$$Bel(A) = \sum_{B \in 2^\Theta, B \subseteq A} m(B) \tag{3.19}$$

$$Pl(A) = \sum_{B \in 2^\Theta, B \cap A \neq \emptyset} m(B) = 1 - Bel(A^c) \tag{3.20}$$

Cada observación hecha por una fuente de información provee evidencia sobre uno o más subconjuntos de todas las proposiciones de interés y es posible representar ambigüedad o ignorancia asignando evidencia a Θ . Esta información es combinada mediante la regla de combinación de Dempster, lo cual es considerado algunas veces una generalización de la regla de Bayes.

3.6.1.1 Regla de Dempster

La Regla de Dempster como se muestra en [6] es la forma en la que las evidencias de diferentes fuentes se combinan, según el modelo de la DST. Para k fuentes de evidencia, la regla de fusión está expresada de la siguiente manera:

$$m(\emptyset) = 0$$

$$m(A) = \frac{\sum_{x_1, x_2, \dots, x_k \in 2^\theta} \prod_{i=1}^k m_i(X_i)}{1 - \sum_{\substack{x_1, x_2, \dots, x_k \in 2^\theta \\ x_1 \cap x_2 \cap \dots \cap x_k = A}} \prod_{i=1}^k m_i(X_i)} \quad \forall A \neq \emptyset \in 2^\theta \quad 3.21$$

Donde k es el número de fuentes de evidencias, m_i es la evidencia que otorga el sensor i -ésimo y X_i es el evento i -ésimo al que se le está dando evidencia. La normalización hecha con el denominador de la ecuación 3.7, sirve para eliminar las partes conflictivas de información entre las fuentes a combinar. El grado de conflicto entre dos fuentes está definido por:

$$k_{12} = \sum_{\substack{X, Y \in 2^\theta \\ X \cap Y = \emptyset}} m_1(X) m_2(Y) \quad 3.22$$

Cuando dos fuentes tienen evidencias que no son coincidentes $k_{12} = 1$ se dice que los cuerpos de evidencia están en *completa contradicción*. Tal caso sucede cuando $\text{Bel}_1(A)=1$ y además $\text{Bel}_2(A^c) = 1$, para algún $A \subset \Theta$. En general, cuando el conflicto entre las fuentes se vuelve alto (cuando k_{12} se acerca a 1) la combinación hecha con la regla de Dempster muestra poca lógica, a lo que se le conoce como resultado contraintuitivo.

La teoría evidencial fue desarrollada como una alternativa de la teoría de probabilidad, debido a que algunas situaciones no pueden ser modeladas con las herramientas de la probabilidad, como es el caso de la ignorancia. En la teoría evidencial los grados de creencia se asignan a subconjuntos, en lugar de a elementos del dominio de referencia o hipótesis. En su funcionamiento, la teoría evidencial no depende de probabilidades *a priori* para la combinación de fuentes de información. Para ejemplificar la

diferencia entre cómo se consideran las creencias en la teoría probabilista y como se hace en la teoría evidencial, consideremos el caso en el que dos amigos apuestan sobre un partido de fútbol, uno de ellos conoce de fútbol y sabe que la probabilidad de que gane el equipo A es mayor a la probabilidad de que gane el equipo B, pero el otro no se sabe de fútbol y en un modelo probabilista se tendría por fuerza que asumir que la probabilidad de ganar es la mismo para ambos equipos. En la teoría evidencial se puede representar la confianza que tiene cada participante en su apuesta, ya que se puede representar la ignorancia.

3.6.2 Teoría de Dezert-Smarandache

Debido a las limitaciones de la Teoría Dempster-Shafer, Jean Dezert y Florentin Smarandache propusieron en el 2002 un nuevo modelo que terminara con las limitaciones de la DST [39]. Descrito de forma general, el objetivo era tener un marco de discernimiento menos restringido y una regla de fusión que tratara con fuentes conflictivas.

El resultado fue un modelo matemático con tres características que no tiene el modelo DS.

- ❖ La combinación de evidencia que refuta el tercer principio lógico o principio del tercer medio excluido en relación a los elementos pertenecientes al conjunto potencia de Θ en la DST. El tercer principio lógico establece que algo “es o no es”. La DSMT no requiere de este principio, por lo que se permite considerar conceptos que se relacionan con la lógica difusa, que no tienen interpretación absoluta, como grande/pequeño, caliente/frío, etcétera.
- ❖ Se propone una nueva regla de combinación
- ❖ Los elementos del marco de discernimiento no son necesariamente exhaustivos y exclusivos.

En algunos problemas donde es necesario combinar información los conceptos involucrados no tienen una división marcada, tales como grande/pequeño, caliente/frío. Por

ello el modelo DSMT propone un modelo donde las hipótesis o elementos del marco de discernimiento θ_i $i=1, \dots, n$ sólo tienen el requerimiento de ser exhaustivos, es decir, se pueden traslapar. A este modelo se le denomina *modelo libre DSMT* y se denota como $\mathcal{M}^f(\Theta)$.

El modelo DSMT además se considera cuando el marco de discernimiento Θ es dinámico o varía con el tiempo. En ese caso algunos elementos de Θ podrían ser exclusivos en un momento, pero después no existir en otro. Por ello el modelo $\mathcal{M}^f(\Theta)$ puede restringirse para dar origen a un *modelo híbrido DSMT* $\mathcal{M}(\Theta)$ que es idéntico al modelo DS. El *modelo híbrido* se define introduciendo restricciones sobre algunos elementos del *modelo libre*; pueden utilizarse varios tipos de restricciones, dependiendo del problema que se esté abordando, por ejemplo, un elemento de Θ puede definirse como vacío si se conoce la naturaleza de tal elemento y se sabe que no aporta información en ese momento.

Los fundamentos de la DSMT, en cuanto a combinación de información, son diferentes al resto de los trabajos propuestos para resolver los problemas de la DST, en cuanto a la forma de manejar incertidumbre, imprecisión y conflictos, que además promete cubrir una variedad más amplia de problemas.

3.6.2.1 Conjunto Hiperpotencia

Así como la DST tiene un conjunto potencia sobre el cual se consideran las evidencias de las fuentes, la DSMT considera un conjunto hiperpotencia D^Θ , que se construye a partir de un conjunto de n elementos exhaustivos $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$, llamado *marco*. Los elementos de D^Θ dependen del problema que se quiera resolver, estos elementos pueden verse como proposiciones simples.

El conjunto hiperpotencia D^Θ , es construido a partir de θ siguiendo las tres reglas siguientes:

1. $\emptyset, \theta_1, \theta_2, \dots, \theta_n \in D^\Theta$

2. Si $A, B \in D^\Theta$, entonces $A \cap B \in D^\Theta$ y $A \cup B \in D^\Theta$
3. Ningún otro elemento pertenece a D^Θ , excepto los obtenidos mediante 1 y 2.

Esta definición del conjunto hiperpotencia D^Θ cumple con propiedades diferentes a las del conjunto potencia en la DST. La cardinalidad de D^Θ , cuando Θ tiene n elementos, está acotada por 2^{2^n} , por lo que $|D^\Theta| \geq |2^\Theta|$

Con estas reglas se pueden listar un par de ejemplos de conjuntos hiperpotencia:

$$\text{Para } \Theta = \{\theta_1\}, D^\Theta = \{\emptyset, \theta_1\}$$

$$\text{Para } \Theta = \{\theta_1, \theta_2\}, D^\Theta = \{\emptyset, \theta_1, \theta_2, \theta_1 \cap \theta_2, \theta_1 \cup \theta_2\}$$

La cardinalidad del conjunto hiperpotencia sigue la secuencia de los números de Dedekind: 1, 2, 5, 19, 167, 7580, 7828353, 56130437228687557907787..., para un mayor estudio de la cardinalidad del conjunto hiperpotencia véase [6] en la parte de “generación del conjunto hiperpotencia”.

3.6.2.2 Funciones de creencia generalizadas

Partiendo de un marco general Θ se define la función $m: D^\Theta \rightarrow [0,1]$, llamada función de *asignación de creencia básica generalizada* (gbba), aplicada sobre un cuerpo de evidencia $A \in D^\Theta$:

$$m(\emptyset) = 0$$

$$\sum_{A \in D^\Theta} m(A) = 1$$

3.23

Las funciones de *creencia* y *plausibilidad generalizadas* también son usadas, tal como en la DST, pero sus definiciones se hacen usando el conjunto hiperpotencia D^Θ en lugar de D^θ .

$$Bel(A) = \sum_{B \in D^\theta, B \subseteq A} m(B) \quad 3.24$$

$$Pl(A) = \sum_{B \in D^\theta, B \cap A \neq \emptyset} m(B) = 1 - Bel(A^c) \quad 3.25$$

3.6.2.3 Regla de combinación clásica DS_m

Cuando se está considerando el modelo $\mathcal{M}^f(\Theta)$, la regla de combinación usada para un marco Θ con k elementos está definida como:

$$m_{m^{f(\Theta)}}(A) = \sum_{\substack{x_1, x_2, \dots, x_k \in D^\theta \\ x_1 \cap x_2 \cap \dots \cap x_k = A}} \prod_{i=1}^k m_i(X_i) \quad 3.26$$

3.6.2.4 Regla de combinación híbrida DS_m

En algunos problemas de fusión es necesario considerar restricciones sobre los conjuntos a los que les es asignada evidencia, por lo que el modelo $\mathcal{M}^f(\Theta)$ no es suficiente y se tiene que usar en su lugar el modelo híbrido DS_m $\mathcal{M}(\Theta)$ que incluye restricciones que mantienen la integridad de los resultados.

Para este modelo híbrido la regla de combinación para dos o más fuentes está definida para cualquier $A \in D^\Theta$ por las siguientes funciones:

$$m_{\mathcal{M}(\Theta)}(A) = \phi(A)[S_1(A) + S_2(A) + S_3(A)] \quad 3.27$$

$$S_1(A) = \sum_{\substack{x_1, x_2, \dots, x_k \in D^\Theta \\ x_1 \cap x_2 \cap \dots \cap x_k = A}} \prod_{i=1}^k m_i(X_i) \quad 3.28$$

$$S_2(A) = \sum_{\substack{x_1, x_2, \dots, x_k \in \emptyset \\ [U=A] \wedge [[U \in \Theta] \wedge [A = I_t]]}} \prod_{i=1}^k m_i(X_i) \quad 3.29$$

$$S_3(A) = \sum_{\substack{x_1, x_2, \dots, x_k \in D^\Theta \\ x_1 \cup x_2 \cup \dots \cup x_k = A \\ x_1 \cap x_2 \cap \dots \cap x_k \in \emptyset}} \prod_{i=1}^k m_i(X_i) \quad 3.30$$

donde $\phi(A)$ es la *función característica de no-vacio* de un conjunto A definida como:

$$\phi(A) = \begin{cases} 1 & \text{si } A \notin \emptyset \\ 0 & \text{otro caso} \end{cases} \quad 3.31$$

donde \emptyset es el conjunto de todos los elementos del conjunto hiperpotencia que han sido forzados a ser vacíos debido a las restricciones del modelo \mathcal{M} .

Las variables U e I_t se definen como:

$$U = u(X_1) \cup u(X_2) \cup \dots \cup u(X_k) \quad 3.32$$

$$I_t = \theta_1 \cup \theta_2 \cup \dots \cup \theta_n$$

donde a su vez la función $u(X)$ es la unión de todos los *singletons* θ_i que componen $X \in D^\emptyset$.

Cada una de las funciones S tiene un propósito dentro del modelo $\mathcal{M}(\Theta)$: $S_1(A)$ corresponde a la regla clásica de combinación de la DSMT para k fuentes independientes basada en el modelo libre. $S_2(A)$ representa la masa de todos los conjuntos vacíos, que es transferida a las ignorancias. $S_3(A)$ transfiere la suma de los conjuntos relativamente vacíos a los conjuntos no-vacíos.

Tomando en cuenta que el cálculo de las evidencias se tiene que realizar para cada elemento del conjunto hiperpotencia, tenemos que la complejidad de la DSMT es $O(2^{2^n})$

3.7 FUSIÓN MULTICÁMARA PARA SEGUIMIENTO DE OBJETOS BASADA EN TEORÍA EVIDENCIAL

En el trabajo de Esteban García [4] fusiona la información proveniente de dos cámaras para tener mayor certeza de que el objeto que está siguiendo se encuentra en una posición de la escena, sobre la escena se han definido áreas de interés, sobre las cuales se sitúa el objeto de interés.

Las cámaras son independientes la una de la otra, es decir que las observaciones que hace una cámara respecto a la escena no dependen de lo que la otra cámara dice de la misma escena, entonces cada cámara otorga valores de creencia sobre el objeto a seguir,

con estos valores se crean los marcos de discernimiento que sirven para hacer la fusión de información más tarde.

Los valores de creencia se calculan de acuerdo al eje vertical perteneciente al objeto, sin embargo el largo de este eje depende del ángulo que la cámara tiene respecto al plano piso, si la cámara está mas cerca del plano piso el eje vertical es más grande y si la cámara está sobre el objeto el eje vertical es más pequeño, esto se conoce como incertidumbre, ya que si la cámara está mas próxima al plano piso, es más complicado saber sobre que parte del plano piso se encuentra al objeto, entre más bajo esté la cámara mayor incertidumbre se tiene.

Una vez obtenido el marco discernimiento, se procede a hacer la fusión de información, posteriormente se toma una decisión con el resultado de la fusión.

El enfoque utilizado por Esteban García [4] utiliza cámaras fijas y por esto, presenta problemas cuando el objeto se ocluye parcial o totalmente, además, con las cámaras estáticas se abarca una porción más pequeña de la escena que con las cámaras en movimiento. En este trabajo se utilizan cámaras en movimiento para lograr abarcar una mayor área de la escena, sin embargo, con este movimiento de las cámaras es posible lograr ver objetos que están ocluidos. A diferencia del trabajo de E. García en este trabajo se utiliza acumulación de evidencias, con lo cual es posible tener una mayor certeza sobre el objeto que se está siguiendo.

Capítulo 4

Resumen

En este capítulo se presenta la propuesta de solución para lograr el seguimiento de los objetos con sensores móviles, se muestran ejemplos con imágenes en algunos de los pasos seguidos para la solución del problema.

PROPUESTA DE SOLUCIÓN

4.1 ESQUEMA DEL METODO PROPUESTO

Para alcanzar el objetivo propuesto se presenta el siguiente método, con el cual se pretende alcanzar el objetivo del seguimiento confiable de objetos.

1. Adquisición de imágenes
2. Identificación del objeto a seguir
3. Eliminación de ruido
4. Cálculo de la homografía entre cámaras
5. Evaluación de la porción visible del objeto
6. Calculo de su nivel de creencia
7. Fusión de creencias

Para el punto uno se determinó utilizar dos cámaras para hacer la adquisición de las imágenes, las secuencias de imágenes fueron adquiridas en formato mpeg, posteriormente cada cuadro fue guardado en formato jpg.

En el punto dos, para identificar al objeto a seguir se utilizaron dos métodos, uno es por diferencias de imágenes y el otro es por medio de segmentación de color.

En el punto tres, para la eliminación de ruido se utilizó un filtro de tamaño, logrando así eliminar las regiones de tamaño menor que el objeto encontrado.

En el punto cuatro, para el cálculo de la homografía se utilizaron dos métodos, el primero consiste en encontrar 4 puntos de control que se correspondan entre los campos de vista de las cámaras, con las coordenadas de estos puntos se encuentra una matriz de transformación que se aplica a cada punto de una imagen para hacer la homografía. El segundo consiste en realizar transformaciones de una imagen, las transformaciones consisten en una rotación y una trasformación.

En el punto cinco, se determina que parte del objeto está oculto, esto se realiza mediante el cálculo del eje horizontal del objeto, mientras mayor sea la parte del objeto que se observa el eje horizontal se hace más grande y con este cálculo se tiene más certeza sobre el objeto.

En el punto seis, para el cálculo de la creencia del objeto, se divide la imagen en rejillas, se obtiene el eje vertical del objeto y se calculan los pixeles de este eje, haciendo una intersección entre el eje y las rejillas se determina cuantos pixeles pasan por cada rejilla, estos pixeles son normalizados y de esa forma se obtienen las creencias.

En el punto siete, se realiza la fusión de creencias obtenidas mediante los pasos anteriores, esto se logra mediante la implementación del método de fusión DSMT. Finalmente, se toma el valor mayor de los resultados para hacer la decisión.

4.2 ARREGLO EXPERIMENTAL

Se tienen varias cámaras dispuestas alrededor de una escena, las cámaras se encuentran en posiciones distintas de tal forma que las vistas son distintas también. Estas cámaras están en movimiento, el movimiento se realiza para un experimento en línea recta y para otro alrededor del objeto a seguir, en ambos casos a una velocidad constante. Las cámaras se mueven sobre un riel haciendo que el campo de vista cambie mientras las cámaras se mueven.

En la escena se encuentran objetos móviles que se les muestran a las cámaras, el objetivo es que el sistema logre seguir a estos objetos y mediante una fusión de información se logre una mejor decisión sobre el resultado del sistema.

4.3 EXPERIMENTOS REALIZADOS

4.3.1 Movimiento de las cámaras

4.3.1.1 Movimiento horizontal

Se tomaron imágenes del repositorio PETS [9] que se encuentra disponible en la red. Este repositorio cuenta con secuencias de imágenes con varias cámaras observando una escena con objetos en movimiento. En especial se utilizó una secuencia de imágenes tomadas con dos cámaras dispuestas a 90° una de la otra, estas cámaras observan la misma escena al mismo tiempo, es decir, están alineadas temporal y espacialmente. Las imágenes de estas secuencias no fueron tomadas con cámaras móviles, por lo cual fue necesario crear su movimiento virtualmente, haciendo una translación a las imágenes de 5 píxeles entre cada cuadro de tal forma que se apreciara un movimiento tanto vertical como horizontal (Fig. 4.1).

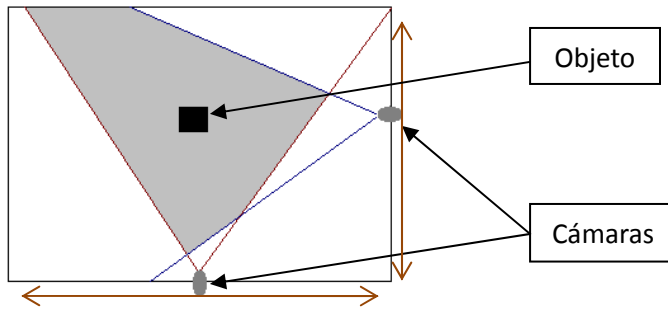


Figura 4.1. Movimiento horizontal de las cámaras

Con la simulación de movimiento horizontal o en línea recta se tienen limitaciones que con el movimiento real de las cámaras no se tendrían, por ejemplo que el campo de visión no cambia realmente con la simulación sin embargo el proceso que se utilizó en la simulación asemeja muy cercanamente el movimiento real de las cámaras. Dado que la simulación del movimiento no es del todo fiel al movimiento real, y dado que no se tuvo los recursos suficientes para realizar el experimento real en un entorno más grande, se hicieron experimentos en pequeña escala del movimiento horizontal de las cámaras. Los algoritmos utilizados para la detección de movimiento en la simulación fueron los mismos que se utilizaron en los experimentos con movimiento real y estos algoritmos funcionaron correctamente.

4.3.1.2 Movimiento circular

Se construyó una plataforma donde colocar las cámaras y lograr que se movieran en forma circular alrededor de la escena. Las cámaras están separadas por un ángulo de 90° entre ellas y por 1.60 m del centro de la escena. Las cámaras rotan alrededor de la escena con una velocidad constante (Fig. 4.2).

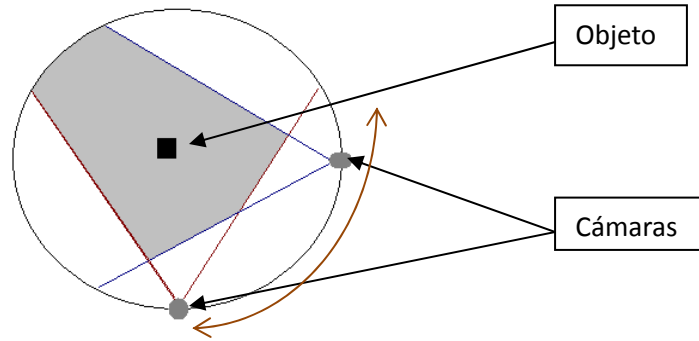


Figura 4.2. Movimiento circular de las cámaras

4.3.1.3 Detección de movimiento con cámaras en movimiento horizontal

Para la segmentación de los objetos se utilizó el método conocido como resta de imágenes, para lograr esto, se obtuvieron imágenes de la escena para cada cámara en la cual no existiera ningún objeto en movimiento, estas son las plantillas que se utilizan para hacer la resta con las imágenes donde se muestran objetos en movimiento.

El movimiento de la cámara es continuo, sin embargo las imágenes son tomadas en intervalos constantes de tiempo, es decir, son tomadas discretamente, entonces el movimiento que realiza la cámara desde el punto inicial hasta el punto final o de retorno está dividido en n tomas de imágenes o fotogramas. Para realizar la resta de imágenes y detectar el movimiento de los objetos, se hace la resta de la imagen tomada en la posición actual de la cámara i -ésima, contra la plantilla i -ésima, es decir, existe una plantilla por cada posición posible de la cámara.

De ésta forma es posible detectar cambios de una imagen a otra, sin embargo existen regiones de la imagen que debido al cambio en la intensidad de luz o por la calidad de la cámara se pueden detectar como movimiento en la imagen. Estas regiones pueden alterar el comportamiento del sistema, por lo que es necesario hacer un procesamiento posterior para eliminar el ruido.

Como ya se comentó anteriormente, las imágenes pertenecientes al repositorio PETS, fueron tomadas con sensores estáticos, por lo que se implementó un algoritmo para lograr simular el movimiento de las cámaras.



Figura 4.3 Plantilla



Figura 4.4 Imagen de la cámara 2



Figura 4.5 Imagen resultante de la resta de las imágenes 4.3 y 4.4

Este tipo de movimiento simulado se utilizó para mostrar que es posible realizar el seguimiento del objeto, la correlación y fusión de datos con cámaras móviles. También que este movimiento de las cámaras sirve para observar objetos ocultos (ver sección 5.3).

4.3.1.4 Detección de objetos con cámaras en movimiento circular

Para este caso se utilizó una segmentación basada en color (sección 3.1.1.2), dado que el fondo no es estable, es decir, presenta mucho movimiento (Fig. 4.6). Se utiliza este tipo de segmentación, porque se sabe que el objeto que se está siguiendo es de un color específico (amarillo). El movimiento de los sensores no afecta la utilización de éste método de segmentación ya que el objeto que se quiere detectar o segmentar es de un color definido.

Dado que las imágenes utilizadas fueron tomadas en el exterior, las condiciones no son controladas, por lo tanto es necesario hacer un procesamiento posterior a la segmentación para eliminar el ruido causado por estas condiciones.

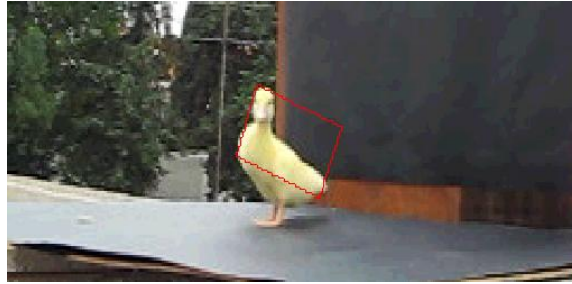


Figura 4.6 Detección por color

4.4 ELIMINACIÓN DE RUIDO

Dado que existen cambios en la iluminación entre las plantillas utilizadas y las imágenes tomadas por los sensores durante el movimiento de los mismos, existen pixeles que no tienen valor 0 después de la resta de imágenes o después de la umbralización, por lo tanto, se debe realizar una eliminación del ruido, para que el ruido no altere los procedimientos posteriores a este paso.

El ruido presente en las imágenes posteriores a la segmentación es un ruido conocido como sal y pimienta, éste tipo de ruido se caracteriza por ocurrencias de valores negros y blancos al azar en la imagen[24], existen técnicas para la eliminación de este tipo de ruido, tal como el filtro de tamaño, el cual toma las regiones resultantes del paso anterior y selecciona aquellas regiones que cumplen con la condición de que su área es mayor a un valor determinado[24]; el valor para cada una de las diferentes secuencias varía, ya que en ocasiones los sensores se encuentran muy lejos de la escena y los objetos que deben

seguirse son pequeños, por el contrario en ocasiones los sensores están cerca de la escena y el objeto a seguir es más grande.

Se utilizó el filtro de tamaño porque es sabido que el objeto de interés tiene un tamaño determinado, por lo tanto cualquier objeto que sea mayor o menor al tamaño del objeto de interés no es tomado en cuenta. Por ejemplo si el tamaño del objeto de interés es de t píxeles, el filtro de tamaño eliminará cualquier objeto que sea menor a t y cualquier objeto que sea mayor a t .

Se utilizaron distintos valores para el filtro de tamaño, éstos se presentan en la tabla 4.1.

Tipo de secuencia	Área en píxeles	
	Valor mínimo	Valor máximo
1	150	2500
2	500	1500
3	500	2700

Tabla 4.1 Áreas utilizadas para el filtro de tamaño.

4.4 CORRESPONDENCIA DE DATOS

Es necesario hacer que los objetos detectados en uno de los sensores se hagan corresponder con los objetos detectados en el otro sensor, esto con el fin de que la información sea fusionada correctamente y no se tenga el problema de fusionar la información que no corresponde a los mismos objetos.

Cuando un sensor S_1 genera información acerca de los objetos que está observando, se necesita saber qué información generada por el otro sensor S_2 corresponde con la primera, es decir, si el objeto i del sensor S_1 es el objeto j del sensor S_2 . Una vez que se sabe que el objeto i del sensor S_1 y el objeto j del sensor S_2 es el mismo objeto, se dice que los objetos se corresponden. Para este propósito los objetos se representan con datos como

su área, perímetro, coordenadas de su centro, cuando se logra saber que los datos asociados con el i del sensor S_1 y los datos del objeto j del sensor S_2 son los mismos dentro de un rango, entonces se dice que se tiene una correspondencia de datos.

El movimiento de los sensores hace que la correspondencia de los datos sea distinta de la que se emplea cuando las cámaras están estáticas; cuando las cámaras están en movimiento, el cálculo de la correlación debe hacerse cada vez que las cámaras se mueven ya que el movimiento hace que los parámetros que se utilizan para calcular la correlación cambien. En el caso de la homografía, las coordenadas de los puntos de control cambian y por tanto es necesario recalcular la matriz que representa esta homografía.

4.6 CREACIÓN DEL MARCO DE DISCERNIMIENTO

Con los objetos obtenidos del paso anterior, se procede a encontrar el eje vertical de los mismos, esto se logra al encontrar el rectángulo más pequeño que encierre al objeto, posteriormente se encuentra el punto medio de los lados horizontales, entre estos puntos medios se traza una línea que los une, esta línea se conoce como el eje vertical. En la figura 4.7 se muestran el rectángulo más pequeño que encierra al objeto (rojo) y el eje vertical del objeto (verde).



Figura 4.7 Rectángulo y eje vertical del objeto

La imagen se divide en celdas de 20 X 20 píxeles, al conjunto de estas celdas se le denomina rejilla, con esta rejilla es posible crear un marco de discernimiento. Se toma el

eje vertical de cada objeto y se toman las celdas por las que pasa el eje. A partir del conjunto de celdas por las cuales pasa el eje vertical se crea el marco de discernimiento. Si el eje vertical del objeto pasa por n celdas, entonces el marco de discernimiento tendrá n elementos. En la figura 4.8 se puede observar como los elementos del marco de discernimiento θ corresponden a las celdas por las que pasa el eje vertical del objeto.

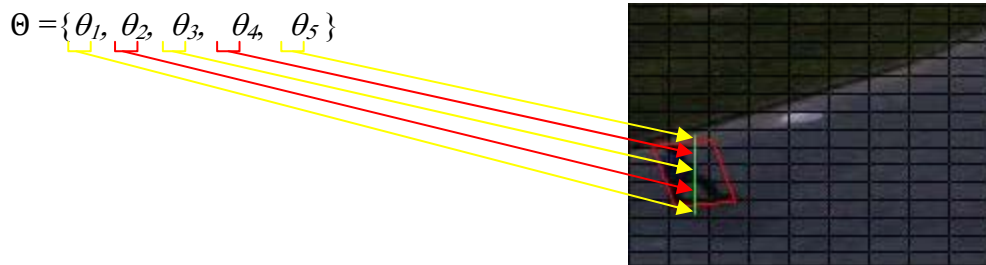


Figura 4.8 Creación del marco de discernimiento.
Cada celda es un elemento del marco.

Al utilizar los momentos estadísticos para determinar la parte visible del objeto se utilizaron los dos primeros momentos, al primero corresponde al área del objeto y el segundo al perímetro del objeto, sin embargo este enfoque no permite crear un marco de discernimiento en donde colocar las observaciones de los sensores, por lo que se prefirió utilizar el eje horizontal para determinar que se veía una mayor parte del objeto. El cálculo del eje vertical es el mismo que el cálculo para el eje vertical, solamente que el eje horizontal es trazado horizontalmente y corresponde a la línea media horizontal que pasa por en medio del objeto.

Con el eje horizontal, a medida que el objeto comienza a verse en mayor proporción, el eje vertical crece logrando determinar que una mayor parte del objeto es visible. Esto es posible ya que las cámaras se mueven horizontalmente y el objeto oculto se descubre con este tipo de movimiento.

4.7 CREACIÓN DE MASAS

Las masas son evidencias (valores) que son necesarias para la fusión de información, se toma el marco de discernimiento creado anteriormente por cada cámara, y se crean valores que representan las evidencias que otorgan las cámaras a esos eventos, en seguida se explica la forma en que se obtuvieron estos valores.

Estos valores nos dan evidencia perteneciente a objetos que no son parte del fondo de la imagen, con estos valores se harán posteriormente operaciones que den como resultado mejores evidencias sobre los objetos.

Se utilizaron dos formas para encontrar las masas o evidencias de los objetos en la escena:

1. Dado el eje vertical del objeto y las rejillas en las que se dividió la imagen, se hace un conteo de los pixeles del eje vertical perteneciente a cada rejilla, estos valores se normalizan y se toman como las masas de cada región.

Como ejemplo, supongamos que el marco de discernimiento contiene 3 elementos

$$\Theta = \{\theta_1, \theta_2, \theta_3\} \quad 4.1$$

los cuales representan las celdas del eje vertical del objeto, si las celdas tienen 20 pixeles de alto y θ_1 abarca 10 pixeles, θ_2 abarca 20 pixeles y θ_3 abarca 15 pixeles, entonces tenemos que los valores del marco de discernimiento son los siguientes:

$$\Theta = \{10, 20, 15\} \quad 4.2$$

Normalizando tenemos:

$$\Theta = \left\{ \frac{10}{45}, \frac{20}{45}, \frac{15}{45} \right\} \quad 4.3$$

$$\Theta = \{0.222, 0.444, 0.333\} \quad 4.4$$

Entonces ahora tenemos que las masas del marco de discernimiento son las que se muestran en la ecuación 4.3

2. Cada objeto encontrado es una región dentro de la imagen, como región puede tener varias características que se pueden utilizar, en este caso se utilizaron los momentos estadísticos 1 y 2 como las masas del objeto.

Tomemos el siguiente caso como ejemplo, dado el marco de discernimiento siguiente:

$$\Theta = \{\theta_1, \theta_2\} \quad 4.5$$

El marco de discernimiento en este caso tiene solamente dos elementos ya que se están utilizando solamente dos momentos estadísticos. Utilizando Matlab para obtener los momentos estadísticos tenemos los siguientes resultados.

$$\Theta = \{0.8654, 0.1345\} \quad 4.6$$

Entonces, las masas del marco de discernimiento son las que se muestran en la ecuación 4.6.

4.8 CREACIÓN DEL CONJUNTO HIPERPOTENCIA

Del marco de discernimiento se hace un conjunto con una cardinalidad más grande, el cual está formado por uniones e intersecciones de los elementos del marco de discernimiento, este conjunto es importante ya que con él se hace la fusión.

Se tomaron solamente unos elementos del conjunto hiperpotencia, ya que varios de estos elementos no aportan información a la fusión, tal es el caso de la intersección de los elementos en los extremos del marco de discernimiento, éstos no son tomados en cuenta. Estos elementos no tienen sentido, ya que si el eje vertical es muy grande, los extremos no tienen

Como ejemplo tenemos el siguiente caso, donde el marco de discernimiento original es:

$$\Theta = \{\theta_1, \theta_2, \theta_3\} \quad 4.7$$

A partir de este conjunto se forma el conjunto hiperpotencia, el cual consta del vacío, los elementos de Θ , así como de las uniones e intersecciones de Θ ; por lo tanto tenemos:

$$D^\Theta = \{\phi, \theta_1, \theta_2, \theta_3, \theta_1 \cup \theta_2, \theta_1 \cup \theta_3, \theta_2 \cup \theta_3, \theta_1 \cap \theta_2, \theta_1 \cap \theta_3, \theta_2 \cap \theta_3, (\theta_1 \cup \theta_2) \cap \theta_3, (\theta_1 \cup \theta_3) \cap \theta_2, (\theta_2 \cup \theta_3) \cap \theta_1, (\theta_1 \cap \theta_2) \cup \theta_3, (\theta_1 \cap \theta_3) \cup \theta_2, (\theta_2 \cap \theta_3) \cup \theta_1\}, \quad 4.8$$

$$\theta_1 \cup \theta_2 \cup \theta_3, (\theta_1 \cap \theta_2) \cup (\theta_1 \cap \theta_3) \cup (\theta_2 \cap \theta_3)$$

Como se mostró en la sección 3.6.2.1 la cardinalidad de Θ es menor a la cardinalidad de D^Θ , donde $|\Theta| = 2$ y $|D^\Theta| = 16$. Sin embargo, no todas las proposiciones del conjunto hiperpotencia pueden utilizarse en el caso del eje vertical ya que no tienen sentido. Como se muestra en la figura 4.8, los elementos del marco de discernimiento pertenecen a las celdas que ocupa el eje vertical, y están ordenadas, es decir que el elemento uno corresponde a la celda inferior y el último elemento pertenece a la celda superior; por esto, las intersecciones entre elementos que no son continuos no se toman en cuenta, ya que esta intersección no tiene sentido, así mismo, por tanto se pueden eliminar del conjunto hiperpotencia. Eliminando los elementos que no tienen sentido tenemos:

$$D^\Theta = \{\phi, \theta_1, \theta_2, \theta_3, \theta_1 \cup \theta_2, \theta_2 \cup \theta_3, \theta_1 \cap \theta_2, \theta_2 \cap \theta_3, (\theta_1 \cup \theta_2) \cap \theta_3, (\theta_2 \cup \theta_3) \cap \theta_1, (\theta_1 \cap \theta_2) \cup \theta_3, (\theta_2 \cap \theta_3) \cup \theta_1\}, \quad 4.9$$

$$\theta_1 \cup \theta_2 \cup \theta_3$$

Con esto se logran eliminar elementos y los cálculos se hacen más sencillos y más rápidos.

4.9 FUSIÓN DE INFORMACIÓN

Una vez que se tienen las masas y el conjunto hiperpotencia creados se procede a hacer la fusión de estas masas.

El módulo de fusión se creó en Matlab, ya que este lenguaje permite operaciones sobre matrices de manera nativa, la fusión está basada en el modelo híbrido de la teoría Dezert-Smarandache. De forma sencilla se pueden realizar operaciones con imágenes, lo que es vital para realizar el seguimiento de objetos.

Se llamará fusión inter-sensor a la fusión que ocurre dentro de cada sensor y fusión intra-sensor a la fusión que ocurre con la información de un sensor contra la información de otro sensor. La fusión inter-sensor es la que se realiza en cada paso de la fusión con acumulación, es cuando la información del paso anterior se fusiona con la información del paso actual y se queda ahí para el paso siguiente. La fusión intra-sensor es aquella que se realiza en cada paso de la fusión sin acumulación, es la que permite obtener el resultado de las observaciones al final de la fusión.

Es necesario mencionar que dado que los sensores están en movimiento, es necesario reflejar esta ventaja sobre los sensores que no pueden moverse y que la fusión en ambos casos da diferentes resultados. Si se hace fusión con los sensores estáticos, los resultados darán buenos resultados pero en ciertos casos, por ejemplo cuando un sensor no ve al objeto que se está siguiendo, entonces la fusión no tiene razón de ser. Sin embargo, al utilizar sensores móviles, se tiene un mayor campo de visión y es posible seguir al objeto por más tiempo que con los sensores fijos.

Aunque se puede suponer que para hacer la fusión con sensores en movimiento es el mismo método que con sensores estáticos, es necesario hacer notar que no precisamente se tiene el mismo modelo.

- Cuando se tienen sensores estáticos se toman imágenes de la escena cada t intervalo de tiempo y la imagen es sobre la misma escena. Esta información se fusiona en el tiempo t y se tiene un resultado global.
- Cuando se tienen sensores móviles, se toman las imágenes cada t intervalo de tiempo, la diferencia es que cada imagen es una escena que ha cambiado de posición, este cambio se utiliza a favor de la fusión, ya que se fusionan las imágenes $t-1$, t y $t+1$ en cada sensor y posteriormente se hace una fusión global intra-sensor. Con esto se logra una mayor evidencia del objeto.
- Se utiliza el movimiento de los sensores como técnica para abordar el problema de oclusión de los objetos. A diferencia del trabajo de García [4] en el cual se trabaja con sensores estáticos y no resuelven este problema.

Puede suceder que el objeto en la imagen $t-1$ no aparezca en la imagen, sin embargo, con el movimiento de las cámaras el objeto sale a la vista de la cámara en el tiempo t y $t+1$ con lo que al finalizar la fusión intra-sensor la evidencia sobre el objeto tiene sentido.

Capítulo 5

Resumen

En este capítulo se muestran los resultados obtenidos al implementar la propuesta de solución planteada en el capítulo anterior. La solución utiliza fusión DSMT para el seguimiento. Se muestra también una comparación entre los datos obtenidos utilizando fusión DSMT y fusión Bayesiana.

Resultados experimentales

La propuesta de solución descrita en el capítulo previo fue implementada en un sistema prototipo de hardware/software, que permitió mostrar la efectividad de la propuesta en situaciones reales, aún cuando se implementara a una escala pequeña con respecto al escenario. Se puso atención en que esta limitación no le restara generalidad al sistema. En general solamente se necesita realizar ajustes menores para escenarios de mayores dimensiones.

Utilizando los esquemas mencionados en la sección anterior se logró hacer la correspondencia de datos por medio de una matriz de proyección homográfica en el caso de las secuencias con la disposición y el movimiento de los sensores como se ejemplifica en la figura 4.1. La matriz de proyección homográfica se construyó gracias a puntos de control

que fueron dados por la gente que tomó la secuencia de PETS, estos puntos son aquellos que son fácilmente reconocibles tanto en el sensor 1 como en el sensor 2 y que corresponden al mismo punto. Este tipo de correlación ha sido utilizado anteriormente para aplicaciones similares, como en [3], donde se emplea para correlacionar los objetos que se están siguiendo. Sin embargo, a diferencia de [3], el cálculo de la homografía debe realizarse en cada movimiento de las cámaras por los motivos mencionados en el párrafo anterior.

Para las secuencias tomadas con los sensores dispuestos y con el movimiento ejemplificado en la figura 4.2, la correspondencia de datos se logró haciendo transformaciones, éstas son una rotación de 90° y una traslación de 10 píxeles, aplicadas a la imagen del sensor 1 para que correspondieran con la imagen del sensor 2. Las transformaciones aplicadas a las imágenes pueden no funcionar si el ángulo de separación entre los sensores cambia; sin embargo, este problema puede resolverse haciendo la rotación en función de los grados de separación entre los sensores; la traslación aplicada a las imágenes es de 10 píxeles. Las transformaciones que se realizan a las imágenes no se ven afectadas por el movimiento de los sensores, ya que estos conservan el ángulo de separación y la posición sobre la plataforma durante el movimiento.

En este tipo de secuencias se utilizaron transformaciones para realizar la correspondencia de datos, debido a que para este caso no se contaba con puntos de control para realizar la homografía como en el caso anterior; con esta forma de correlacionar los datos también fue necesario recalcular la matriz de transformación en cada movimiento de las cámaras.

En las tablas de las secciones 5.1 y 5.2, se muestran los resultados obtenidos al fusionar la información de las cámaras con movimiento horizontal y circular respectivamente, en la primera columna se muestran las imágenes obtenidas por la cámara 1 y en la segunda columna se muestran las imágenes obtenidas por la cámara 2, en la tercera columna se muestran los marcos de discernimiento tanto de la cámara uno, de la cámara dos y del resultado de la fusión. En la tercera columna se muestran tres elementos, m_1 , m_2 y m :

- m1: es el marco de discernimiento creado por el sensor1, se muestran las creencias asignadas por el sensor a cada elemento del marco de discernimiento.
- m2: es el marco de discernimiento creado por el sensor2, se muestran las creencias asignadas por el sensor a cada elemento del marco de discernimiento.
- m: es el marco de discernimiento al final de la fusión, se muestran las creencias finales de la fusión.

En las figuras 5.1 y 5.2 se dan ejemplos de la correspondencia de los datos. En la figura 5.1 se proyectó el eje del objeto detectado en el sensor 1 hacia el sensor 2; en la figura 5.2 se proyectó el recuadro que envuelve al objeto de la imagen del sensor 1 hacia la imagen del sensor 2.

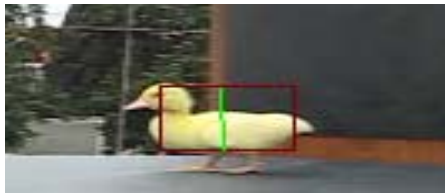


a) Eje vertical del objeto, sensor 1.

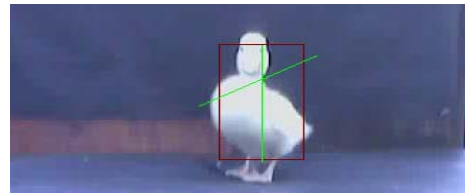


b) Eje vertical del objeto en el sensor 2 y proyección del eje vertical del objeto del sensor 1.

Figura 5.1 Eje vertical del objeto, secuencia con movimiento horizontal.



a) Eje vertical del objeto, sensor 1.



b) Eje vertical del objeto en el sensor 2 y proyección del eje vertical del objeto del sensor 1.

Figura 5.2 Eje vertical del objeto, secuencia con movimiento circular.

5.1 PRUEBAS CON LAS CÁMARAS EN MOVIMIENTO HORIZONTAL

Para las pruebas con las cámaras en movimiento horizontal se utilizaron 3 secuencias de imágenes con 600 fotogramas (cuadros) en cada una de las secuencias. El escenario incluyó objetos en movimiento (personas, animales, etcétera), así como objetos estáticos (árboles, edificios, autos, etc.), el sistema operó con condiciones de iluminación naturales (*outdoor*).

En la tabla 5.1 se muestra parte de una secuencia y los resultados obtenidos. El objeto seguido es una persona que aparece a partir de la fila 5. En la primer columna se tienen las imágenes que observa la cámara 1, en la segunda columna las imágenes que observa la cámara 2, en la tercer columna se tienen los siguientes datos: “m1” corresponde a los datos que obtuvo la cámara 1 a partir de la imagen analizada; es el marco de discernimiento de la cámara 1, “m2” corresponde a los datos que obtuvo la cámara 2 a partir de la imagen analizada; es el marco de discernimiento de la cámara 2, “m” es la salida del proceso de fusión.

En la tabla 5.1 se muestran los resultados de una parte de una secuencia tomada de PETS, esta secuencia fue tomada con cámaras estáticas. Sin embargo, se simuló el movimiento horizontal de las cámaras por medio de software, esto se logró tomando una subimagen en cada imagen. Las imágenes originales son de 768 X 576 píxeles, las subimágenes son de 500 X 576 píxeles.

Recordemos que el marco de discernimiento se forma con la intersección del eje vertical con las rejillas de la imagen, las evidencias se obtienen al contar el número de píxeles del eje vertical que pasan por cada rejilla, como se explicó en la sección anterior.

Tabla 5.1 Secuencia 1 y resultado de la fusión sin acumulación en el tiempo.

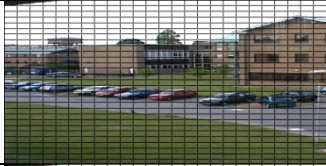

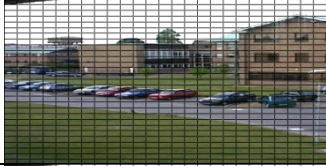

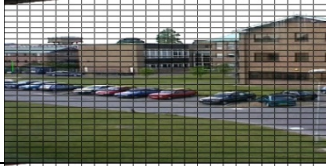

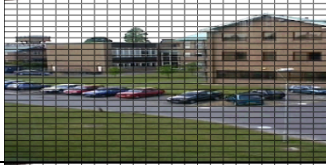

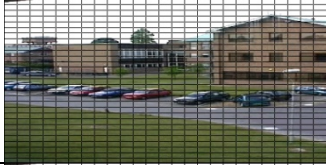

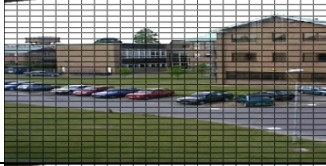



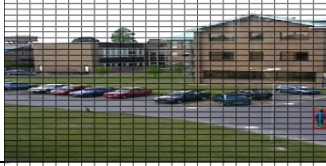

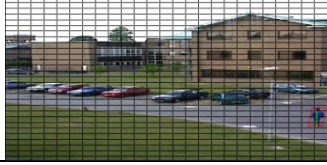

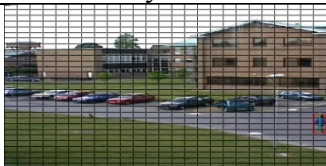

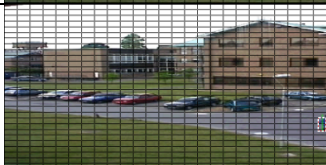

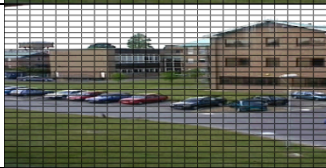

Número de cuadro	Cámara 1	Cámara 2	Marcos de discernimiento y resultado de la fusión (Cada elemento del marco corresponde a una rejilla)
1			$m1 = \{\}$ $m2 = \{\}$ $m = \{\}$
2			$m1 = \{\}$ $m2 = \{\}$ $m = \{\}$
3			$m1 = \{\}$ $m2 = \{\}$ $m = \{\}$
4			$m1 = \{\}$ $m2 = \{\}$ $m = \{\}$
5			$m1 = \{\}$ $m2 = \{ 0.1808 \ 0.1647 \}$ $m = \{ 0.1808 \ 0.1647 \}$
6			$m1 = \{\}$ $m2 = \{ 0.2363 \ 0.2363 \ 0.2184 \ 0.3090 \}$ $m = \{ 0.2363 \ 0.2363 \ 0.2184 \ 0.3090 \}$
7			$m1 = \{\}$ $m2 = \{\}$ $m = \{\}$
8			$m1 = \{ 0 \ 0.3598 \ 0.3312 \ 0 \ 0.3090 \}$ $m2 = \{ 0.1808 \ 0.1647 \ 0.1808 \ 0.1647 \ 0.3090 \}$ $m = \{ \ 0 \ 0.2213 \ 0.2181 \ 0 \ 0.0955 \}$
9			$m1 = \{ 0.2424 \ 0.2243 \ 0.2243 \ 0.3090 \}$ $m2 = \{ 0.2363 \ 0.2363 \ 0.2184 \ 0.3090 \}$ $m = \{ 0.0573 \ 0.1953 \ 0.1858 \ 0.0955 \}$

Tabla 5.1 Secuencia 1 y resultado de la fusión sin acumulación en el tiempo. (continuación)

10			$m1 = \{0.1830 \quad 0.1693 \quad 0.1693$ $0.1693 \quad 0.3090\}$ $m2 = \{0.1727 \quad 0.1727 \quad 0.1727$ $0.1727 \quad 0.3090\}$ $m = \{0.0316 \quad 0.1350 \quad 0.1350$ $0.0292 \quad 0.0955\}$
11			$m1 = \{\}$ $m2 = \{0.1759 \quad 0.1759 \quad 0.1759$ $0.1633 \quad 0.3090\}$ $m = \{0.1759 \quad 0.1759 \quad 0.1759$ $0.1633 \quad 0.3090\}$
12			$m1 = \{\}$ $m2 = \{0.1727 \quad 0.1727 \quad 0.1727$ $0.1727 \quad 0.3090\}$ $m = \{0.1727 \quad 0.1727 \quad 0.1727$ $0.1727 \quad 0.3090\}$

En la tercer columna de la tabla 5.1 se pueden ver las creencias que dan las cámaras, $m1$ son las creencias de la cámara 1 y $m2$ son las creencias de la cámara 2, m es el resultado de la fusión de las creencias. Podemos ver que cuando los dos sensores aportan información sobre las imágenes, el resultado de la fusión tiene un elemento que es mayor al resto, esto significa que la fusión determinó que el objeto tiene una mayor evidencia de encontrarse en la celda que corresponde a ese elemento. Cuando alguno de los dos sensores no aporta información, entonces el resultado de la fusión es la información que aportó el sensor que observó algo, ya que el otro sensor se declara como ignorante. Cuando ocurre que el sensor no detectó ningún objeto entonces no asigna valores a los elementos del marco de discernimiento, podemos ver en la tabla 5.1 en las filas 1,2,3 y 4 que ninguno de los dos sensores ha detectado objeto alguno, por lo tanto no asignan valores a los elemento del marco de discernimiento.

Podemos observar en los resultados que la fusión de los sensores logró su propósito, al final de los cálculos se obtuvieron mayores evidencias sobre algunos de los elementos del marco, con lo cual se puede originar una decisión general sobre el objeto. Se dan más detalles de estos resultados en la sección 5.5.

5.2 PRUEBAS CON LAS CÁMARAS EN MOVIMIENTO CIRCULAR

Para las pruebas con las cámaras en movimiento circular se utilizaron 2 secuencias con 200 fotogramas (cuadros) y una secuencia más con 150 fotogramas. Las imágenes utilizadas en estas secuencias presentan objetos en movimiento, en este caso un pato y las hojas de los árboles en el fondo de las imágenes, así como condiciones de luz naturales.

En la tabla 5.2 se muestra una secuencia con sensores en movimiento; el movimiento y disposición de los sensores es circular.

Tabla 5.2 Secuencia 2 y resultado de la fusión sin acumulación en el tiempo.

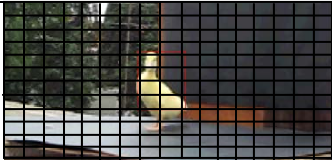
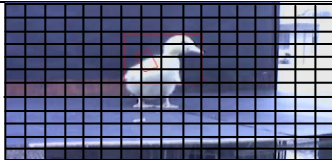
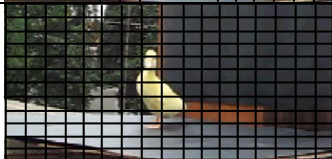

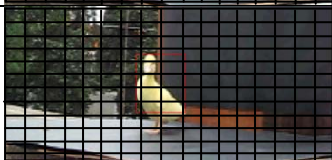



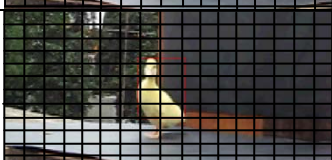
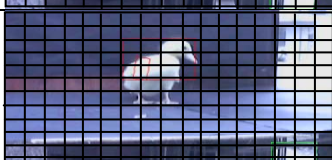
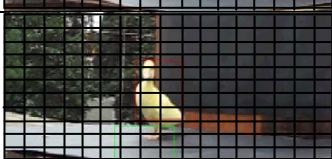

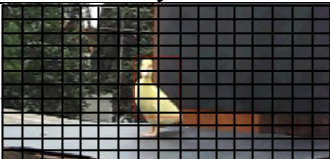
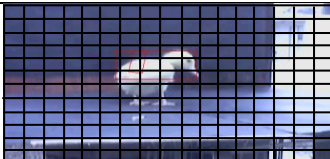
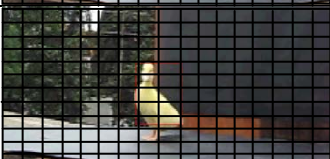
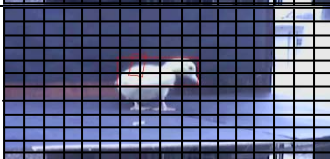
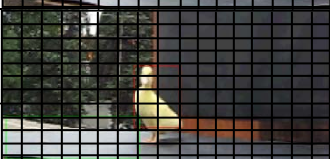

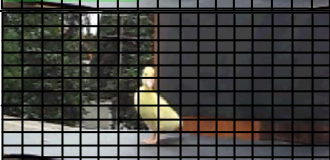
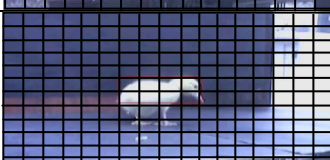
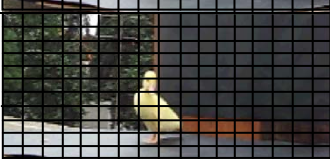



Número de cuadro	Cámara 1	Cámara 2	Marcos de discernimiento y resultado de la fusión (Cada elemento del marco corresponde a una rejilla)
1			$m1 = \{ 0.1727 \ 0.1727 \ 0.1727 \ 0.1727 \ 0.3090 \}$ $m2 = \{ 0.1727 \ 0.1727 \ 0.1727 \ 0.1727 \ 0.3090 \}$ $m = \{ 0.0298 \ 0.1366 \ 0.1366 \ 0.0298 \ 0.0955 \}$
2			$m1 = \{ \}$ $m2 = \{ 0.1799 \ 0.1704 \ 0.1704 \ 0.1704 \ 0.3090 \}$ $m = \{ 0.1799 \ 0.1704 \ 0.1704 \ 0.1704 \ 0.3090 \}$
3			$m1 = \{ 0.1727 \ 0.1727 \ 0.1727 \ 0.1727 \ 0.3090 \}$ $m2 = \{ 0.1799 \ 0.1704 \ 0.1704 \ 0.1704 \ 0.3090 \}$ $m = \{ 0.0311 \ 0.1355 \ 0.1355 \ 0.0294 \ 0.0955 \}$
4			$m1 = \{ 0.1727 \ 0.1727 \ 0.1727 \ 0.1727 \ 0.3090 \}$ $m2 = \{ 0.1773 \ 0.1682 \ 0.1773 \ 0.1682 \ 0.3090 \}$ $m = \{ 0.0715 \ 0.3138 \ 0.3240 \ 0.0678 \ 0.2229 \}$
5			$m1 = \{ 0.1727 \ 0.1727 \ 0.1727 \ 0.1727 \ 0.3090 \}$ $m2 = \{ 0.1773 \ 0.1682 \ 0.1773 \ 0.1682 \ 0.3090 \}$ $m = \{ 0.0715 \ 0.3138 \ 0.3240 \ 0.0678 \ 0.2229 \}$
6			$m1 = \{ 0.1727 \ 0.1727 \ 0.1727 \ 0.1727 \ 0.3090 \}$ $m2 = \{ 0.1773 \ 0.1682 \ 0.1773 \ 0.1682 \ 0.3090 \}$ $m = \{ 0.0715 \ 0.3138 \ 0.3240 \ 0.0678 \ 0.2229 \}$

Tabla 5.2 Secuencia 2 y resultado de la fusión sin acumulación en el tiempo.(Continuación)

7			$m1 = \{ 0.1930 \ 0.1660 \ 0.1660 \ 0.1660 \ 0.3090 \}$ $m2 = \{ 0.1767 \ 0.1767 \ 0.1767 \ 0.1608 \ 0.3090 \}$ $m = \{ 0.0799 \ 0.3169 \ 0.3169 \ 0.0626 \ 0.2238 \}$
8			$m1 = \{ 0.1930 \ 0.1660 \ 0.1660 \ 0.1660 \ 0.3090 \}$ $m2 = \{ 0.1767 \ 0.1767 \ 0.1767 \ 0.1608 \ 0.3090 \}$ $m = \{ 0.0799 \ 0.3169 \ 0.3169 \ 0.0626 \ 0.2238 \}$
9			$m1 = \{ 0.1859 \ 0.1596 \ 0.1859 \ 0.1596 \ 0.3090 \}$ $m2 = \{ 0.1814 \ 0.1641 \ 0.1814 \ 0.1641 \ 0.3090 \}$ $m = \{ 0.0786 \ 0.2943 \ 0.3433 \ 0.0611 \ 0.2227 \}$
10			$m1 = \{ \}$ $m2 = \{ 0.1814 \ 0.1641 \ 0.1814 \ 0.1641 \ 0.3090 \}$ $m = \{ 0.1814 \ 0.1641 \ 0.1814 \ 0.1641 \ 0.3090 \}$
11			$m1 = \{ \}$ $m2 = \{ \}$ $m = \{ \}$
12			$m1 = \{ \}$ $m2 = \{ 0.1814 \ 0.1641 \ 0.1814 \ 0.1641 \ 0.3090 \}$ $m = \{ 0.1814 \ 0.1641 \ 0.1814 \ 0.1641 \ 0.3090 \}$

En la tabla 5.2 tenemos una secuencia de imágenes con las cámaras en movimiento circular y la detección de los objetos por medio de segmentación por color, en algunos casos no logra detectar al objeto (pato) ya que la parte que se segmentó por el color fue tan pequeña que fue eliminada por el filtro de tamaño aplicado posteriormente. Sin embargo se puede observar que cuando ambos sensores aportan información, la fusión logra acumular las evidencias en un elemento, haciendo posible la decisión final, dando así más certeza de la observación de las cámaras. Un ejemplo de esto es el cuadro 9 donde $m1 = \{ 0.1859 \ 0.1596 \ 0.1859 \ 0.1596 \ 0.3090 \}$, $m2 = \{ 0.1814 \ 0.1641 \ 0.1814 \ 0.1641 \ 0.3090 \}$, $m = \{ 0.0786 \ 0.2943 \ 0.3433 \ 0.0611 \ 0.2227 \}$, en m se puede observar que el tercer elemento tiene una mayor evidencia haciendo que la decisión caiga en ese elemento.

En la columna 3 de la tabla 5.2 podemos ver los resultado de la fusión, donde m_1 son las creencias de la cámara 1, m_2 son las creencias de la cámara 2 y m es el resultado de la fusión. Los valores de m representan las rejillas por las que pasa el eje vertical, por lo tanto el valor mayor en el resultado indica que la acumulación de evidencias se hizo en ese punto, que normalmente en esta secuencia, corresponde al centro del objeto; con esto se puede determinar que el objeto fue encontrado en ambos sensores y con qué valor de certeza.

Cuando uno de los sensores no logra detectar al objeto, el valor final de la fusión es igual a los valores que proporciona el sensor que detectó al objeto. Podemos ver que en la penúltima fila, no se ha detectado ningún objeto, es por esto que no hay valores en la tercera fila, ninguna de las cámaras logró dar evidencia sobre el objeto. El problema en este cuadro fue que el filtro de tamaño eliminó las regiones que se habían detectado, la segmentación de los objetos no fue buena tampoco y por esta razón los sensores no dieron creencias.

5.3 PRUEBAS CON LAS CÁMARAS EN MOVIMIENTO PARA EVITAR LA OCLUSIÓN DEL OBJETO

En la tabla 5.3 se muestra una secuencia con los objetos ocultos; se intenta detectar el cubo de Rubik que se encuentra oculto en las imágenes 1 y 4 en la tabla, en estas imágenes los sensores no dan valores para crear los marcos de discernimiento ya que no encuentran al objeto, sin embargo, al mover los sensores se comienza a observar al objeto, lo cual muestra que es posible con el movimiento de los sensores, abordar el problema de oclusión parcial de los objetos.

En la tabla 5.3 se tienen los siguientes elementos: en la primera columna se tienen las imágenes obtenidas por la cámara 1, en la segunda columna se tienen las imágenes

obtenidas por la cámara 2, en la tercera columna se tiene la segmentación del cubo de Rubik a partir de las imágenes de la cámara 1, en la cuarta columna se tiene la segmentación del cubo de Rubik a partir de las imágenes de la cámara 2 y en la quinta columna se tienen las creencias que dio la cámara 1 las cuales son m_1 , m_2 son las creencias que dio la cámara 2 y m corresponde al resultado de la fusión.









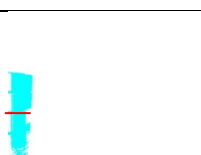
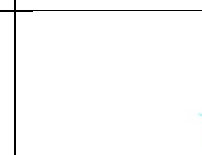


Cámara 1	Cámara 2	Detección en la cámara 1	Detección en la cámara 2	Marco de discernimiento y resultado de la fusión
				$m_1 = \{ \}$ $m_2 = \{ \}$ $m = \{ \}$
				$m_1 = \{0.3054, 0.5600, 0.1345, \}$ $m_2 = \{0.2028, 0.4100, 0.3871\}$ $m = \{ 0.3506, 0.5100, 0.1394 \}$
				$m_1 = \{0.6303, 0.3696\}$ $m_2 = \{0.5022, 0.4977\}$ $m = \{0.6819, 0.3181\}$
				$m_1 = \{ \}$ $m_2 = \{ \}$ $m = \{ \}$

Tabla 5.3 Secuencia 3 y resultado de la fusión sin acumulación en el tiempo, utilizando el eje horizontal para evaluar la parte visible del objeto.

En la secuencia presentada en la tabla 5.3 se muestra que el movimiento de las cámaras puede ayudar a observar objetos que están ocultos parcialmente. En la fila 2 se puede observar que no se detectó el cubo de Rubik, sin embargo con el movimiento, en la imagen siguiente se logra detectar el cubo, esto se puede ver en la fila 3 de la tabla donde también se reconoce un objeto aparte del cubo de rubik pero el cual es descartado por el

algoritmo de etiquetación, este es un fenómeno que se presentó en esta parte del experimento que no se ha estudiado a profundidad. Para cuantificar la parte visible del objeto se utilizaron en un principio los momentos estadísticos, sin embargo con ellos no se podía generar un marco de discernimiento por lo cual se prefirió utilizar el eje vertical del objeto para tener un marco de discernimiento donde colocar las evidencias.

En la tabla 5.3 se puede observar como el eje horizontal del objeto es encontrado (columnas 3 y 4) y cómo los sensores otorgan evidencias al marco de discernimiento (columna 5). Se puede observar que cuando las cámaras se mueven y se logra observar una mayor parte del objeto, tanto el eje horizontal como el marco de discernimiento crecen.

A diferencia del enfoque donde se utilizan cámaras estáticas, como en [3], el enfoque con cámaras móviles puede trabajar y dar una solución a la oclusión parcial de los objetos de forma casi automática. Como consecuencia, se puede decir que el enfoque utilizado en este trabajo es superior al enfoque que utiliza solo cámaras estáticas como [3].

5.4 ACUMULACIÓN DE EVIDENCIA A TRAVÉS DE VARIOS CUADROS

Hasta el momento se han mostrado los resultados de la fusión de sensores en cada cuadro, sin embargo, los sensores pueden acumular información a través del tiempo y esta información fusionada puede ser útil al tomar la decisión. Podemos tomar una de las imágenes como la imagen t y las subsecuentes como $t+1, t+2, \dots, t+n$. De esta forma, se acumuló información del objeto que estaba siendo detectado, los resultados se muestran en la tabla 5.4.

Para hacer los experimentos con las secuencias se tomaron en cuenta 3 cuadros consecutivos para hacer la acumulación de evidencias; posteriormente se hizo la fusión con estos datos y se inició la acumulación nuevamente.

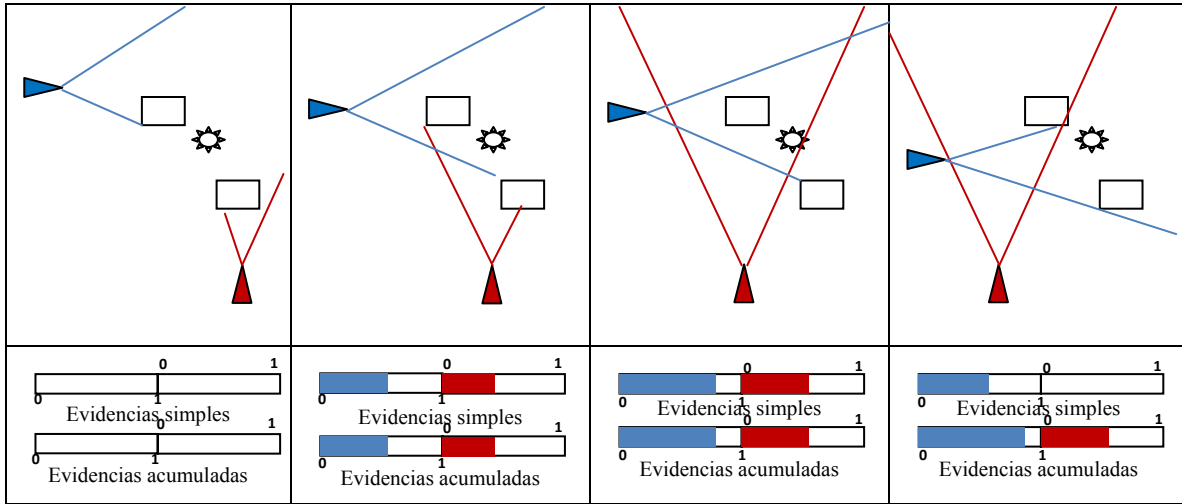


Figura 5.3 Ejemplo de la acumulación de evidencias. Los recuadros representan las evidencias que los sensores obtienen. Se muestra cómo se acumula la evidencia.

En la figura 5.3 se muestra un diagrama de cómo con el movimiento y la acumulación de la evidencia se tiene una mayor certidumbre sobre las observaciones. Este diagrama representa el experimento descrito en la sección 5.3. Podemos ver en la figura 5.3 un ejemplo de cómo se van acumulando las evidencias en cada sensor, en los recuadros superiores se tienen diagramas que representan las cámaras (triángulos) y sus campos de visión (líneas que salen de los triángulos), así como el objeto a detectar (estrella) y los objetos que lo ocultan (rectángulos); en los recuadros inferiores se muestran las evidencias, tanto las que otorgan los sensores en cada cuadro, como las que se acumulan en el tiempo, si el rectángulo está completamente lleno, significa que el sensor tiene completa certeza de que el objeto está ahí.

En los recuadros inferiores, los rectángulos de color azul representan las evidencias dadas por la cámara azul y los de color rojo las evidencias dadas por la cámara roja. En el primer recuadro inferior puede notarse que las cámaras no arrojan evidencia del objeto, en el segundo inferior las cámaras arrojan evidencia, se puede observar que las evidencias acumuladas son igual a las evidencias simples porque no había habido evidencia pasada del objeto. En el tercer recuadro inferior tenemos en las evidencias simples las evidencias de las cámaras en ese momento son representadas; sin embargo, en las evidencias acumuladas se puede ver que la evidencia de que el objeto está en la escena ha aumentado, ya que la

información del cuadro pasado se acumuló con la información del cuadro presente, en el cuarto recuadro inferior se puede ver que la cámara roja no ve más el objeto y la azul si lo ve, en éste caso la fusión de la evidencia simple arrojaría la información de la cámara roja pero, sin embargo con la evidencia acumulada se puede tener mayor certeza de que el objeto está ahí.

Para hacer la acumulación de evidencias se hace una fusión de información dentro de cada sensor en cada cuadro.

Ciertamente, puede ocurrir que la evidencia se acumule y en algún momento el objeto no esté más ahí. Sin embargo, se pueden tomar medidas contra eso, se puede acumular evidencia durante unos cuantos cuadros y después comenzar desde cero, esto es lo que se hace en este trabajo. Sin embargo, puede hacerse que cuando ninguno de las dos cámaras vea más al objeto se comience la acumulación desde cero; también puede utilizarse un híbrido con estas dos técnicas, es decir, si el objeto sigue apareciendo en los cuadros y se ha alcanzado el límite de evidencias a acumular, se comenzará desde cero, pero si el objeto no es visto por ambas cámaras entonces se comienza desde cero.

	t	$t + 1$	$t + 2$
Fusión en un solo punto	[0.9106 0.0894]	[0.9819 0.0181]	[0.9106 0.0894]
Fusión en el tiempo	[0.000 0.000]	[0.000 0.000]	[0.9982 0.0018]

Tabla 5.4 Resultados de la fusión de la primera terna de la tercera secuencia de imágenes.

En la tabla 5.4 se muestran las evidencias obtenidas por los sensores al seguir un objeto, en la primer fila se tienen as evidencias simples, sin acumular, y en la segunda fila se tienen los resultados de la fusión con acumulación, en la primera y segunda columna de la segunda fila no se tienen valores ya que se están acumulando las evidencias en cada sensor. Se puede observar que a medida que se incrementa la fusión en el tiempo, la evidencia se mejora y se acumula pues al moverse los sensores se descubren objetos que

desde una posición fija de los mismos, sería imposible determinar su ubicación (por ejemplo cuando el objeto se mueve paralelo a la línea de visión de un sensor mientras que la línea del otro sensor no puede observarlo). Puede pensarse que la fusión en un solo punto es mejor que la fusión acumulada pero no es así, ya que con la fusión acumulada, el resultado otorga más evidencia sobre un elemento del marco de discernimiento, de esta forma se tiene mayor certeza, ya que los valores se acumulan sobre un elemento haciendo más fácil la decisión.

	Tiempo de fusión por imagen
Fusión sensores fijos	0.03 segundos
Fusión en el tiempo con sensores móviles	0.12 segundos
Fusión Bayesiana	0.01 segundos

Tabla 5.5 Tiempos de ejecución

Los tiempos de ejecución del sistema, que fue implementado en Matlab en su totalidad, son mostrados en la tabla 5.5, considerando una rejilla de 20X20 sobre imágenes de 520X480 píxeles. De esta tabla se infiere que la fusión de sensores en movimiento es más costosa que la fusión con sensores fijos, evidentemente, debido a que la fusión de sensores en movimiento emplea varias imágenes para obtener un resultado final, a diferencia de la fusión sobre sensores fijos que solo procesan la información de dos imágenes (una por cada sensor), pero que tienen las limitaciones antes señaladas. Se muestra también que la fusión Bayesiana es 3 veces más rápida que la fusión con sensores fijos. Estos tiempos pueden ser fácilmente mejorados al programar los algoritmos de fusión en un lenguaje de bajo nivel como C o paralelizar el proceso de fusión.

5.5 COMPARACIÓN ENTRE FUSIÓN EVIDENCIAL Y FUSIÓN BAYESIANA

Con los mismos datos que se obtuvieron al hacer el seguimiento de los objetos se hicieron dos tipos de fusiones; fusión evidencial y fusión bayesiana, los resultados se presentan a continuación.

Es necesario mencionar que para utilizar la fusión bayesiana, es necesario tener probabilidades *a priori* para calcular la probabilidad final del evento, en este trabajo, se utilizó una distribución uniforme para las probabilidades *a priori*, estos datos se introducen en el módulo de fusión Bayesiana y el resultado es la probabilidad de que ese evento ocurra.

Tabla 5.6 Fragmento de la secuencia 1 y resultado de la fusión Bayesiana.

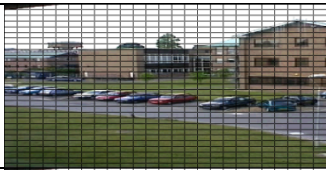

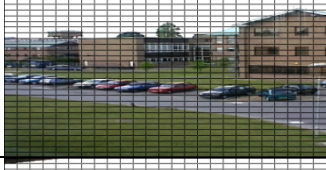

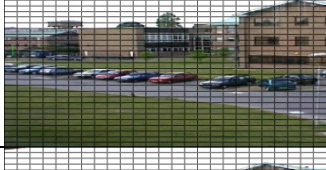

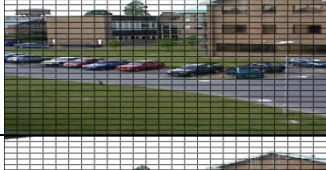

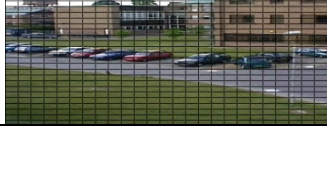
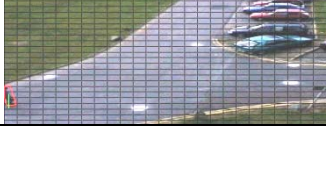










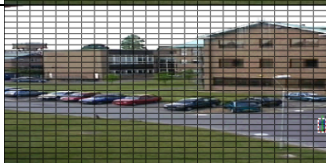

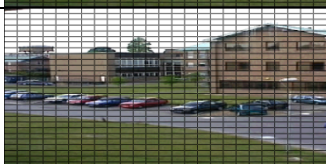

Número de Cuadro	Cámara 1	Cámara 2	Marcos de discernimiento y resultado de la fusión
1			$p1 = \{\}$ $p2 = \{\}$ $p = \{\}$
2			$p1 = \{\}$ $p2 = \{\}$ $p = \{\}$
3			$p1 = \{\}$ $p2 = \{\}$ $p = \{\}$
4			$p1 = \{\}$ $p2 = \{\}$ $p = \{\}$
5			$p1 = \{\}$ $p2 = \{ 0.1808 \quad 0.1647 \}$ $p = \{ 0.1808 \quad 0.1647 \}$

Tabla 5.6 Fragmento de la secuencia 1 y resultado de la fusión Bayesiana. (Continuación)

6			$p1 = \{ \}$ $p2 = \{ 0.2363 \quad 0.2363 \quad 0.2184 \quad 0.3090 \}$ $p = \{ 0.2363 \quad 0.2363 \quad 0.3090 \quad 0.2184 \}$
7			$p1 = \{ \}$ $p2 = \{ \}$ $p = \{ \}$
8			$p1 = \{ 0 \quad 0.3598 \quad 0.3312 \quad 0 \quad 0.3090 \}$ $p2 = \{ 0.1808 \quad 0.1647 \quad 0.1808 \quad 0.1647 \quad 0.3090 \}$ $p = \{ 0 \quad 0.2181 \quad 0.0955 \quad 0 \quad 0.2213 \}$
9			$p1 = \{ 0.2424 \quad 0.2243 \quad 0.2243 \quad 0.3090 \}$ $p2 = \{ 0.2363 \quad 0.2363 \quad 0.2184 \quad 0.3090 \}$ $p = \{ 0.1953 \quad 0.0573 \quad 0.1858 \quad 0.0955 \}$
10			$p1 = \{ 0.1830 \quad 0.1693 \quad 0.1693 \quad 0.1693 \quad 0.3090 \}$ $p2 = \{ 0.1727 \quad 0.1727 \quad 0.1727 \quad 0.1727 \quad 0.3090 \}$ $p = \{ 0.0316 \quad 0.1350 \quad 0.1350 \quad 0.0292 \quad 0.0955 \}$
11			$p1 = \{ \}$ $p2 = \{ 0.1759 \quad 0.1759 \quad 0.1759 \quad 0.1633 \quad 0.3090 \}$ $p = \{ 0.1759 \quad 0.1759 \quad 0.1633 \quad 0.3090 \quad 0.1759 \}$
12			$p1 = \{ \}$ $p2 = \{ 0.1727 \quad 0.1727 \quad 0.1727 \quad 0.1727 \quad 0.3090 \}$ $p = \{ 0.1727 \quad 0.1727 \quad 0.1727 \quad 0.3090 \quad 0.1727 \}$

En la tabla 5.6 se muestran los resultados de la fusión Bayesiana, en la primer columna se muestran los cuadros que ve la cámara 1, en la segunda columna se muestran los cuadros que ve la cámara 2 y en la tercera columna se muestran las probabilidades resultantes de la fusión Bayesiana, en esta columna se tienen 3 datos: $m1$ que representan las probabilidades obtenidas por el sensor 1, $m2$ que representan las probabilidades obtenidas por el sensor 2 y m es el resultado de la fusión Bayesiana.

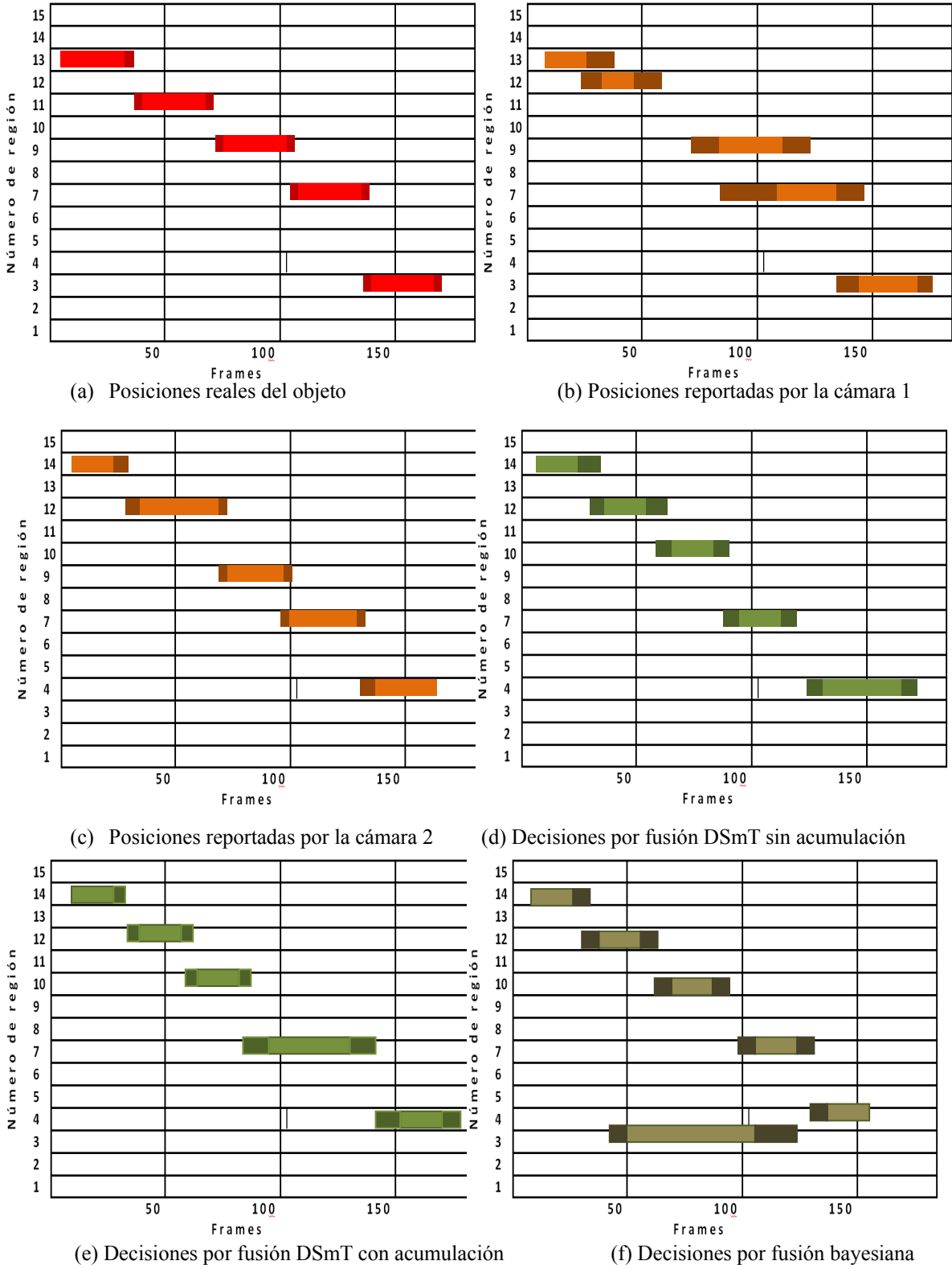


Figura 5.4 Comparación entre las evidencias mostradas por las cámaras y las decisiones tomadas con las técnicas de fusión utilizadas en este trabajo.

En la figura 5.4 los rectángulos representan al objeto, la parte más clara muestra que el objeto ha estado en la misma región durante varios cuadros y la parte más oscura muestra que ha habido un cambio de región. Para tomar la decisión sobre en que región se encuentra el objeto seguido se toma el valor máximo del resultado de la fusión en cualquier modalidad, ya sea DSmT simple, DSmT Acumulada o Bayesiana.

Comparando la figura 5.4 (d) con la figura 5.4 (f) se puede ver que en la fusión bayesiana se tienen errores al decidir en qué región de la imagen se encuentra el objeto, podemos ver que a partir del cuadro 48 y hasta el cuadro 130 la fusión bayesiana dice que el objeto se encuentra en dos lugares distintos al mismo tiempo, esto es debido que las probabilidades finales son iguales en esos puntos y al hacer la fusión se quedan de la misma forma, el que se tengan valores iguales en distintos elementos genera ruido cuando se intenta decidir en qué región se encuentra el objeto exactamente, por otro lado, la fusión DSmT sin acumulación se equivoca menos. Podemos ver también que la fusión DSmT con acumulación tiene mejores resultados que la fusión sin acumulación, esto lo podemos observar en la figura 5.2 (e) y 5.2 (d) si observamos los dos primeros rectángulos (de izquierda a derecha) en (d) se tiene un traslape, es decir, la fusión dice que el objeto está en dos lugares al mismo tiempo, el traslape ocurre por una mayor cantidad de cuadros en (d) que en (e).

Se ha dicho que con el movimiento de las cámaras es posible lidiar con el problema de la oclusión parcial de los objetos, en la tabla 5.3 se muestra un experimento hecho con un objeto ocluido y como el movimiento hace que el objeto se descubra poco a poco. En la figura 5.4 se muestra un gráfico en el cual se puede observar que con el movimiento de las cámaras se obtiene una mayor evidencia sobre el objeto que está ocluido.

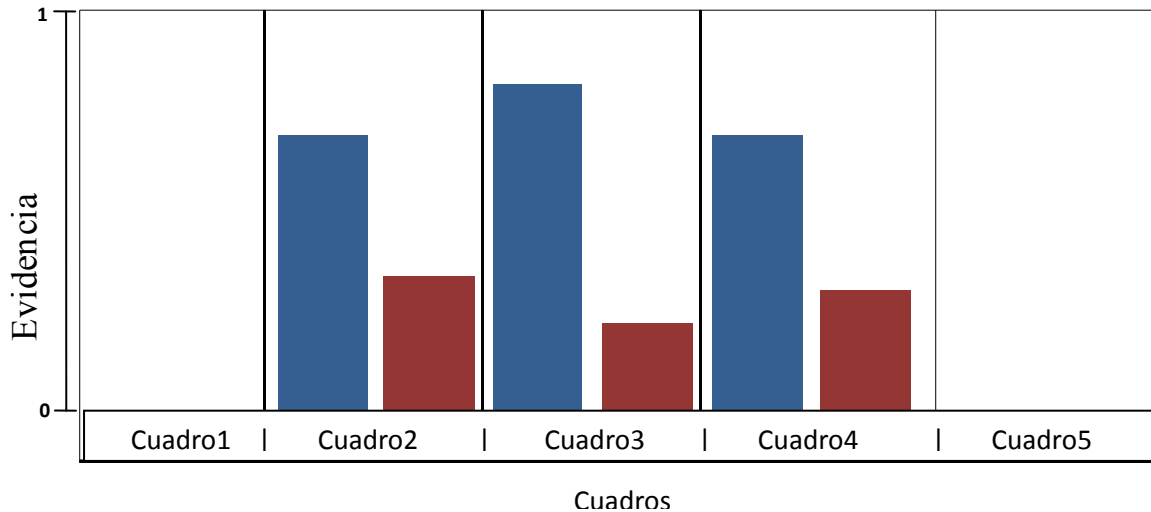


Figura 5.5 Incremento de la evidencia con el movimiento de las cámaras

En la Figura 5.4 las barras representan la evidencia que las cámaras arrojan acerca de una imagen que están observando, las barras azules representan la evidencia del sensor azul y las barras rojas representan las evidencias del sensor rojo, en la primer columna podemos ver que las cámaras no dan evidencia ya que no hay objetos en la escena, en la segunda columna las cámaras logran ver una parte del objeto, en la tercer columna las cámaras ven una mayor parte del objeto; por lo cual las evidencias crecen, en la cuarta columna las cámaras observan una menor porción del objeto y las evidencias decrecen, por último en la quinta columna el objeto desaparece del campo de visión de las cámaras. Las cámaras en este ejemplo se mueven y el objeto está ocluido, dadas estas circunstancias podemos decir que es posible trabajar con la oclusión de los objetos si se utilizan cámaras en movimiento, ya que el movimiento de las cámaras logra hacer visible al objeto ocluido.

5.6 MÉTRICAS DE EVALUACIÓN

Para lograr una evaluación cuantitativa se utilizaron métricas que indican la exactitud de los métodos propuestos. Se utilizaron dos métricas una llamada Taza de

Detección del Seguidor (Tracker Detection Rate - TRDR) y la otra Taza de Falsa Alarma (False Alarm Rate - FAR) [27], éstas se calculan mediante las siguientes fórmulas:

$$TRDR = \frac{TP}{TG} \quad 5.1$$

$$FAR = \frac{FP}{TP + FP} \quad 5.2$$

Donde TP es el número de verdaderos positivos; un verdadero positivo es aquel evento que el sensor dice que ocurre y en realidad ocurre el evento, TG es el número total de regiones por cada imagen donde se encuentran objetos en movimiento de acuerdo a los datos reales FP es el número de falsos positivos; un falso positivo ocurre cuando el sensor indica que existe un evento y en realidad no está ocurriendo. Para este trabajo un evento es que un objeto esté en movimiento o aparezca en la imagen. Con estas métricas tenemos los resultados que se muestran a continuación.

Fuente	TRDR	FAR
Cámara 1	98.5%	60%
Cámara 2	97%	59%
DSmT sin acumulación	93.3%	6%
DSmT con acumulación	96%	6%
Fusión Bayesiana	92.3%	6.5%

Tabla 5.7 Comparación entre los distintos métodos utilizados en este trabajo

Con la tabla 5.7 se puede observar que al utilizar fusión DSmT se obtienen mejores resultados que al utilizar la Fusión Bayesiana, la mejora es de 1% en TRDR y de 0.5% en

FAR. Es necesario mencionar que la fusión DSMT mejora a la fusión Bayesiana debido a los cálculos que realiza y a la redistribución de las creencias no asignadas, sin embargo, la fusión Bayesiana, al ser menos compleja en sus operaciones obtiene los resultados más rápido.

Capítulo 6

Resumen

En este capítulo se describen las conclusiones del trabajo así como el trabajo futuro que se podría realizar.

Conclusiones y trabajo futuro

Se presentó una mejora para realizar la fusión de sensores cuando dichos sensores están en movimiento relativo a la escena que sensan. Los resultados obtenidos señalan que la propuesta mejora la evidencia obtenida respecto a sensores fijos, pues la acumulación de evidencia permite que el sistema almacene y emita mejores evidencias al considerar no solamente el estado actual del escenario, sino además basa sus decisiones en términos de los escenarios previos. Esto es de especial interés cuando los objetos en movimiento se encuentran parcialmente ocultos respecto a una posición fija de los sensores y por tanto la evidencia generada por la fusión es nula o casi nula. Al permitir el movimiento de los sensores se logra evitar la oclusión de manera total o parcial, aun cuando el objetivo pueda permanecer inmóvil respecto al plano de sensado de los sensores en una posición fija.

Los resultados experimentales fueron obtenidos sobre escenas reales, aun cuando se requiere más trabajo en el hardware para construir el sistema real hardware/software que pueda ser empleado en situaciones de vigilancia de tipo comercial o industrial; sin embargo

los principios propuestos aquí son válidos aún sin tener el hardware apropiado pues los resultados de la fusión fueron obtenidos con un prototipo que simula apropiadamente el hardware requerido.

Los resultados de los experimentos muestran que el uso de la fusión por medio de DSMT tiene una mayor precisión al identificar la posición del objeto que la fusión bayesiana, comparando DSMT sin acumulación contra la fusión bayesiana se tiene una mejora del 1%, esto es debido a que la fusión bayesiana no utiliza la información que la DSMT si utiliza, como la incertidumbre de los sensores o la evidencia no asignada; la fusión DSMT toma en cuenta más elementos del conjunto hiperpotencia que la probabilidad bayesiana, de hecho, la fusión bayesiana no trabaja con el conjunto hiperpotencia, ésta trabaja solamente con el conjunto potencia cuya cardinalidad es menor al conjunto hiperpotencia, por ejemplo, toma en cuenta las intersecciones entre los elementos del marco de discernimiento, como $(\theta_1 \cap \theta_2)$ [6], debido a esto, la DSMT puede tratar información incompleta e inexacta, lo que permite rescatar información que con la fusión bayesiana estaría perdida o no se tomaría en cuenta.

6.1 APORTACIONES

Se creó un sistema capaz de fusionar la información de sensores en movimiento con el fin de seguir objetos en movimiento, las características principales de este sistema son las siguientes:

- Utiliza sensores en movimiento.
- Tiene la capacidad de seguir objetos en movimiento.
- Realiza una acumulación de evidencia a través de los cuadros, esto es importante ya que en el trabajo anterior a esta tesis la información no se acumulaba. La fusión de la información acumulada da mejores resultados que la fusión de información en un solo cuadro.
- Utiliza fusión basada en DSMT

6.2 TRABAJO FUTURO

El trabajo presentado en esta tesis es susceptible de mejoras que repercutirán en un sistema más robusto y con una aplicabilidad mayor. Algunas de estas mejoras incluyen:

1. La utilización de más de dos sensores en el sistema: esto ayudaría a tener más evidencia sobre la escena y de ésta manera se tendrían mejores resultados. Los demás sensores que se utilicen en el hardware pueden ser colocados en distintas posiciones, no es necesario que estén en un mismo plano, los sensores pueden colocarse con diferentes planos de sensado, con lo cual se puede observar la escena desde distintos ángulos de visión, logrando así que se acumule mayor evidencia al permitir más grados de libertad al sistema de vigilancia.
2. Disminuir o incrementar el tamaño de la rejilla: si se incrementa el tamaño de la rejilla, es decir, si los cuadros que conforman la rejilla aumentan de tamaño es posible hacer el proceso de fusión más rápido, sin embargo se tiene menos exactitud en la fusión. Si se disminuye el tamaño de la rejilla, es decir, si los cuadros que la conforman disminuyen de tamaño es posible hacer que la fusión sea más exacta pero esto haría más costoso el proceso de fusión. Por ello se propone también:
3. Paralelizar el proceso de fusión, en especial si se desea obtener un sistema de tiempo real que emita decisiones más precisas que aquellas obtenidas con dos sensores fijos y con un tamaño de rejilla muy pequeña.

ANEXO A
PUBLICACIONES OBTENIDAS

1. “Fusion of Multiple Mobile Cameras for Object Tracking”, Jorge Eduardo Xalteno, Leopoldo Altamirano, Carlos Alberto Reyes, Luis Carlos Altamirano, Special Issue in Advances in Pattern Recognition Research in Computing Science 44, 2009, pp 37-44.

Seleccionado como uno de los mejores artículos del Workshop ... e invitado a ser publicado en versión extendida en la revista Computación y Sistemas.

2. “Fusion of Multiple Mobile Sensors for Partially Ocluded Objects”, Jorge Eduardo Xalteno, Luis Carlos Altamirano, Carlos Alberto Reyes, Leopoldo Altamirano, Enviado al journal Computación y Sistemas. México D.F. Diciembre 2009.
3. “Partially-occluded Object Recognition through Appereance Based Models and Sensor Fusion”, Jorge Eduardo Xalteno, Luis Altamirano, Carlos Alberto Reyes, Lepoldo Altamirano, En proceso. Para su envio a Patter Recognition Letters, The Nederlands. 2010.

Referencias

- [1] “Principal Axis-Based Correspondence between Multiple Cameras for People Tracking”, Weiming Hu, Min Hu, et. Al. IEEE Transactions on pattern analysis and machine intelligence April 2006, páginas 663-671.
- [2] “Consistent labeling of tracked objects in multiple cameras with overlapping fields of view”. Sohaib Khan and Mubarak Shah, IEEE Transactions on pattern analysis and machine intelligence, October 2003, páginas 1355-1360
- [3] Esteban García (2008), “Fusión multicámara para seguimiento de objetos basada en Teoría evidencial”. Tesis de maestría INAOE.
- [4] “Decision Level Multiple Cameras Fusion Using Dezert-Smarandache theory”. Esteban Garcia and Leopoldo Altamirano. Lecture Notes in Computing Science Vol.4673. ISBN: 978-3-540-74271-5, ISSN 0302-9743. pp. 117-124. Springer-Verlag. Agosto 2007.
- [5] “Fixed filtering and intelligent control for tracking with mobile sensors”. Dominique Luzeaux, Proceedings of the 29th Southeastern Symposium of System Theory, IEEE. 2007
- [6] “An introduction to the DS_m Theory for the combination of paradoxical, uncertain and imprecise sources of information”, Floretin Smarandache, Jean Dezert, Online paper revised on August 1st 2006,
- [7] Hanbiao Wang, Kung Yao, and Deborah Estrin, "Information-Theoretic Approaches for Sensor Selection and Placement in Sensor Networks for Target Localization and Tracking" (January 1, 2005). Center for Embedded Network Sensing. Technical Reports. Paper 18. <http://repositories.cdlib.org/cens/techrep/18>
- [8] “Improving the selection of feature points for tracking”, Zoran Zivkovic, Ferdinand van der Heijden, In pattern analysis and application vol 7 no. 2 2004 Springer-Verlag.
- [9] <ftp://pets2001.cs.rdg.ac.uk/>
- [10] “Homography Based Parallel Volume Intersection: Toward Real-Time Volume Reconstruction Using Active Cameras” Toshikazu Wada, Xiaojun Wu, Shogo Tokai, et al.

Proceedings of the Fifth IEEE International Workshop on Computer Architectures for Machine Perception, (CAMP '00)

[11] "Look there!" Predicting where to look for motion in active camera network" Ugur Murat Erdem, Stan Sclaroff. 2005 IEEE

[12] "Coupled Object Detection and Tracking from Static Cameras and Moving Vehicles" Bastian Leibe, Konrad Schindler, et al. IEEE Transactions on Pattern Analysis and Machine Intelligence vol. 30 no. 10 October 2008.

[13] "Intrusion Detection and Tracking with Pan-Tilt Cameras". Aridiam Biswas, Pritwijit Guha et al. Second European conference on Computer Vision 1995.

[14] "Approaches to Multisensor Data Fusion in Target Tracking: A Survey", Duncan Smith and Sameer Singh, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING VOL.18 NO. 12 DECEMBER 2006

[15] "A New Relaxation Algorithm and Passive Sensor Data Association," K.R. Pattipatti, S. Deb, Y. Bar-Shalom, and R.B. Washburn, IEEE Trans. Automatic Control, vol. 37, pp. 198-213, Feb. 1992.

[16] "A Neural Network for Data Association," M. Winter and G. Favier, Proc. 1999 IEEE Int'l Conf. Acoustics, Speech, and Signal Processing, vol. 2, pp. 1041-1044, Mar. 1999.

[17] "Multi-Target Tracking Based on Multi-Sensor Information Fusion with Fuzzy Inference," H. Hong, H. Chong-zhao, Z. Hong-Yan, and W. Rong, Control and Decision, vol. 19, no. 3, pp. 272-276, 2004.

[18] "Mathematical Techniques in Multisensor Data Fusion". D. Hall, Artech House, 1992.

[19] "Multi-Sensor Integration and Decision Level Fusion," Kittler, Proc. DERA/IEE Workshop Intelligent Sensor Processing (Ref. No. 2001/050), pp. 6/1-6/6, 2001.

[20] "Multifocus image fusion using the nonsubsampling contourlet transform" [Zhang, Q.](#), [Guo, B.-l.](#) 2009 *Signal Processing* 89 (7), pp. 1334-1346

[21] "Error compensation of GPS using sensor fusion in intelligent vehicle", Chung, B.-M., Yeo, I.-J., Seok, J.-W., Cho, C.-S. 2009 ICARA 2009 - Proceedings of the 4th International Conference on Autonomous Robots and Agents, art. no. 4803944, pp. 278-283

[22] "Pedestrian crossing detection based on evidential fusion of video-sensors", [Hide Abstract Boudet, L.](#), [Midenet, S.](#) 2009 *Transportation Research Part C: Emerging Technologies*, Article in Press

- [23] “Design of multisensor fusion-based tool condition monitoring system in end milling”, [Hide Abstract](#) [Cho, S.](#), [Binsaeid, S.](#), [Asfour, S.](#) 2009 *International Journal of Advanced Manufacturing Technology*, pp. 1-14
- [24] “Machine Vision” Ramesh Jain, Rangachar, rian G. Schunck, McGraw-Hill International editions, Computer Sciencies Series, 1995
- [25] “DIGITAL IMAGE PROCESSING using MATLAB”, Rafael C. Gonzalez, Richard E. Woods, Steven L. Eddins, Pearson Education, 2004
- [26] <http://iie.fing.edu.uy/investigacion/grupos/gti/cursos/egvc/material/tema-7.pdf>
- [27] “Performance Evaluation of Object Detection and Tracking Systems”, Bashir, F. y Porikli, F. (2006). En IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS).
- [28] “Probabilistic Image Sensor Fusion”, Rave K. Sharma, Todd K. Leen, Misha Pavel, Advances in Neural Information Processing Systems 11, The MIT Press, 1999.
- [29] “Sensor Fusion for Video Surveillance” Lauro Snidaro, Gian Luca Foresti, Ruixin Niu, Pramod K. Varshney, Sensor Fusion for Video Surveillance, 7th Conference on Information Fusion 2004 pp. 739-746.
- [30] “Principle of Artificial Inteligence” N. J. Nilson, Palo Alto California. Tioga 1980.
- [30] “Probabilistic Fusion of Gait Features for Biometric Verification”, Bazin, A. I., Middleton, L. and Nixon, M. S. (2005) Proceedings of Eighth International Conference of Information Fusion, pp. 124-131, Philadelphia, PA, USA
- [32] “Multiple camera tracking of interacting and occluded human motion,” S. L. Dockstader and A. M. Tekalp, Proc. IEEE, vol. 89, pp. 1441–1455, Oct. 2001.
- [33] “Multitarget-Multisensor Tracking”, Yaakov Bar-Shalom, Xiao-Rong Li, Aerospace and Electronic Systems Magazine, IEEE, Feb 1996, Volumen 11, Página 41.
- [34] “Dempster–Shafer theory for sensor fusion in autonomous mobile robots”, P.R. Murphy, IEEE Transactions on Robotics and Automation 14(2) (1999) 197–206.
- [35] “Sensor Fusion in Certainty Grids for Mobile Robots”, HP Moravec - AI magazine, 1988 - aaai.org

- [36] “Multi-camera people tracking using evidential filters”, Rafael Muñoz Salinas, R. Medina-Carnicer, F. J. Madrid-Cuevas, International Journal of Approximate Reasoning, 2009, pages 732-749
- [37] “A Bayesian plan-view map based approach for multiple-person detection and Pattern Recognition Volume 41, Issue 12 (December 2008) Pages: 3665-3676 2008
- [38] “Study of the different methods for combining evidence”, Cheng Y., Kashyap R.L., SPIE Vol. 635 Applications of Artificial Intelligence III, pp. 384-393, 1986.
- [39] Foundations for a new theory of plausible and paradoxical reasoning. Dezert, J. 2002). Information and Security Journal, 9.
- [40] Dubois D., Prade H., A Set-Theoretic View of Belief Functions, International Journal of General Systems, pp.193-226, Vol.12, 1986.
- [41] Smets Ph., The combination of evidence in the Transferable Belief Model, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 12, No. 5, pp. 447-458, 1990.
- [42] Yager R.R., Hedging in the combination of evidence, Journal of Information & Optimization Sciences, Analytic Publishing Co., Vol. 4, No. 1, pp. 73-81, 1983.