

Seguimiento 3D para Múltiples Personas en Entornos Multicámara Basado en Filtros de Partículas

Adolfo López Méndez
Mayo 2007
ETSETB

Resumen

El rápido desarrollo de las tecnologías y el aumento de la capacidad de procesado ha potenciado el estudio de tecnologías que tratan de comunicar de un modo más humano las máquinas con las personas. En este aspecto, el seguimiento visual de múltiples individuos es un problema clave en la interpretación automatizada del lenguaje verbal y no verbal. Cualquier posible interpretación pasa por la estimación de la posición de las personas.

Existen muchos trabajos que tratan el seguimiento de objetos y personas con numerosos enfoques y resultados notables, sin embargo la mayoría parten de un enfoque con una sola vista bidimensional. Sin embargo, el presente proyecto parte de la información básica con la que se han desarrollado algunos de los trabajos en el ámbito de la interpretación del lenguaje humano en entornos cerrados en la Universitat Politècnica de Catalunya (UPC): una reconstrucción 3D de la escena. Datos de tales características admiten varios niveles de complejidad, pero en este trabajo se emplea un volumen binario, una forma tridimensional que representa las personas y objetos activos en la sala.

Con respecto a las herramientas con las que se abordan los problemas de seguimiento, existen dos enfoques básicos. El primero se basa en efectuar una descomposición de la escena y efectuar asignaciones con los candidatos del siguiente fotograma empleando criterios de semejanza. El segundo enfoque pretende estimar la nueva posición de cada objetivo con la información a priori, las observaciones disponibles y un modelo dinámico del proceso. Este trabajo ha seguido una estrategia perteneciente al segundo grupo y de eficacia demostrada ya no sólo en el ámbito del seguimiento visual: el filtro de partículas. Para hacernos una idea, esta estrategia consiste en una población de partículas o muestras que, iteración tras iteración, se propagan según una dinámica determinada y sobreviven en las zonas donde hay mayor verosimilitud de ubicarse el objetivo al que se pretende seguir.

El diseño propuesto combina los elementos citados y añade otros modelos con los que finalmente se ha elaborado un algoritmo de sencilla implementación pero que obtiene resultados notables en escenarios de sala inteligente. Como contribuciones significativas de este proyecto podemos destacar la adaptación de los filtros de partículas a datos basados en voxels y una propuesta de bloqueo para tratar las interacciones entre múltiples personas. Es presente trabajo explica de modo sencillo los elementos principales que constituyen el algoritmo así como la teoría que lo soporta y da muestras objetivas de los resultados logrados. El trabajo desarrollado es la base de un artículo de conferencia [13] publicado a finales del pasado año 2006 y presentado en el ICASSP '07.

Adolfo López Méndez
Mayo 2007

Agradecimientos

A Nuria, por el cariño y apoyo brindados y porque es, en definitiva, mi mayor inspiración y motivación para elaborar un trabajo digno de escribir su nombre en él.

A mis padres, mi hermano y toda mi familia.

A todos mis amigos (siento no citarlos a todos, pero es que sois unos cuantos), en especial a los que siempre pensaron que podía hacer un trabajo como éste.

A Cristian Cantón, porque que he tenido la suerte de tenerle como director de proyecto y por brindarme la oportunidad de incluir mi trabajo en un artículo de conferencia. En este aspecto agradecer también a Josep Ramón Casas su trabajo e interés, al igual que a todos aquellos que hayan hecho posible el poder realizar un proyecto tan interesante como éste.

A mis compañeros de la sala Phoenix de Vodafone cuya amistad ha sido un capítulo muy importante en esta etapa de mi vida: Joan, Dimas, Pepillo, Jorge, Pol –¡menudos asados!-, Pedro, Masalías, Cristina, María, Carlos, Peter, Hugo, Jessica, Jordi ...

A mis compañeros en estos años de universidad: Albert M., Ángel, Albert. P., Toni, Albert. A., Rengel, Jordi, Guiri, Dani Rojo, Dani Hernández (gracias por las imágenes), Juanjo, Pau, Joaquín ...

A todos los que habéis contribuido a este camino, gracias.

Adolfo López Méndez

Índice

1 Introducción.....	8
1.1 Escenario	8
1.2 Estado del arte	9
1.3 Organización.....	11
2 Seguimiento y Filtro de Partículas	12
2.1 Seguimiento Bayesiano No Lineal	12
2.2 El filtro de Kalman	14
2.3 El filtro de partículas	15
2.3.1 SIS PF	16
2.3.2 SIR PF	18
2.3.3 Filtro de Partículas aplicado a múltiples objetivos.....	19
3 Seguimiento Multi-Persona en Salas Inteligentes.....	21
3.1 Un entorno Multicámara: La sala CHIL.....	21
3.2 Datos en el Escenario	24
3.2.1 Volúmenes	24
3.3 Aplicación al seguimiento Multipersona 3D	26
3.3.1 Clasificación	26
3.3.2 Filtrado del escenario.....	27
3.3.3 El filtro de partículas en el entorno CHIL	29
3.3.3.1 Propagación de las partículas	30
3.3.3.2 Cálculo de pesos	31
3.3.3.3 Seguimiento de múltiples personas y bloqueo de partículas.....	35
3.3.3.4 Remuestreo	37
4 Resultados y evaluación	40
4.1 Métricas	40
4.2 Algoritmo de seguimiento basado en Kalman.....	43
4.3 Algoritmo basado en filtro de partículas	45
4.4 Comparativa de remuestreo y asignación de partículas a voxels	51
5. Conclusiones.....	54
5.1 Cumplimiento de objetivos.....	54
5.2 Trabajo Futuro	55
Anexo – Estudio del espacio de color en el seguimiento basado en filtro de partículas	56
A.1 Obtención de la Información de color	56
A.2 Filtro de partículas con información de color.....	58
A.3 Dificultades.....	59
Bibliografía.....	62

Lista de ilustraciones

Fig. 1.1 Distribución de probabilidad multimodal	9
Fig. 2.1 Propuesta de distribución y cálculo de pesos	16
Fig. 2.2 Las partículas con más peso, al constituir puntos de mayor probabilidad, son remuestreadas con mayor cardinalidad, mientras que las que tienen poco peso se consideran negligibles y “mueren”	17
Fig. 2.3 Efecto de concentración de las partículas (representadas en rojo). En azul se han pintado los voxels. Se muestran los fotogramas 401, 421, 441 y 461. Al cabo de 60 fotogramas todas las partículas se concentran en unos puntos muy concretos.	18
Fig. 3.1 Diagrama esquemático de la Sala CHIL	21
Fig. 3.2 Imagen obtenida desde la cámara cenital	22
Fig. 3.3 Origen de coordenadas de la sala CHIL.....	22
Fig. 3.4 Modelo puntual de cámara.....	23
Fig. 3.5 Proyección en el plano focal	23
Fig. 3.6 Fotograma con su máscara correspondiente.....	25
Fig. 3.7 En azul se muestran los volúmenes reconstruidos. Las elipsoides señalan las zonas donde trabaja cada filtro de partículas.....	27
Fig. 3.8 Filtrado en la zona de puertas. Los conjuntos de puntos en azul son los voxels de “foreground”. Las imágenes de la izquierda no emplean filtrado mientras que las de la derecha sí lo utilizan. El ruido añadido por el movimiento de las puertas puede dar lugar a detecciones confusas.....	28
Fig. 3.9 La secuencia de arriba no emplea apertura morfológica mientras que la de abajo sí.	29
Fig. 3.10. Diagrama de bloques del filtro de partículas.....	30
Fig. 3.11 Tras el muestreo, varias partículas (puntos rojos) pueden ocupar un mismo voxel. El primer paso del filtro consiste en redibujar las partículas como se muestra.....	31
Fig. 3.12 Imágenes correspondientes a 80 iteraciones del algoritmo. En la imagen superior se emplea un criterio de media de voxels, mientras que en la inferior se utiliza la estrategia de asimilación de la superficie	33
Fig. 3.13 Análisis de cuatro entornos de partícula para el cálculo de su peso. En azul se marcan los voxels de “foreground”.....	34
Fig. 3.14 Zonas sobre las que se remuestrean las partículas de los filtros de partículas....	35
Fig. 3.15 Vista cenital de dos ejecuciones con 600 partículas y arista de voxel de 2 cm ..	36
Fig. 3.16 Gráficos de distancia de remuestreo en función del número de partículas y el tamaño de voxel. a)2cm. b) 3cm c) 5cm.....	38
Fig. 3.17 Distancia de remuestreo en cm para los distintos ejes. Se toman como medidas de sala 400x500x210 cm.	39
Fig. 4.1 Casuísticas de seguimiento y evaluación de las mismas. a) La hipótesis generada se aleja del objeto al que sigue. En el fotograma k+1 supone una penalización en el MOTP, mientras que en el k+2 se produce una pérdida asociada al objeto y una falsa detección asociada a la hipótesis, penalizando el MOTA.. b)Intercambio entre hipótesis y objetivos. La línea discontinua muestra la trayectoria definida por cada una de las hipótesis (círculos rojo y naranja). Nótese que al hacer las asignaciones objetivo-hipótesis según un criterio de mínima distancia, sólo se producen dos intercambios (uno por cada hipótesis) mientras que si las correspondencias se efectuasen teniendo en cuenta los pares que se han dado más veces contaría 5 intercambios (3 asociadas a la hipótesis roja y 2 a la naranja).	43
Fig. 4.2 Si contabilizamos pérdidas fotograma a fotograma obtenemos 75% en los tres primeros y en en k+5, 100% en k+3 y k+4 y 0% en los tres últimos. La tasa de error	

cometido por pérdidas para la secuencia, según este cómputo, es 56%. Si hacemos el cálculo como se propone, tenemos 20 pérdidas sobre 27 objetivos presentes, lo que da una tasa de error por pérdidas de 74,1%. Una tasa de error de 56% no refleja el mal comportamiento que ofrece visualmente el seguimiento ilustrado.	43
Fig. 4.3 MOTP medio obtenido en todos los seminarios procesados para distinto número de partículas y tamaño de voxel. En magenta se muestran los resultados obtenidos mediante el algoritmo basado en Kalman.....	46
Fig. 4.4 MOTA medio obtenido en todos los seminarios procesados para distinto número de partículas y tamaño de voxel. En magenta se muestran los resultados obtenidos mediante el algoritmo basado en Kalman.....	46
Fig. 4.4. Ejecución con 600 partículas y tamaño de voxel de 2cm. El algoritmo resuelve con éxito el cruce de dos asistentes al seminario del 06-07-2005.	48
Fig. 4.7 Ejemplos de falsa detección en una ejecución con 1000 partículas y voxels de 2 cm sobre el seminario del 20-07-2005. La chaqueta colgada por uno de los asistentes es detectada en un instante concreto como un objetivo. En la parte inferior derecha, un objeto depositado constituye un conjunto conexo de voxels que atrae al filtro de partículas amarillo. Este último caso está causado además por la falta de un volumen conexo en la posición en que se sienta la persona objetivo.....	50
Fig. 4.8 Ejecución con 1000 partículas y voxels de 2 cm sobre el seminario del 20-07-2005. Los asistentes se concentran alrededor de la mesa ubicada en la parte superior izquierda. A pesar de las falsas detecciones, el algoritmo mantiene una consistencia notable con las hipótesis acerca de la posición de cada persona.	50
Fig. 4.9. Fotogramas 800, 900, 1000 y 1100 de una secuencia de IBM no evaluada con las métricas CLEAR. Los únicos intercambios que se producen son debidos a que los asistentes salen de la zona en que es posible reconstruir volúmenes. Si se garantiza que la reconstrucción es suficientemente sólida y guarda una cierta correlación con la forma a la que representa, puede plantearse un filtrado de los voxels para una altura inferior a un cierto umbral, eliminando así gran parte de los volúmenes de las sillas.	51
Fig. A.1. Ambigüedad espacial. Dos posibles asignaciones de color en la escena.	56
Fig. A.2. Ambigüedad cromática. El voxel superior tiene asignaciones distintas en las dos escenas debido a las occlusiones del mismo. Podría considerarse que el voxel no es fotoconsistente.	56
Fig. A.3 Rayo de Bresenham 2D.....	57
Fig. A.4. Fotograma de seminario y su correspondiente coloreado de voxel.....	59
Fig. A.5. Métrica de Bhattacharyya entre un histograma en el instante k y k-1. a) Calculado para los filtros 1 y 2 (las dos primeras personas en entrar a la sala). b) Calculado para los filtros 3, 4 y 5. La métrica rara vez supera el 0.75 por lo que los histogramas no se pueden validar. Los ceros suelen ocurrir porque no se han podido extraer ningún histograma.	60
Fig. A.6. El gráfico muestra la métrica de Bhattacharyya para el filtro 2 y el producto de las métricas 1 y 2, a modo de medida de correlación. El resultado indica que los factores que hacen fallar al recurso de color son mayormente externos al algoritmo y afectan prácticamente por igual a todos los filtros.	61

Lista de Tablas

Tabla 4.1. Resultados de la evaluación del algoritmo basado en filtro de Kalman.....	44
Tabla 4.2 MOTP medio obtenido en los seminarios evaluados	45
Tabla 4.3 MOTA medio obtenido en los seminarios evaluados.....	45
Tabla 4.4 Resultados cuantitativos para un tamaño de voxel de 2cm.	47
Tabla 4.5. Resultados comparativos entre propuestas de distancia de remuestreo y de redistribución de partículas.....	52

1 Introducción

1.1 Escenario

El seguimiento visual es una funcionalidad básica dentro de sistemas de video-vigilancia, de reconocimiento de la actividad humana o la robótica, por citar unos pocos. Dentro de estos campos destaca el crecimiento en los últimos años del desarrollo de aplicaciones de nuevos interfaces hombre-máquina, con proyectos como el estadounidense VACE [29] o el Europeo CHIL (Computers in the Human Interaction Loop) [11] que está financiado por la European Commission's Sixth Framework Programme. Dentro de este último, la UPC tiene un papel relevante.

El objetivo del proyecto CHIL es dar un salto cualitativo en la manera en que usamos los ordenadores. La meta es crear entornos en los que los ordenadores sirvan a personas que interactúan con otras y que no deberían estar pendientes del manejo de la máquina. Ello significa que los ordenadores, en vez de procesar de una manera aislada, en función de órdenes directas del ser humano, interactúen con la persona o personas presentes en un cierto entorno e interpreten las necesidades que puedan ser atendidas. Las aplicaciones CHIL basan su funcionamiento en aproximaciones del concepto de percepción humana para poder atender a personas con las mínimas órdenes.

Para conseguir este objetivo, se han definido una serie de temas claves a desarrollar:

- Interfaces de usuario multimodales que permitan observar, reconocer e interpretar todos los datos e indicios disponibles que permitan explicar las interacciones e intenciones humanas.
- Un conjunto de servicios basados en la interpretación del comportamiento humano. Estos servicios deben entregar información al usuario de manera adecuada. Dichos servicios pueden incluir mejorías en la intercomunicación entre personas, soporte a la memoria humana, provisión de ayuda y datos en reuniones, así como otros servicios y aplicaciones que pueden derivar de una herramienta tan potente.
- Una infraestructura de soporte para los servicios CHIL con procesado autonómico, software de automantenimiento, arquitectura flexible y orientado a integrar numerosos dispositivos de manera inintermitente y dinámica.

Gracias al desarrollo de los procesadores, hoy en día es posible computar grandes cantidades de información en tiempo real, lo que permite el desarrollo de los tres puntos anteriores. Debido a ello las aplicaciones que tratan de interpretar el lenguaje verbal y no verbal están evolucionando rápidamente. Se espera que el paso de interacciones sólo persona-ordenador a la mejora de la interpretación del comportamiento del ser humano aumente en gran medida la productividad y disminuya la frustración en el manejo de computadoras.

1.2 Estado del arte

El desarrollo de aplicaciones que permitan reconocer interacciones entre las personas basándose en la observación a través de las señales de sensores multimodales es objetivo común de varias especialidades científicas, como el procesado de audio, el procesado de imagen, o la interpretación del lenguaje humano. Entre ellas procesado de imagen juega un importante papel, a través de varias líneas de investigación, entre las que podríamos destacar el reconocimiento de caras, el análisis de gestos o el seguimiento de múltiples personas, que es el tema central de este trabajo.

El seguimiento multipersona es difícil dada la complejidad que supone modelar los movimientos e interacciones de los seres humanos, aún estando en un entorno limitado y conocido como una sala. Además hay otras dificultades, como occlusiones, cambios de iluminación o aparición de objetos espurios (elementos dentro de la sala que no hay que seguir, como sillas, percheros u otros objetos del mobiliario).

Algunos enfoques actuales de seguimiento visual se basan en estrategias clásicas como el filtro de Kalman [14, 19, 27] o en algoritmos de obtención de hipótesis acerca de la ubicación del objetivo mediante la extracción de características, como por ejemplo el histograma de color, la reconstrucción 3D o los contornos de una persona [3, 28]. El problema fundamental de las primeras es que no constituyen un enfoque multimodal en lo que a distribución de probabilidad se refiere. Si dibujásemos la distribución de probabilidad del movimiento de las personas en una sala fotograma a fotograma veríamos que los valores obtenidos no se agruparían entorno a una esperanza. Aparecerían diversos picos o modos. Es posible aproximar distribuciones multimodales empleando varias distribuciones unimodales, pero no es una solución robusta al problema de seguimiento dado que dichas aproximaciones siguen siendo una aproximación muy limitada del problema.

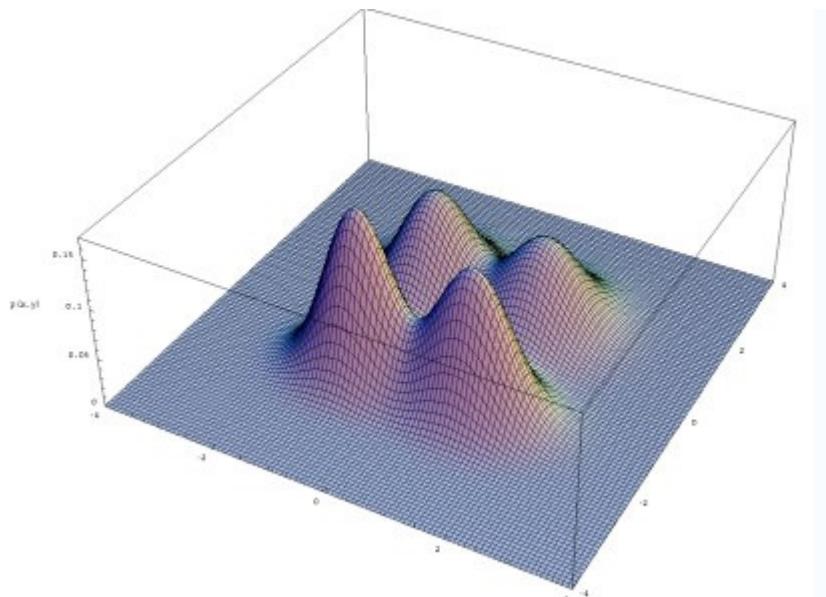


Fig. 1.1 Distribución de probabilidad multimodal.

En cuanto al análisis de los espacios de características y especialmente los de color, podemos destacar que no proporcionan información suficientemente robusta para la

identificación de objetivos, dados los cambios de iluminación, oclusiones, reducida definición de las cámaras o el simple y posible hecho de que dos personas tengan vestimentas de colorido muy similar o igual. Desde hace unos años se viene desarrollando una estrategia diferente, el filtro de partículas, que está basado en métodos de Monte Carlo. Sus numerosas aplicaciones abarcan también algoritmos de seguimiento 2D y 3D, pues se ha demostrado que es una tecnología que permite tratar procesos cuya dinámica se puede modelar mediante una función de densidad de probabilidad (*pdf*) multimodal de manera más efectiva que las soluciones clásicas.

Cuando el proyecto CHIL ha sido descrito (apartado 1.1), se ha definido la necesidad de disponer de un entorno de unas características concretas. Este entorno es lo que llamamos sala inteligente [33]. Una sala inteligente es una habitación que dispone de un conjunto de sensores no invasivos (no emiten o irradian señales al interior de la sala) y cuyo cableado no se encuentra, en general, en el interior del espacio del habitáculo. Estos sensores son típicamente cámaras y micrófonos que captan información de lo que ocurre en la sala para ser procesada y permitir la interpretación de la escena por parte de un computador.

La contribución de este proyecto al ámbito del seguimiento visual de múltiples personas en entornos con múltiples vistas se puede dividir en los siguientes bloques:

- En el entorno de trabajo de este proyecto en concreto, en la sala CHIL de la UPC, que describiremos más detalladamente en el capítulo 3, el algoritmo de seguimiento se hace directamente sobre datos en 3D, sin ningún tipo de marcador corporal, simplemente con una reconstrucción del mundo tridimensional a través de múltiples vistas 2D. Estos datos tridimensionales son un conjunto de voxels, cubos de un tamaño pequeño y determinado cuyos ejes están alineados con las coordenadas de referencia.
- Los datos 3D sólo modelan las formas de los elementos de la escena. A partir de un conjunto de máscaras binarias se obtiene una representación volumétrica de las personas, así como de algunos objetos, presentes en sala. Esto significa que la información de color y de contorno no se explota en el algoritmo.
- El algoritmo está basado en Filtros de Partículas y el presente trabajo añade una nueva combinación de los elementos que los definen adaptándose al problema y a los datos disponibles.

El análisis tridimensional añade ventajas e inconvenientes a considerar. La ventaja de hacer un *tracking* 3D a partir de datos 3D es que no hay cambios de escala y resulta geométricamente más ventajoso localizar un objetivo. En 2D, el hecho de acercarse o alejarse de una posición determinada constituye un cambio en el tamaño aparente en la imagen proyectada por el objetivo a seguir. Sin embargo en 3D la complejidad de cómputo de los datos aumenta. La mínima unidad de información en la imagen 2D es el píxel, mientras que en 3D es el voxel. El píxel es una unidad puntual, y como tal facilita su procesado (por ejemplo, resulta trivial asignar un color a un píxel dada una imagen 2D). El voxel es una unidad con un cierto tamaño (la asignación de color no es sencilla ya queda cara podría tener un color, que puede ser un voxel de superficie o interior, etc.)

El **objetivo** básico de este proyecto es diseñar un algoritmo de seguimiento de múltiples personas en un entorno cerrado con múltiples vistas, basándonos en tecnologías de filtros de partículas. Este algoritmo ha de estar orientado al voxel y tiene que presentar una cierta calidad y precisión evaluables de modo objetivo, que permita demostrar que en términos de

calidad supera a otras soluciones, como Kalman. Finalmente, se busca que el diseño sea de implementación sencilla y que sea una propuesta abierta y escalable, es decir, que su desarrollo no se limite a la versión presentada en este trabajo sino que permita un ajuste en sus parámetros y la posibilidad de añadir otras características y modalidades para la estimación. Se asume que su implementación en tiempo real no es asequible actualmente. A pesar de ello se busca dotar al algoritmo de la mayor eficiencia posible.

En conjunto, las particularidades mencionadas hacen que el algoritmo a desarrollar constituya un enfoque novedoso a un problema ya planteado en otros trabajos de seguimiento en general, presentando también una especificidad destacada en el ámbito del proyecto CHIL.

1.3 Organización

El presente escrito pretende explicar de modo sencillo el punto de partida del seguimiento, como se ha adaptado al entorno de sala inteligente disponible y qué resultados ofrece el diseño propuesto. El trabajo se divide en los siguientes apartados:

El Capítulo 2 introduce al lector en la teoría del seguimiento bayesiano en general y de los filtros de partículas como solución más adecuada. En primer lugar se define formalmente el problema. A continuación se revisa la solución clásica de Kalman y sus limitaciones como aproximación al problema real. Finalmente se define el filtro de partículas como aproximación general y robusta del seguimiento de personas. Dentro de este subapartado se analizan las soluciones SIS (*Sequential Importance Sampling*) y SIR (*Sampling Importance Resampling*), así como el planteamiento de una solución conjunta n -dimensional y el enfoque alternativo con n filtros independientes.

El Capítulo 3 muestra los detalles de enfoque que presenta el diseño del algoritmo. Para ello se describe previamente la sala inteligente de la UPC. Se analiza la obtención de los datos que emplea el algoritmo: los volúmenes reconstruidos a partir de la redundancia 2D que ofrecen las múltiples cámaras. Por último se entra en detalle con los elementos de diseño que caracterizan al algoritmo, tales como la clasificación de los volúmenes reconstruidos, el filtrado de datos innecesarios y los parámetros que gobiernan el filtro de partículas. Este último subapartado merece especial atención dado que constituye la información básica que define al algoritmo en el aspecto de seguimiento.

El Capítulo 4 muestra los resultados obtenidos a través de los experimentos llevados a cabo. Se define para ello las métricas que valoran la calidad objetiva del seguimiento y que son empleadas en las evaluaciones CLEAR organizadas en el contexto por los miembros del proyecto CHIL.

Finalmente el Capítulo 5 expone las conclusiones obtenidas y el trabajo futuro que plantea la línea de investigación seguida en este proyecto.

2 Seguimiento y Filtro de Partículas

El seguimiento es un problema que consiste básicamente en determinar la posición de uno o varios objetivos en cada instante. En el caso particular del vídeo, los instantes vienen determinados por los sucesivos fotogramas.

Generalmente se definen los algoritmos de seguimiento visual en dos tipos, los de Representación y Localización del Objetivo y los de Filtrado y Asociación de Datos [28]. A grandes rasgos, los primeros se basan en la obtención de características que permitan rastrear la escena en busca del objetivo que se asimile más a dichos parámetros. El segundo grupo, de mayor complejidad en general, se basa en incorporar información a priori de la escena –típicamente basada en modelos de movimiento- y la observación de los estados para evaluar la verosimilitud de distintas hipótesis acerca de la posición del objetivo.

Dentro de los algoritmos de Filtrado y Asociación, es cada vez más frecuente que se incluyan elementos de no linealidad y no Gaussianos para modelar con precisión la dinámica del sistema físico. El movimiento de una persona a lo largo del tiempo es un proceso multimodal, es decir, no existe un movimiento con una cierta tendencia en velocidad media y varianza de dicha celeridad alrededor de la media, si no que existen varios patrones. Es por ello que un modelo Gaussiano y lineal es una aproximación poco precisa del problema.

Uno de los métodos de más importancia dentro de las aplicaciones mencionadas es el filtro de partículas, basado en los métodos secuenciales de Monte Carlo. En pocas palabras, el filtro de partículas utiliza un conjunto de puntos para modelar densidades de probabilidad complejas y constituye una generalización de los métodos basados en el filtro de Kalman, ya que puede ser aplicado en cualquier modelo de estados.

En el presente capítulo se formula el problema de seguimiento y se revisa la solución clásica de Kalman, para posteriormente desarrollar el Filtro de Partículas en dos de sus variantes: SIS y SIR. Para ello se ha tomado como referencia el artículo de Arulampalan, Maskell, Gordon y Clapp [1] cuya lectura recomendamos a aquellos que quieran introducirse de un modo más detallado en seguimiento Bayesiano y filtros de partículas.

2.1 Seguimiento Bayesiano No Lineal

Para definir el problema de seguimiento hay que considerar, en primer lugar, un modelo para la evolución del estado de un objetivo. Sea la secuencia de estados $\{x_k, k \in \mathbb{N}\}$; su evolución vendrá dada por la ecuación de estado:

$$x_k = f_k(x_{k-1}, v_{k-1}) \quad (2.1)$$

donde f_k es una función no lineal del estado x_{k-1} y v_{k-1} es un proceso que modela ruido. El objetivo del seguimiento es estimar x_k recursivamente a partir de una serie de observaciones ruidosas:

$$z_k = h_k(x_k, n_k) \quad (2.2)$$

Esta es la ecuación de medida, donde h_k es una función en general no lineal y n_k un proceso estocástico modelando el ruido en la observación. En particular, buscamos estimaciones de x_k basadas en un conjuntos de medidas disponibles $z_{1:k} = \{z_i, i=1, \dots, k\}$, donde k denota el instante de tiempo.

Desde una perspectiva Bayesiana [1], el problema de seguimiento es calcular recursivamente qué grado de verosimilitud tiene un estado x_k en ese mismo instante k , dados los datos disponibles en las observaciones realizadas hasta este momento. Dicho de otro modo, se necesita estimar la densidad de probabilidad $p(x_k | z_{1:k})$. Se asume que la *pdf* inicial $p(x_0 | z_0) \equiv p(x_0)$, también llamada densidad a priori (*prior*), es conocida (siendo z_0 un conjunto vacío de medidas u observaciones) y que, por lo tanto, hay que hacer estimaciones recursivas a partir de las probabilidades anteriores al estado actual en dos pasos: predicción y actualización.

Suponiendo que disponemos de la *pdf* requerida en el estado $k-1$, $p(x_{k-1} | z_{k-1})$, el paso de predicción constituye hallar la densidad a priori mediante la ecuación de Chapman-Kolmogorov:

$$p(x_k | z_{1:k-1}) = \int p(x_k | x_{k-1}) p(x_{k-1} | z_{1:k-1}) dx_{k-1} \quad (2.3)$$

donde se asume la secuencia de estado como un proceso de Markov de orden 1, dado que consideramos $p(x_k | x_{k-1}, z_{1:k-1}) = p(x_k | x_{k-1})$. El modelo probabilístico de la evolución de estado $p(x_k | x_{k-1})$ viene dado por la ecuación de estado y la estadística de ruido conocida, v_{k-1} . En el contexto del seguimiento, esta densidad de probabilidad acostumbra a ser un modelo de movimiento determinado.

En el instante k disponemos de la medida z_k por lo que nos servirá para efectuar el siguiente paso, actualizar la estadística a priori (*prior*) a través de la regla de Bayes:

$$p(x_k | z_{1:k}) = \frac{p(z_k | x_k) p(x_k | z_{1:k-1})}{p(z_k | z_{1:k-1})} \quad (2.4)$$

Donde tenemos la verosimilitud multiplicada por la estadística a priori y normalizada por una constante obtenida según:

$$p(z_k | z_{1:k-1}) = \int p(z_k | x_k) p(x_k | z_{1:k-1}) dx_k \quad (2.5)$$

Así pues, el paso de actualización depende directamente de la verosimilitud $p(z_k | x_k)$ definida en la ecuación de medida y la estadística de ruido n_k .

Las relaciones presentadas en (2.3) y (2.4) son la base de la solución Bayesiana óptima al problema de seguimiento, pero no constituyen más que una formulación conceptual. Ésta no es calculable en general de manera analítica, aunque existen soluciones en casos muy concretos que consideran ciertas restricciones, como el filtro de Kalman. Cuando la solución analítica es realmente intratable, entonces se recurre a soluciones subóptimas como los filtros de partículas, que dan una aproximación muy efectiva.

2.2 El filtro de Kalman

El filtro de Kalman asume que la densidad a posteriori es, en cada instante, Gaussiana y por lo tanto queda completamente definida por su media y su covarianza. Si $p(x_{k-1} | z_{1:k-1})$ es Gaussiana, se puede demostrar que $p(x_k | z_{1:k})$ también lo es. Para ello se asume que v_{k-1} y n_k , de las ecuaciones de estado y de medida, tienen distribuciones Gaussianas de parámetros conocidos y que tanto $f_k(x_{k-1}, v_{k-1})$ como $h_k(x_k, n_k)$ son funciones lineales conocidas de sus respectivas variables (x_{k-1}, v_{k-1}, x_k y n_k). Siendo así, las ecuaciones de estado y de medida pueden ser reformuladas como sigue:

$$x_k = F_k x_{k-1} + v_{k-1} \quad (2.6)$$

$$z_k = H_k x_k + n_k \quad (2.7)$$

F_k y H_k son las matrices que definen las funciones lineales. Llamaremos Q_{k-1} y R_k a la covarianza de los procesos v_{k-1} y n_k respectivamente. Además consideramos que dichos procesos tienen media cero y son estadísticamente independientes. El filtro de Kalman puede enfocarse desde un punto de vista Bayesiano siguiendo las ecuaciones 2.3 y 2.4 [1], aunque también podemos abordar el problema mediante la minimización del error cuadrático. El filtro de Kalman puede ser enunciado como la siguiente relación recursiva:

$$p(x_{k-1} | z_{1:k-1}) = N(x_{k-1}; m_{k-1|k-1}; P_{k-1|k-1}) \quad (2.8)$$

$$p(x_k | z_{1:k-1}) = N(x_k; m_{k|k-1}; P_{k|k-1}) \quad (2.9)$$

$$p(x_k | z_{1:k}) = N(x_k; m_{k|k}; P_{k|k}) \quad (2.10)$$

Donde $N(x; m; P)$ es la densidad Gaussiana de la variable x , de media m y covarianza P . Aplicando 2.3 y 2.4 obtendremos:

$$m_{k|k-1} = F_k m_{k-1|k-1} \quad (2.11)$$

$$P_{k|k-1} = Q_{k-1} + F_k P_{k-1|k-1} F_k^H \quad (2.12)$$

$$m_{k|k} = m_{k|k-1} + K_k (z_k - H_k m_{k|k-1}) \quad (2.13)$$

$$P_{k|k} = P_{k|k-1} - K_k H_k P_{k|k-1} \quad (2.14)$$

Donde K_k es la llamada ganancia de Kalman y tiene la siguiente expresión:

$$K_k = P_{k|k} H_k^H (H_k P_{k|k} H_k^H + R_k)^{-1} \quad (2.15)$$

El superíndice H denota el transpuesto conjugado de la matriz. Este algoritmo iterativo es la solución óptima al problema de seguimiento con suposiciones altamente restrictivas. Ningún algoritmo es capaz de hacerlo mejor que el filtro de Kalman en un entorno lineal y Gaussiano. El problema que se plantea es, evidentemente, que el seguimiento multipersona no es un problema lineal ni Gaussiano, por lo que esta aproximación clásica no resulta efectiva [1].

2.3 El filtro de partículas

Los filtros de partículas [1] son algoritmos basados en los métodos de Monte Carlo. La idea que hay tras ellos es aproximar la densidad de probabilidad a posteriori con un conjunto discreto de muestras o partículas con pesos asociados, ofreciendo la posibilidad de convertir integrales intratables en sumatorios calculables de manera relativamente sencilla. Conforme el número de partículas crece, dicha aproximación se asemeja más a la función continua que describe la probabilidad a posteriori.

De un modo más formal, sea $\{x_{0:k}^i, w_k^i\}_{i=1}^{N_s}$ un conjunto de medidas aleatorias que caracterizan la densidad a posteriori, donde $\{x_{0:k}^i; i = 0, \dots, N_s\}$ son partículas, $\{w_k^i; i = 0, \dots, N_s\}$ los pesos asociados a cada una de estas partículas y $\{x_{0:k}; x_k, j = 0, \dots, k\}$ es el conjunto de estados en cada instante k . Entonces la densidad a posteriori puede ser aproximada tal y como hemos dicho, resultando la siguiente expresión:

$$p(x_{0:k} | z_{1:k}) \approx \sum_{i=1}^{N_s} w_k^i \delta(x_{0:k} - x_{0:k}^i) \quad (2.16)$$

Los pesos w_k^i son calculados a través del principio de “Importance Sampling”[1], que nos permite obtener muestras cuando la *pdf* a aproximar es muy compleja. Supongamos que $p(x)$ es una densidad de estas características, pero que existe una $r(x)$ evaluable tal que $p(x) \propto r(x)$. Además, obtendremos el conjunto de muestras $x^i \sim q(x)$, $i = 1, \dots, N_s$ de la llamada densidad de importancia $q(\cdot)$, con tal de evaluar los pesos asociados a cada partícula. Formalmente, una aproximación de la densidad $p(\cdot)$ la podemos obtener como:

$$p(x) \approx \sum_{i=1}^{N_s} w^i \delta(x - x^i) \quad (2.17)$$

obteniendo los pesos según la siguiente expresión:

$$w^i \propto \frac{r(x^i)}{q(x^i)} \quad (2.18)$$

La conclusión resultante es que si obtenemos un conjunto de N_s muestras a partir de la densidad de importancia $q(\cdot)$ en el estado $x_{0:k}$, es posible conformarlas atribuyéndoles los pesos que nos darán la aproximación de $p(x_{0:k}|z_{1:k})$:

$$w_k^i \propto \frac{p(x_{0:k}^i | z_{1:k})}{q(x_{0:k}^i | z_{1:k})} \quad (2.19)$$

Una representación gráfica ayuda a ver este concepto. Suponiendo que, como hemos dicho anteriormente, no podemos obtener muestras de $p(\cdot)$, las dibujaremos según $q(\cdot)$. En la figura 2.1 se puede observar dicho proceso, en el cual las partículas se dibujan según la densidad de importancia para luego evaluar sus pesos mediante la densidad a posteriori. Una partícula más grande indica que su peso es mayor, dado que coincide con valores altos de la pdf , mientras que a las pequeñas les ocurre lo contrario (cabiendo la posibilidad de que desaparezcan). Claramente se puede ver que el peso de cada partícula es su medida de verosimilitud en el estado k .

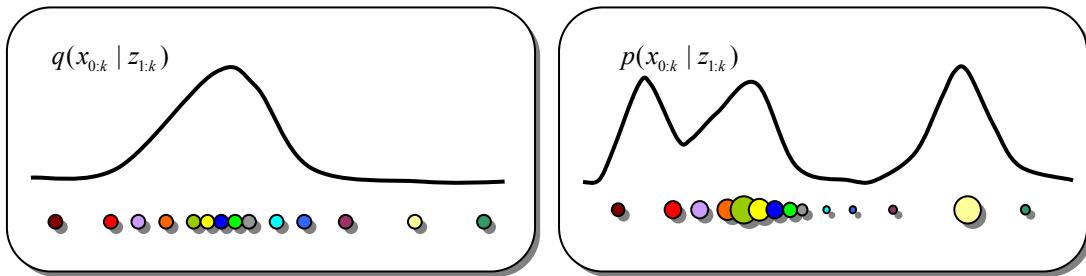


Fig. 2.1 Propuesta de distribución y cálculo de pesos

En los trabajos dedicados a los filtros de partículas podemos hallar gran número de variaciones de la teoría mostrada hasta ahora. Dichas variaciones pretenden acomodar el filtro de partículas a la aplicación para la cual ha sido diseñado. Veremos también, que en el presente trabajo se propone una variación acorde con el seguimiento multi-persona multi-vista 3D en el entorno de salas inteligentes. Previamente, cabría revisar dos de las versiones de filtro de partículas más importantes con el fin de profundizar en la implementación del algoritmo básico: SIS PF (*Sequential Importance Sampling*) y SIR PF (*Sampling Importance Resampling*) [1]

2.3.1 SIS PF

En el apartado anterior hemos introducido el principio de Importance Resampling para ver como se obtiene un peso en un estado determinado, pero se ha obviado el hecho de que estamos ante un algoritmo iterativo. Si las muestras se obtienen como la propagación de las existentes en el estado anterior, esto es, $q(x_{0:k}|z_{1:k})=q(x_k|x_{0:k-1},z_{1:k})q(x_{0:k-1}|z_{0:k-1})$ entonces podemos aplicar 2.4 a la definición de los pesos del filtro de partículas, asumiendo además que estamos ante un sistema de estados Markoviano de orden 1:

$$\begin{aligned} w_k^i &\propto \frac{p(z_k | x_k^i) p(x_k^i | z_{k-1})}{q(x_k^i | x_{k-1}^i, z_k) q(x_{k-1}^i | z_{k-1}) p(z_k | z_{k-1})} = \frac{p(z_k | x_k^i) p(x_k^i | x_{k-1}^i) p(x_{k-1}^i | z_{k-1})}{q(x_k^i | x_{k-1}^i, z_k) q(x_{k-1}^i | z_{k-1}) p(z_k | z_{k-1})} \\ w_k^i &\propto w_{k-1}^i \frac{p(z_k | x_k^i) p(x_k^i | x_{k-1}^i)}{q(x_k^i | x_{k-1}^i, z_k) p(z_k | z_{k-1})} \end{aligned} \quad (2.20)$$

Así pues, en el filtro SIS se propagan secuencialmente las muestras que se dibujan en función de la densidad de importancia y el conjunto de pesos asociados a las mismas.

Un problema importante de los filtros basados en esta técnica es la degeneración de las partículas que lo componen, causado por el incremento estocástico de la varianza de los pesos. Tras varias iteraciones se observará como tan sólo una partícula sobrevive, es decir, todas las partículas menos una tienen pesos negligibles, tiendiendo a 0. En [26] podemos hallar una medida fiable de degeneración de partículas. Dado que no puede ser evaluada de manera exacta, se utiliza una aproximación:

$$N_{\text{eff}} = \frac{1}{\sum_{i=1}^{N_s} (w_k^i)^2} \quad (2.21)$$

Dados pesos normalizados, esta medida, también conocida como tamaño efectivo de muestreo, va de 1 hasta N_s , siendo la unidad el valor que indica que tan sólo una partícula tiene peso. A través del valor de esta expresión y con la ayuda de un umbral adecuado se puede combatir efectivamente la degeneración de partículas. Para ello, la técnica más empleada es el Remuestreo. Básicamente, consiste en reubicar todas las partículas y sus pesos asociados en un nuevo conjunto cuyos pesos sean todos iguales:

$$\{x_k^i, w_k^i\} \rightarrow \left\{x_k^i, \frac{1}{N_s}\right\} \quad (2.22)$$

Con el remuestreo se busca hacer sobrevivir aquellas partículas o muestras más verosímiles, es decir, de más peso, regenerándolas en n nuevas partículas por cada muestra antigua, y asignándoles un mismo peso a todas ellas. En (2.22) se considera el peso normalizado al número total de partículas empleadas.

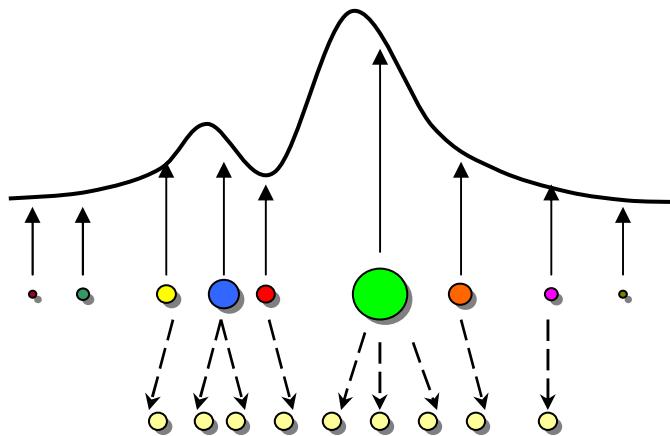


Fig. 2.2 Las partículas con más peso, al constituir puntos de mayor probabilidad, son remuestreadas con mayor cardinalidad, mientras que las que tienen poco peso se consideran negligibles y “mueren”.

El remuestreo soluciona la degeneración de partículas, pero aún así se puede observar un efecto de concentración alrededor de aquellas que poseen más peso (fig 2.3). Esto se traduce en que el conjunto de muestras se ubica en puntos muy concretos del objetivo y reduciendo así la efectividad de la estimación. Un algoritmo de filtro de partículas que

permite evitar esa concentración de partículas es el SIR (*Sampling Importance Resampling*), que veremos a continuación.

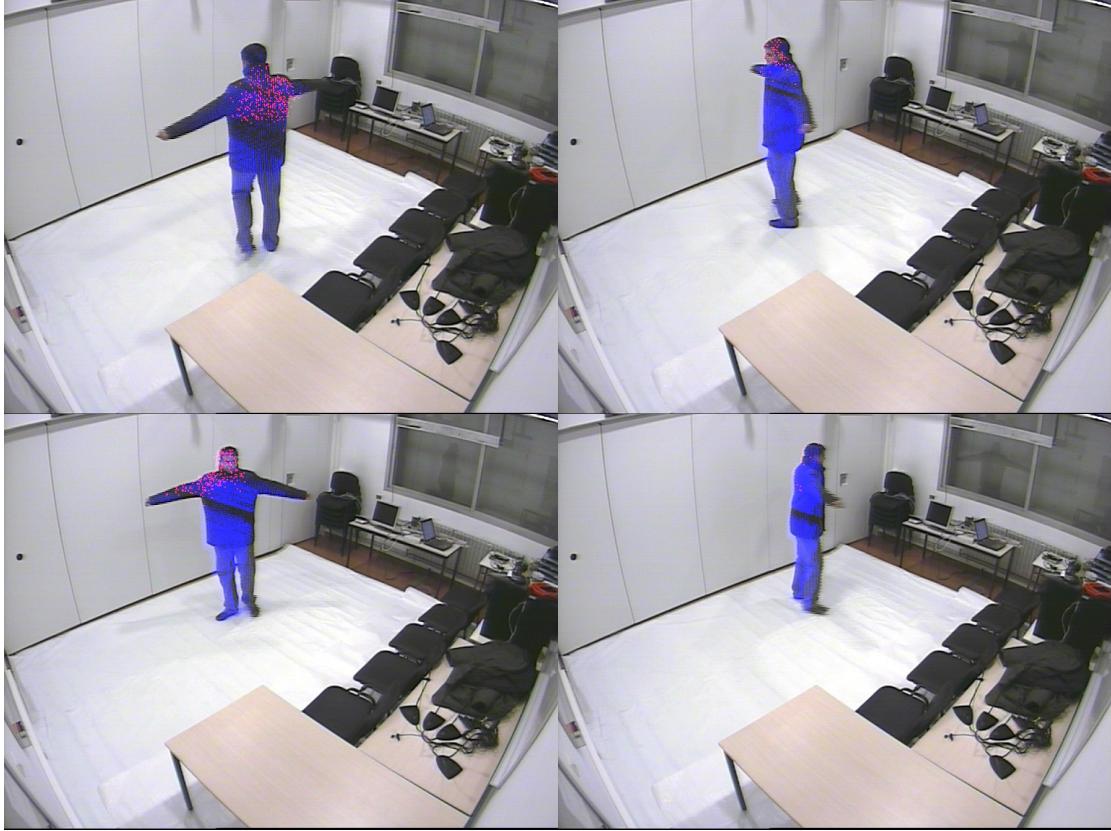


Fig. 2.3 Efecto de concentración de las partículas (representadas en rojo). En azul se han pintado los voxels. Se muestran los fotogramas 401, 421, 441 y 461. Al cabo de 60 fotogramas todas las partículas se concentran en unos puntos muy concretos.

2.3.2 SIR PF

El llamado filtro de partículas SIR no es más que un caso particular del SIS, en el cual escogemos adecuadamente su densidad de importancia y los instantes de remuestreo. La $pdf q(x_k|x_{k-1}, z_k)$ pasa a ser la densidad de transición $p(x_k|x_{k-1})$, de modo que sustituyendo en 2.20 obtenemos:

$$w_k^i \propto w_{k-1}^i \frac{p(z_k | x_k^i)}{p(z_k | z_{k-1})} \Leftrightarrow w_k^i \propto w_{k-1}^i p(z_k | x_k^i) \quad (2.23)$$

Por otra parte, el remuestreo se efectúa en cada iteración, por lo que se evita la concentración de partículas en un punto determinado a lo largo del tiempo. Se puede observar que, al aplicar esta modificación, la expresión de los pesos se simplifica:

$$w_k^i \propto p(z_k | x_k^i) \quad (2.24)$$

ya que tras el remuestreo todos los pesos valen $1/N_s$. El problema que plantea el uso de este algoritmo es que la densidad de importancia es independiente de la observación por lo que los estados se analizan a ciegas, sin conocimiento previo de los mismos. Esto se

traduce en que la eficiencia de estos algoritmos va directamente relacionada con el número de partículas que emplean.

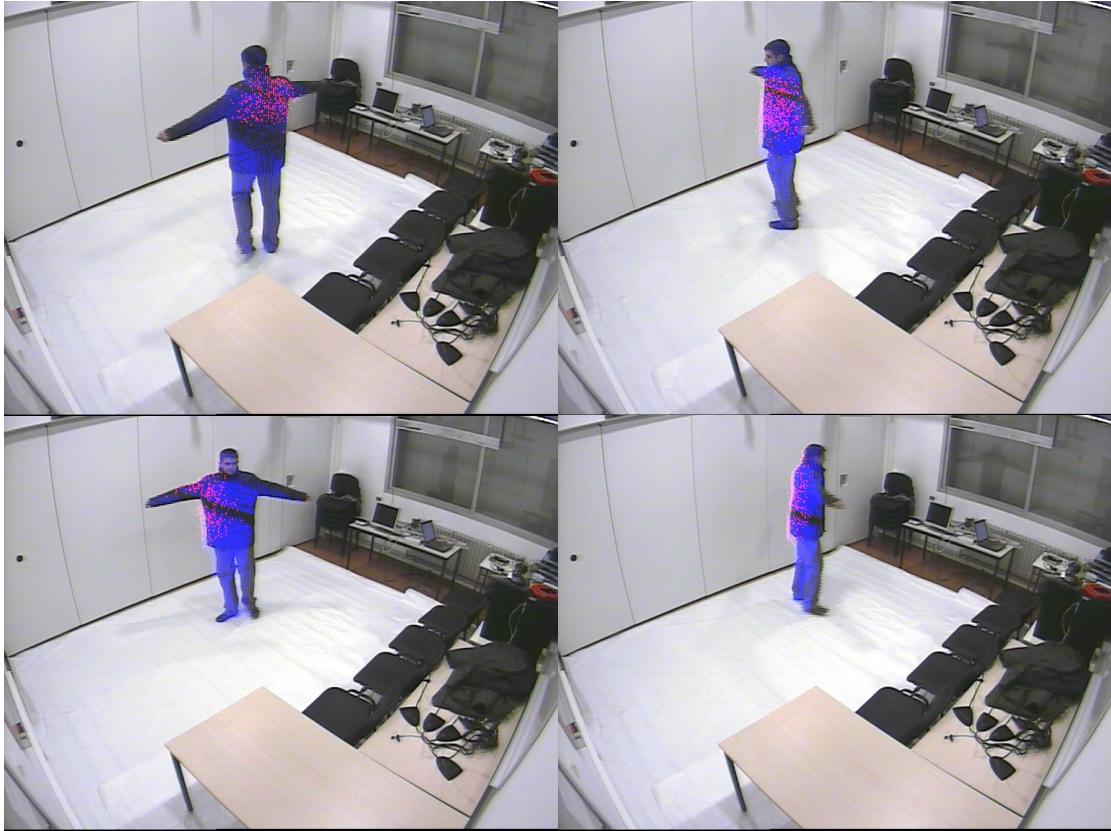


Fig 2.4 Con el método SIR se evita el efecto de concentración de partículas con el tiempo mediante el remuestreo en cada iteración. Se muestran los fotogramas 401, 421, 441, 461, como en la Fig 2.3.

2.3.3 Filtro de Partículas aplicado a múltiples objetivos.

Hasta ahora se ha analizado el algoritmo de filtro de partículas, pero no se ha profundizado en la implicación que tiene aplicarlo al seguimiento en general. Antes de extenderlo a múltiples objetivos presentaremos las pautas que se necesitan seguir para una sola persona.

Cada partícula puede representar una serie de características del sistema a estimar, que constituyen el modelo del objetivo. El modelo más sencillo para el seguimiento es tener cada partícula representando una coordenada del espacio 3D (x , y , z). En este caso, estaríamos formulando un filtro de partículas para un solo objetivo y una única magnitud a estimar, su posición. Esta magnitud tiene una dimensión que viene dada por su naturaleza, en este caso tridimensional. Si se consideran otras magnitudes o características, la dimensionalidad de la partícula aumenta. Si, por ejemplo, añadiésemos al modelo la velocidad y el color RGB cada partícula modelaría un vector de nueve dimensiones (x , y , z , v_x , v_y , v_z , R , G , B). Con el modelo definido se actúa del modo siguiente:

- Se generan muestras a partir de una densidad de importancia $q(\cdot)$ (ver apartado 2.3), que puede ser, para el caso del SIR PF, un modelo de movimiento $p(x_k|x_{k-1})$ que reubica las partículas.
- Se evalúan los pesos w_k^i (2.20) con una pdf modelando la verosimilitud de que una región del espacio pertenezca a una persona presente en la sala. Una vez calculados

w_k^i ya disponemos de una aproximación de la densidad a posteriori. La esperanza de esta aproximación es la hipótesis acerca de la posición de la persona (y de la velocidad, el color dominante o cualquier otra magnitud que incluya el modelo).

- Se remuestrean las partículas para mantener el tamaño efectivo de muestreo (2.21).

El problema del seguimiento de múltiples personas requiere consideraciones adicionales. En [2] se presenta una solución óptima, el filtro de partículas conjunto (*Joint PF*). En esta estrategia si m es el número de objetivos a seguir y empleamos el modelo de objetivo basado en un único punto tridimensional entonces cada partícula debe contener las coordenadas de todos los objetivos presentes ($x_1, y_1, z_1, \dots, x_m, y_m, z_m$). Aún siendo la solución óptima, plantea problemas de implementación. El primero es que la carga computacional crece exponencialmente con la dimensión del sistema de estados (el número de objetivos). Por otra parte, para el caso particular de seguimiento multi-persona que se trata en este trabajo, resulta poco adaptativo y, en consecuencia, ineficiente, en el sentido de que para un número variable de objetivos se requiere la verificación y modificación de las dimensiones de la partícula (e incluso del número de partículas) en cada iteración, en función de la cantidad de personas presentes en la sala.

Las dificultades que entraña el filtro de partículas conjunto requieren formular el problema para hallar una aproximación subóptima, válida para el problema de seguimiento de múltiples personas. Un posible asunción es considerar que los objetivos a seguir son independientes [2], lo cual no es cierto, ya que presentan interacciones y pueden posicionarse muy cerca unos de otros, condicionando sus trayectorias – aunque, por el hecho de tener datos tridimensionales nunca se ocultan ocupando aparentemente el mismo espacio. Esta interpretación permite utilizar un conjunto de m filtros, cada uno para un solo objetivo. Este planteamiento de filtros separados ahorra coste computacional y permite alta adaptabilidad a situaciones de variación de número de personas del entorno con un clasificador adecuado. El inconveniente que presenta la solución subóptima es que, al haber asumido independencia entre los objetivos, se requiere de un modelado de interacciones que permita evitar intercambios o fusión. Entendemos por intercambio la situación en que el sistema tiene identificados a las personas A y B y en la siguiente iteración intercambia dichas denominaciones (y por tanto las trayectorias temporales de A y B). Por fusión entendemos el evento que tiene lugar cuando dos objetivos cercanos forman un volumen conexo de forma que de tener dos personas identificadas se pasa a detectar una sola. En el capítulo 3 veremos la aplicación del algoritmo de filtro de partículas al seguimiento en entornos de sala inteligente e introduciremos los modelos para las interacciones en este escenario.

3 Seguimiento Multi-Persona en Salas Inteligentes

En el presente capítulo se describirá de forma sucinta la configuración disponible en la Sala Inteligente de la UPC. De la definición de este entorno pasaremos a analizar, en la sección 3.2, qué datos y espacios de características se explotan. Finalmente, en el apartado 3.3 se analizará de qué modo se implementa el seguimiento 3D con múltiples vistas y personas, basándonos en la teoría revisada en el capítulo 2.

3.1 Un entorno Multicámara: La sala CHIL

El entorno de sala inteligente sobre el que se ha desarrollado y probado el algoritmo es la sala CHIL del Campus Nord de la UPC. Se trata de una habitación que dispone de un conjunto de cámaras calibradas y sincronizadas y *arrays* de micrófonos[21].

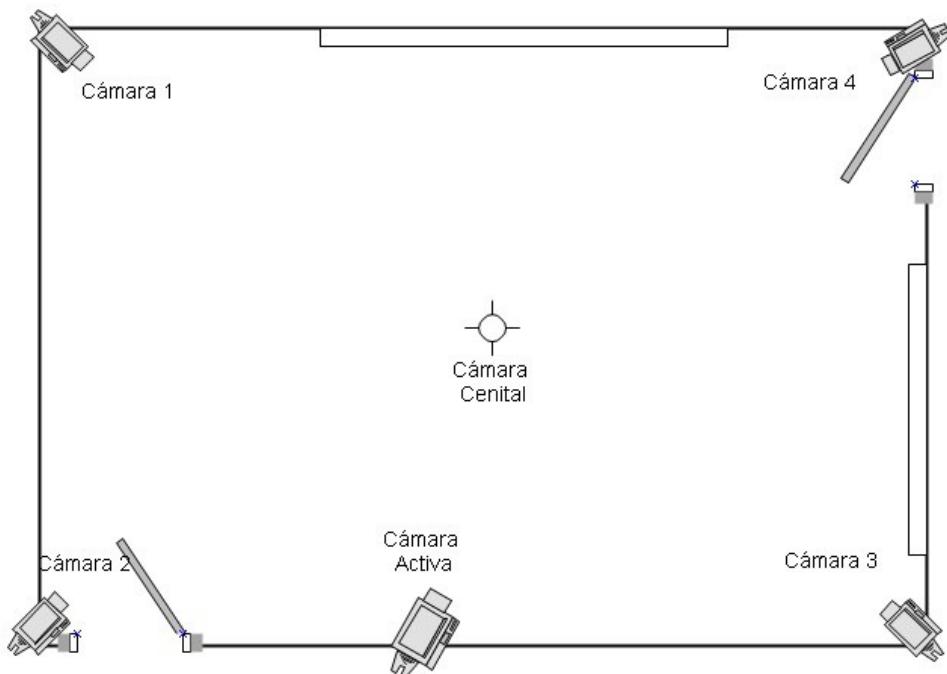


Fig. 3.1 Diagrama esquemático de la Sala CHIL.

En la Fig 3.1 podemos observar la disposición de cuatro cámaras fijas y una cámara activa que no será utilizada en el presente trabajo. Para completar el conjunto de cámaras empleado falta mencionar la existencia de una quinta cámara fija, la cámara cenital, que proporciona una vista completa de la disposición de la sala en el plano XY, tal y como se puede ver en la Fig 3.2. Todas las cámaras tienen una referencia numérica única. Las fijas en las esquinas van de la 1 a la 4, siguiendo las posiciones que se ven en 3.1, mientras que la cámara 5 es la cenital.



Fig. 3.2 Imagen obtenida desde la cámara cenital

El origen de coordenadas de la sala es el punto que se ve en la Fig 3.3 a través de la imagen que reporta la cámara 3.



Fig. 3.3 Origen de coordenadas de la sala CHIL

Al ser fijas las cámaras cubren una región determinado del mundo 3D. Dichas posiciones pueden ser mapeadas al sistema de coordenadas de las imágenes obtenidas por las cámaras. Para ello se modela la cámara como un punto (centro de cámara) y un plano de proyección situado a distancia f .

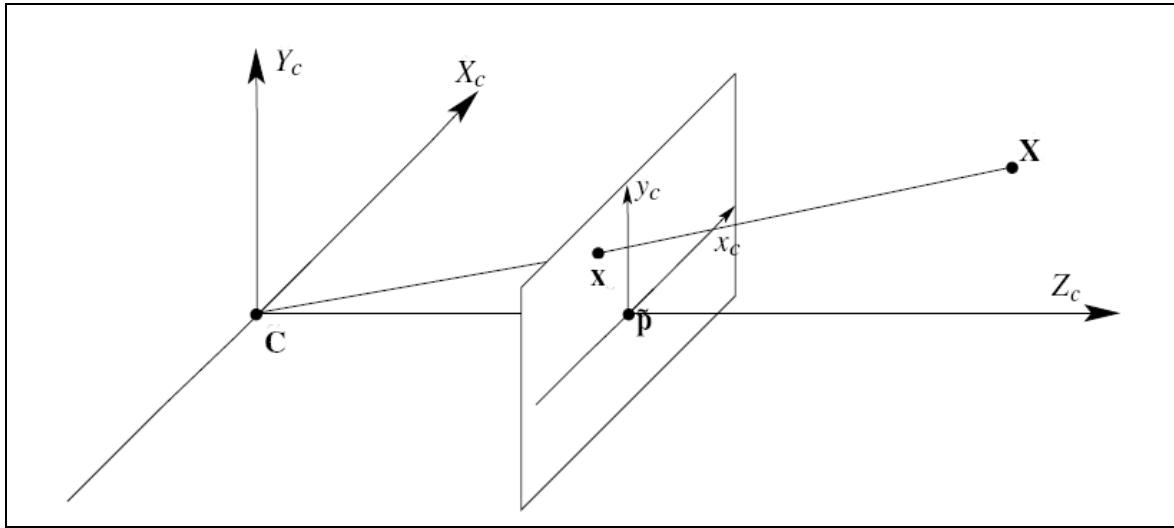


Fig. 3.4 Modelo puntual de cámara.

La consecuencia de estas definiciones, tal y como se ve en la figura 3.3, es que automáticamente podemos relacionar las coordenadas del mundo 3D, el sistema de coordenadas propio de cada cámara y las coordenadas 2D en el plano de proyección. La proyección de las coordenadas 3D respecto al centro de cámara en el plano viene dada por:

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} \rightarrow \begin{pmatrix} x_c \\ y_c \end{pmatrix} = \begin{pmatrix} f \frac{X_c}{Z_c} \\ f \frac{Y_c}{Z_c} \end{pmatrix} \quad (3.1)$$

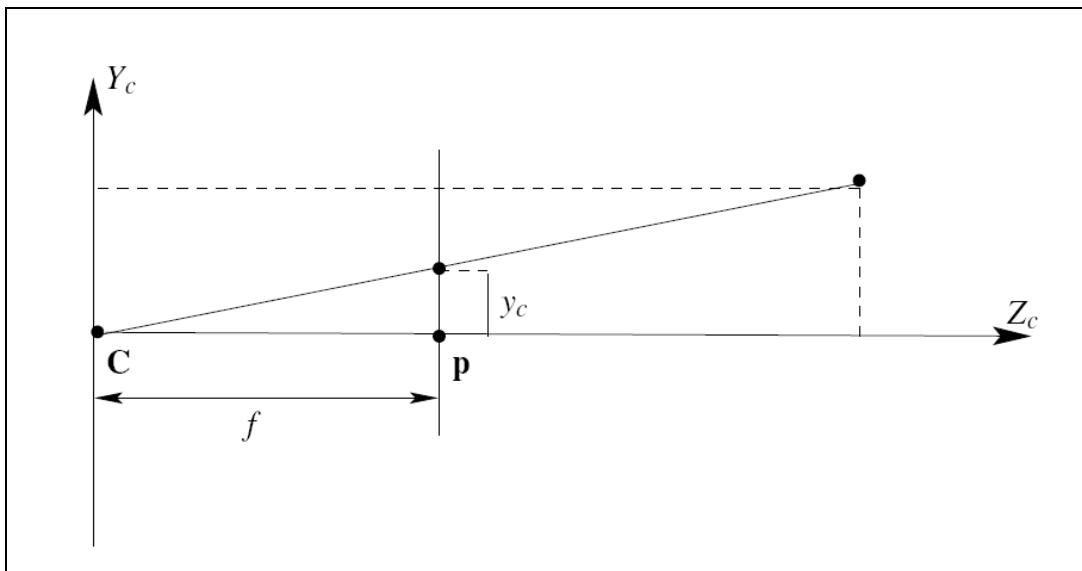


Fig. 3.5 Proyección en el plano focal

Obsérvese que para $Z_c=0$ la función no está definida. Para tener completamente definida la proyección, se recurre a coordenadas homogéneas, resultando la siguiente expresión:

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \\ T_c \end{pmatrix} \rightarrow \begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} = \begin{pmatrix} f \frac{X_c}{Z_c} \\ f \frac{Y_c}{Z_c} \\ 1 \end{pmatrix} \quad (3.2)$$

Con estas dos expresiones se definen los casos más generales de proyección de coordenadas en una cámara modelada puntualmente. Todos los casos particulares, que derivan de estos, y que modelan escalamientos o desplazamientos del origen de coordenadas del plano focal, se resumen en lo que llamamos parámetros de calibración, que se recogen en la matriz K.

Por otra parte, hay que tener en cuenta que el sistema de coordenadas del mundo 3D, que en nuestro caso particular es un punto concreto de la sala CHIL (ver fig. 3.4), estará desplazado y rotado respecto al sistema adaptado de las cámaras. Dicho esto, sea \bar{x} el vector de coordenadas en el plano de proyección de una determinada cámara; sea X el vector de coordenadas homogéneas de un punto 3D de la sala, C las coordenadas del origen adaptado a la cámara (respecto al origen de la sala). La expresión que mapea X en \bar{x} es:

$$\bar{x} = \bar{K}R(\bar{I} | -\bar{C})\bar{X} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} R_1 & R_2 & R_3 & -C_x \\ R_4 & R_5 & R_6 & -C_y \\ R_7 & R_8 & R_9 & -C_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ T_c \end{pmatrix} \quad (3.3)$$

donde R es la matriz ortogonal de rotación. En la sala CHIL, todas las cámaras que usaremos llevan asociados todos los parámetros de rotación, translación y calibración necesarios para mapear coordenadas del mundo 3D en cada una de ellas. Una vez introducidos estos conceptos, veremos como aprovecharlos para obtener los datos necesarios para efectuar el seguimiento 3D.

3.2 Datos en el Escenario

El sistema de múltiples cámaras del que dispone la sala CHIL nos permite explotar diversas características visuales del entorno para obtener datos cuya manipulación nos permita interpretar el comportamiento básico de las personas que hay dentro. El seguimiento multipersona diseñado requiere de una reconstrucción del mundo 3D para poder trabajar. A continuación analizaremos cómo se obtiene y de qué manera nos puede ayudar.

3.2.1 Volúmenes

Como hemos dicho anteriormente, para cualquier instante de tiempo, toda escena es vista desde cinco puntos, creando una redundancia de la información que permitirá discernir entre las partes de la escena consistentes de las que no lo son. A partir de esta selección se

triangulan los puntos 2D de las imágenes de las cámaras para obtener una reconstrucción 3D única de la escena [34].

En primer lugar se define la unidad mínima en términos 3D, el voxel. El voxel es un cubo orientado según los ejes de la sala, cuyas dimensiones son la unidad de longitud mínima de precisión en la que nos moveremos en x , y o z . Dado que el voxel, por definición, tiene un volumen determinado, fijamos un punto del mismo (p. ej. el centroide) como referencia, de modo que tenemos una referencia infinitesimal para una pequeña región de la sala o, lo que es lo mismo, cada voxel se identifica con una única coordenada 3D.

Las imágenes de todas estas cámaras han sido segmentadas empleando un algoritmo basado en la técnica de Stauffer-Grimsom [20] de aprendizaje del fondo (*background learning*), de modo que para cada fotograma en cada cámara existe una máscara binaria que determina qué píxeles representan el escenario (fig. 3.6) –y por lo tanto no son posibles objetivos a seguir– y cuáles no. Cada voxel se proyecta en todas las cámaras de la sala, obteniendo de un punto 3D del mundo 5 pares de coordenadas 2D. Con el resultado binario de estas proyecciones sobre las máscaras binarias se determina la consistencia espacial del voxel [34], classificándolo como “*foreground*” (objeto activo) o “*background*” (objeto de fondo o escenario). En más bajo nivel, consideramos los voxels de fondo como voxels a 0 y el resto diferente de 0 (típicamente con el valor binario 1), denotando de modo booleano el resultado del procedimiento comentado.

Tal y como se puede apreciar en la Fig 3.6, en las máscaras binarias aparecen numerosos espurios, ya sean causados por cambios de iluminación, como los píxeles blancos que aparecen en el suelo, o simplemente, porque se detectan objetos presentes en la sala, que no tienen que ser localizados por el algoritmo, puesto que está enfocado al seguimiento de personas. En el primer caso, los cambios de iluminación suelen ser únicos para cada cámara, por lo que no dan regiones consistentes en el espacio 3D. El segundo caso requiere del diseño de herramientas para eliminarlos de la escena 3D, ya que pueden causar errores en el seguimiento.



Fig. 3.6 Fotograma con su máscara correspondiente.

3.3 Aplicación al seguimiento Multipersona 3D

En este apartado se adaptará la teoría revisada en el capítulo 2. Esta es la contribución del algoritmo diseñado en este proyecto, que procesa los volúmenes reconstruidos, tal y como se explica en 3.2, para generar hipótesis acerca de la posición de las personas en la sala. En primer lugar cabrá hacer una revisión del tratamiento previo que reciben los conjuntos de voxels con los que trabaja el algoritmo. En los subapartados sucesivos se detallará como se han diseñado los elementos más importantes del filtro de partículas.

3.3.1 Clasificación

El primer tratamiento de datos a efectuar es la detección de conectividades entre voxels de “*foreground*” para hallar conjuntos que puedan estar identificando a personas en la sala. Con este propósito nos podríamos servir de herramientas de otro nivel (Bases de Datos de los asistentes, detectores de entrada de personas,...) pero se ha decidido dotar al algoritmo de una capacidad básica de clasificación de conjuntos de voxels como posibles objetivos.

Para ello simplemente se analizan los conjuntos de voxels conexos mirando su cardinalidad (en términos de número de voxels de cada conjunto conexo) y su altura sobre el suelo en coordenadas reales del espacio 3D. El algoritmo considera que el número de voxels multiplicado por su tamaño (en cm) ha de ser superior a 3000. En lo referente a la altura, se ha fijado un mínimo de 60 cm entre el voxel superior y el inferior, teniendo en cuenta que es suficiente para discriminar posible mobiliario desplazado en la sala (p. ej. sillas) permitiendo detectar a personas sentadas o cuya reconstrucción sea parcial debido a occlusiones.

Una vez se considera que el conjunto de voxels conexos es un posible objetivo se crea un filtro de partículas partiendo del centro de masas. Para evitar inicializar un filtro en cada iteración se añade como restricción que no puede haber filtros de partículas con su centroide ubicado a menos de 60 cm del centro de masas de voxels conexos para los que anteriormente se ha creado un PF.

En la fig 3.7 podemos apreciar la inicialización de un filtro de partículas desde dos cámaras. En los fotogramas superiores el clasificador de volúmenes conexos sólo ha detectado una posible persona, por lo que se ha inicializado un filtro de partículas. El modelo que hemos empleado para las personas a seguir se reduce básicamente a estimar el centroide de las mismas. Este punto está marcado con la cruz verde. En las imágenes inferiores se ha inicializado un segundo filtro (elipsoide amarilla) ya que el conjunto de voxels conexos que resulta de la reconstrucción 3D de esa persona en concreto se ajusta a las características de cardinalidad y altura antes citadas.

Se puede ver claramente, a tenor del ejemplo revisado, que un algoritmo más desarrollado de clasificación puede dar lugar a un seguimiento más preciso, dado que una identificación rápida y correcta de una persona mejorará las estadísticas asociadas al seguimiento global dentro de la sala, como veremos en el capítulo 4.

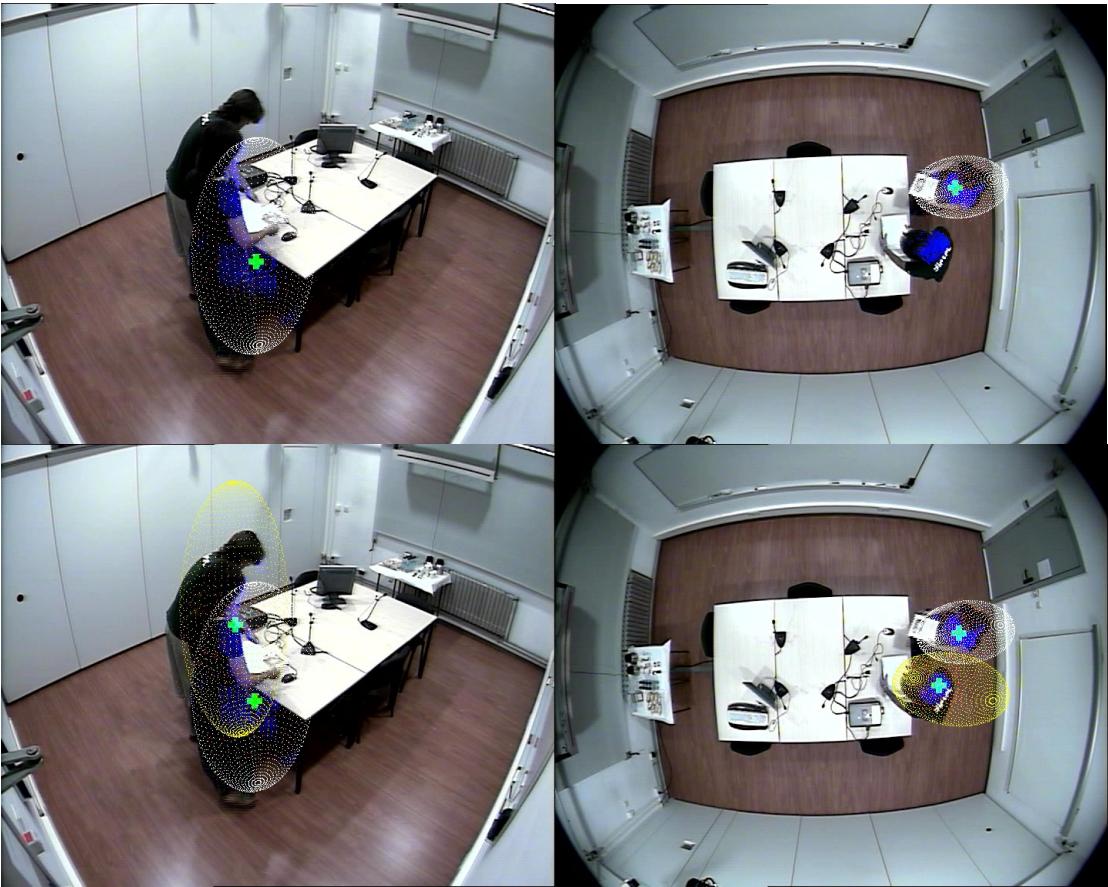


Fig. 3.7 En azul se muestran los volúmenes reconstruidos. Las elipsoides señalan las zonas donde trabaja cada filtro de partículas.

3.3.2 Filtrado del escenario

El análisis de conjuntos de voxels ha de hacerse asegurando la eliminación de posibles espurios. En el apartado 3.2.1 hemos definido que se entiende por espurio en el entorno de sala inteligente con reconstrucción por explotación de la redundancia información 2D. Los casos más graves y que más afectan al seguimiento son los causados por desplazamiento de objetos dentro de la sala. Al mover un objeto del mobiliario se genera una diferencia con el fondo o “background” que queda reflejado en la máscara binaria. Por lo tanto, aparecerá un volumen en la región que ocupaba dicho objeto y la nueva región que ocupa. El ejemplo más notorio de este fenómeno ocurre en las zonas de entrada de la sala. Al abrir y cerrar las puertas se generan volúmenes que no constituyen ningún tipo de información. Como señal de ruido que constituyen requieren ser filtrados. Por ello todo voxel que aparezca en un entorno de las puertas es considerado como objeto de fondo, para que no interfiera en el seguimiento.

A parte de las puertas en la sala hay sillas, portátiles, proyectores y objetos de esta índole que pueden dificultar el seguimiento. En la fig. 3.9 se ilustra un ejemplo de la influencia de estos objetos. Puede apreciarse como en la fila superior el filtro representado en color amarillo queda “enganchado” en un portátil que hay encima de la mesa, dado que el algoritmo considera que ese conjunto de voxels puede ser parte de la representación volumétrica de un blanco. Al mismo tiempo, al quedar el volumen de la persona que acaba de ser perdida por el algoritmo “atrae” a las partículas del filtro violeta. También se puede apreciar como el filtro verde marca la posición en la hay un sillón y no la persona que ha movido dicha silla. De todos modos, este último caso viene dado por un error en la

reconstrucción 3D ya que sólo aparece un volumen entorno a dicha silla. Dado que estos objetos no se hallan fijos en una posición concreta como las puertas, y que pueden estar conectados a volúmenes de personas, se ha empleado una apertura morfológica 3D.

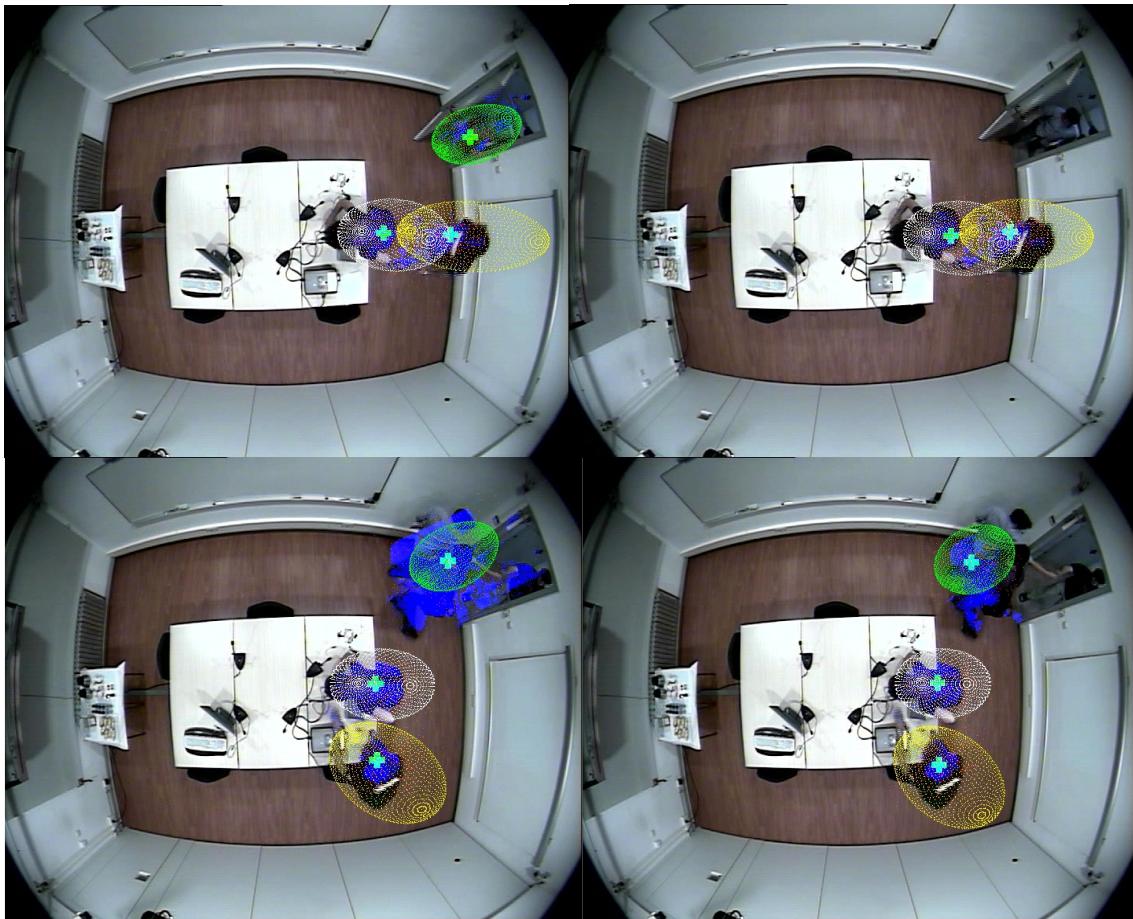


Fig. 3.8 Filtro en la zona de puertas. Los conjuntos de puntos en azul son los voxels de “foreground”. Las imágenes de la izquierda no emplean filtro mientras que las de la derecha sí lo utilizan. El ruido añadido por el movimiento de las puertas puede dar lugar a detecciones confusas.

Considerando que los objetos más molestos deberían tener una altura notablemente inferior a la de las personas, un elemento estructurante unitario en X e Y y con la altura suficiente en Z podría discriminar volúmenes correspondientes a objetos, tales como el portátil de la imagen recientemente mencionada. En la práctica, una apertura con elemento estructurante de una altura aproximada de unos 50 cm elimina prácticamente todos los volúmenes reconstruidos y esto es debido a que la reconstrucción no es suficientemente sólida, existen agujeros debidos a problemas de calibración y limitaciones del propio algoritmo de reconstrucción de la escena 3D ante occlusiones. Sin embargo, esta técnica, empleada con un elemento estructurante de un tamaño adecuado, permite reducir la solidez de volúmenes espurios. Así, se puede apreciar en las imágenes inferiores de la fig. 3.9, donde se aplica una apertura morfológica con un elemento estructurante de $1 \times 1 \times 5$ voxels, cada uno de tamaño $2 \times 2 \times 2$ cm, como el filtro sigue a la persona y no queda “enganchado” en el portátil. Existe otra diferencia importante entre los frames superiores e inferiores. En los primeros se ha producido un intercambio de objetivos (el filtro verde y el naranja están en objetivos distintos arriba y abajo), lo cual es uno de los problemas más frecuentes en el seguimiento de múltiples individuos.

La apertura también reduce ligeramente el tamaño y solidez de los volúmenes que representan a las personas, de modo que reduce la verosimilitud geométrica de aquellas regiones del espacio donde varios volúmenes se fusionan (por proximidad de las personas). Esto permite reducir la tasa de intercambios de objetivo redundando en un seguimiento más consistente. La desventaja es que está técnica, al distorsionar los volúmenes que también pueden ser considerados válidos, puede afectar a la precisión de la estimación de la posición de la persona. Aún teniendo en cuenta este problema, el algoritmo final ha sido evaluado con apertura morfológica, como veremos en el capítulo 4.

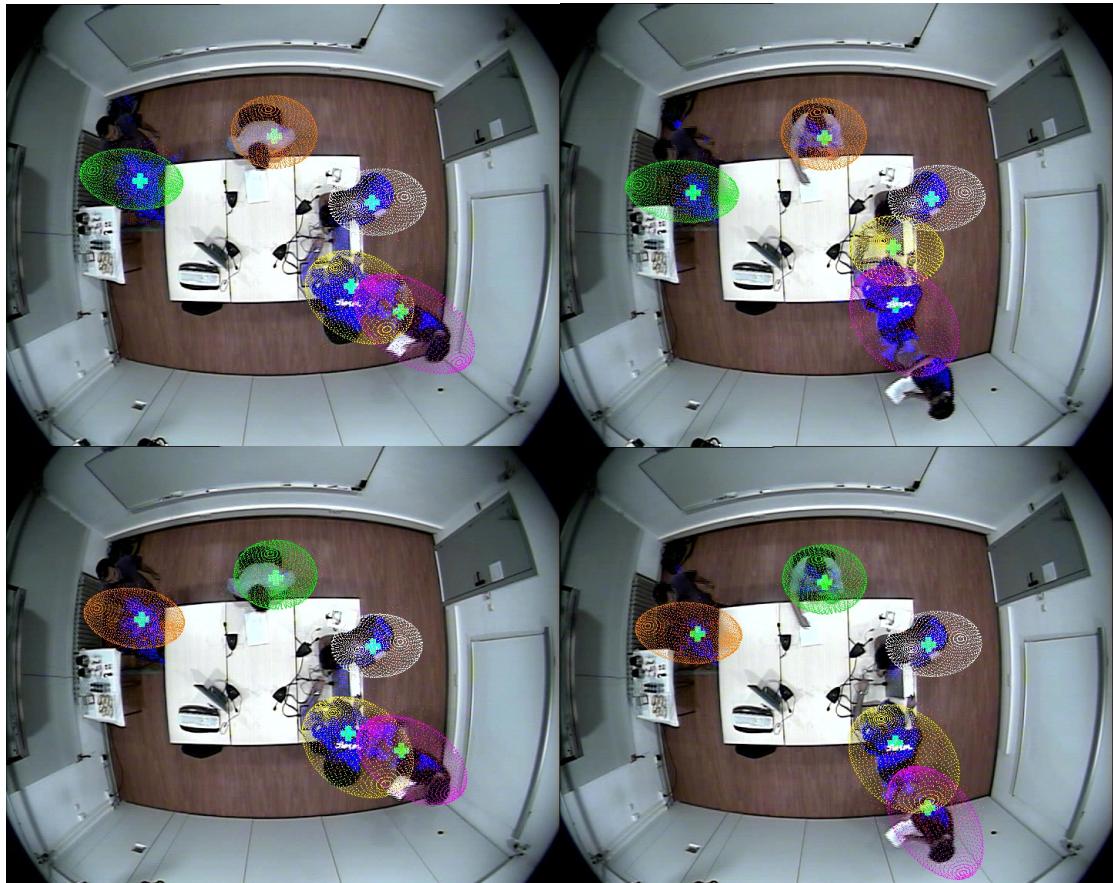


Fig. 3.9 La secuencia de arriba no emplea apertura morfológica mientras que la de abajo sí.

3.3.3 El filtro de partículas en el entorno CHIL

En las imágenes mostradas en este capítulo se ha podido apreciar el aspecto visual de los filtro de partículas empleados. El tipo de filtro escogido, dadas las razones expuestas en 2.3, es el SIR. Podemos dividir el algoritmo iterativo en tres pasos básicos:

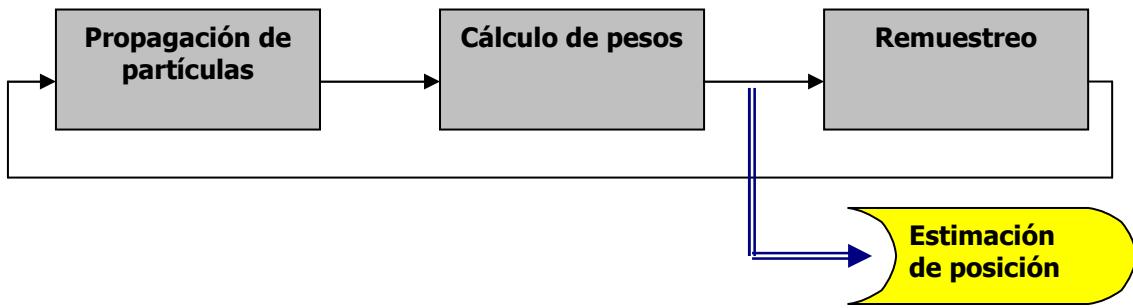


Fig. 3.10. Diagrama de bloques del filtro de partículas.

Como hemos avanzado en el apartado anterior, la salida que da el filtro es una hipótesis acerca del centroide de la persona seguida, que aproxima la media de la densidad a posteriori. Siendo w_k^i los pesos normalizados y teniendo en cuenta al filtro m -ésimo:

$$\mathbb{E}[p(x_k^m | z_k)] \approx X_k^m = \sum_{i=1}^{N_s} w_k^i x_k^i \quad (3.4)$$

En los siguientes apartados se detallarán los tres procedimientos, haciendo hincapié en los enfoques diferenciales respecto a otros trabajos [4, 8, 18].

3.3.3.1 Propagación de las partículas

Tal y como se ha visto en el capítulo 2, los algoritmos de seguimiento bayesianos no lineales en general, y los filtro de partículas en particular, emplean una estadística de transición o propagación para estimar la densidad a priori. En el caso concreto del filtro de partículas SIR la densidad de importancia $q(x_k|x_{k-1}, z_k)$, que se emplea para ubicar las partículas como paso previo al cálculo de pesos, pasa a ser la densidad de transición $p(x_k|x_{k-1})$. Dicho de otro modo, a partir de la estimación en el instante $k-1$, redibujamos las partículas según una densidad de probabilidad asociada a un modelo de propagación para posteriormente calcular sus pesos. Sin embargo, en el caso del seguimiento en una sala de tamaño reducido como la que disponemos en la UPC, las dinámicas de movimiento de las personas son muy limitadas, de modo que en el instante k el objetivo estará en un entorno cercano de la estimación en $k-1$. Por ello se ha considerado que no es necesario emplear un modelo de movimiento, siempre y cuando el remuestreo de partículas, que se efectúa en cada iteración, garantice la diversidad suficiente como para asegurar que se cubre un entorno en el que hallaremos el objetivo en la siguiente iteración. El no incluir un modelo de movimiento es poco habitual, aunque no es la primera propuesta de filtro de partículas que prescinde de un modelo definido [6]. La única reubicación que sufren las partículas consiste en la asignación de una única partícula por voxel. Al trabajar con granularidad de voxel este paso asegura mayor efectividad con el mismo número de partículas o ahorro de coste computacional para resultados equiparables en calidad. Un ejemplo de esta propagación la podemos ver en la fig. 3.11.

Existen ciertos instantes de tiempo en que la concentración de partículas es muy alta y no es posible conseguir una relación voxel-partícula única en todos los casos. Este fenómeno ocurre durante un tiempo limitado tras la inicialización del filtro de partículas con número grande de las mismas. Pasados pocos fotogramas, la diversificación de partículas es

suficiente como para poder reubicarlas en voxels diferentes. Podemos decir que tras el instante de inicialización hay un transitorio en el que algunas partículas caen en voxels ocupados por otras, y por lo tanto aportan información redundante al seguimiento.

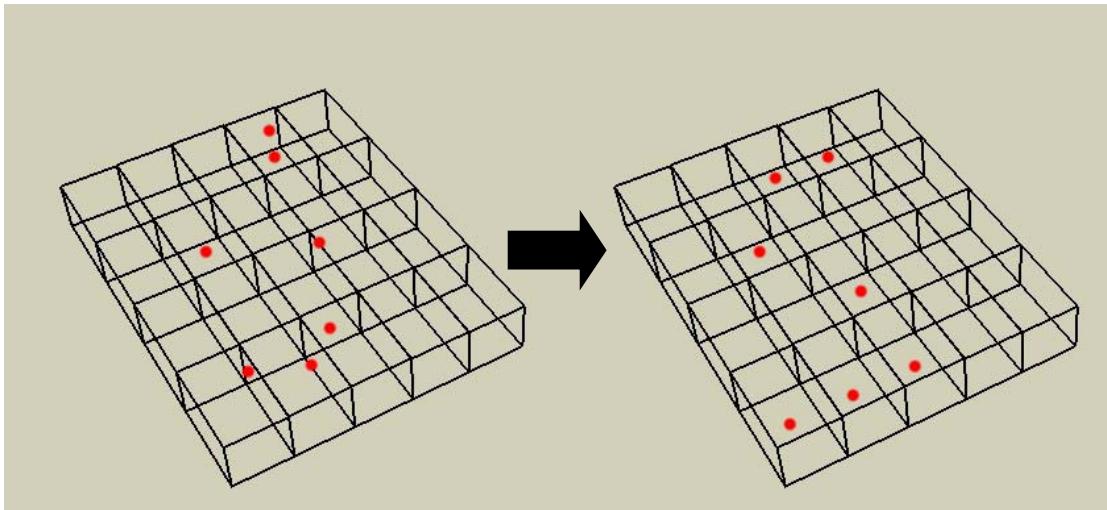


Fig. 3.11 Tras el muestreo, varias partículas (puntos rojos) pueden ocupar un mismo voxel. El primer paso del filtro consiste en redibujar las partículas como se muestra.

3.3.3.2 Cálculo de pesos

Una vez tenemos una relación unívoca entre partículas y voxels se procede al cálculo de pesos. Estos pesos reflejan qué grado de verosimilitud presenta una región del espacio de formar parte del volumen de una persona y más concretamente, del individuo que se estaba siguiendo en $k-1$. Por ello la función $p(z_k|x_k)$ se obtiene bajo el análisis conjunto de la información geométrica de la región espacial bajo estudio y de las posibles interacciones que pueda haber en ese instante k . A continuación presentaremos de qué modo obtenemos y explotamos ambas informaciones.

La verosimilitud de una partícula x_k^i que tienen asociada un voxel (por la propagación expuesta en 3.3.3.1) depende si ese voxel pertenece a un volumen que representa a una persona. La primera y más sencilla aproximación para obtener $p(z_k|x_k^i)$ es asignar un valor positivo a w_k^i si el voxel es de “foreground” o 0 si es de “background”. De manera más formal queda como sigue en la expresión 3.5, donde w_k^i es el peso asociado a la partícula i -ésima en el instante k , N_f es el número de partículas del filtro que caen en voxels de “foreground” y que sirve para normalizar los pesos, v es el voxel asociado a la partícula, Φ es el conjunto de voxels de “foreground” y β el de voxels de “background”.

$$w_k^i \propto \begin{cases} \frac{1}{N_f}; v \in \Phi \\ 0; v \in \beta \end{cases} \quad (3.5)$$

Es fácil determinar que esta función de verosimilitud se verá fuertemente afectada por el ruido de reconstrucción de la escena 3D. Una partícula aislada que caiga en un voxel que, por cambios de iluminación u otro tipo de errores, sea considerado de *foreground* tendrá peso distinto de 0 mientras que una partícula que, pese a estar dentro de la región ocupada por un volumen, caiga en un voxel a 0 no tendrá peso y no será remuestreada. Resulta

evidente que la estrategia más adecuada es utilizar un cálculo basado en medias de los valores de un conjunto de voxels de un entorno de la partícula. Desde el punto de vista de procesado de señal, los errores de reconstrucción sumados al efecto de las herramientas morfológicas comentadas en 3.3.2, añaden ruido a la escena. Puesto que contribuyen a reducir la solidez de los volúmenes y, por analogía, la solidez de las regiones de *foreground*, el ruido introducido es de alta frecuencia, ya que podemos hallar transiciones de 0 a 1 y de 1 a 0 en varias direcciones del espacio. La media, al ser un filtro paso bajo permite combatir este ruido. Esto lleva a formular los pesos del siguiente modo:

$$w_k^i \propto \frac{1}{|C(x_k^i, q)|} \sum_{v \in C(x_k^i, q)} r(v); \quad \text{donde } r(v) = \begin{cases} 1; & v \in \Phi \\ 0; & v \in \beta \end{cases} \quad (3.6)$$

En esta nueva expresión $C(\cdot)$ es una vecindad de la partícula en el dominio de la conectividad q , mientras que $|C(\cdot)|$ indica la cardinalidad de este conjunto. Esta medida de la verosimilitud es más precisa y más robusta ante ruido de reconstrucción. Además, podemos añadir la distancia a los voxels de “*foreground*”:

$$w_k^i \propto \frac{1}{|C(x_k^i, q)|} \sum_{v \in C(x_k^i, q)} r(v) d(x_k^i, v); \quad (3.7)$$

Sin embargo puede observarse mediante experimentos que esta medida de verosimilitud tiende a crear puntos de atracción de las partículas penalizando distribuciones en el espacio que darían más calidad al seguimiento. Estos puntos de atracción surgen en las regiones donde las reconstrucciones presentan mayor solidez. Si la reconstrucción es suficientemente uniforme en cuanto a densidad de voxels de “*foreground*” este criterio da buenos resultados. Sin embargo, en caso de no poder asegurar dicha uniformidad, se puede alterar la media y emplear otro criterio que asegure la dispersión de las partículas.

La siguiente opción estudiada es concentrar el peso en la superficie del volumen. Para ello podríamos calcular derivadas 3D con herramientas de procesado de imagen y obtener una superficie de voxels. Luego simplemente aplicaríamos el criterio de media de la vecindad de partícula y ya dispondríamos de un criterio que diversifica más las posiciones de las muestras. El problema de esta estrategia es, como anteriormente, que la carencia de solidez del volumen reconstruido no permite reducirlos a superficies mediante derivadas.

Modificando la media empleada hasta ahora podemos hallar una aproximación al criterio de superficie. Dado que la media analiza un entorno de la partícula, simplemente podemos definir en qué medida dicho entorno se asemeja a una posible superficie. Idealmente, una partícula ubicada en un voxel de superficie tiene aproximadamente tantos voxels de “*foreground*” como de “*background*” a su alrededor. Por lo tanto, se trata de establecer qué proporciones de voxels de ambos tipos deben aparecer en el entorno de una partícula y en qué medida representan a la superficie del volumen.



Fig. 3.12 Imágenes correspondientes a 80 iteraciones del algoritmo. En la imagen superior se emplea un criterio de media de voxels, mientras que en la inferior se utiliza la estrategia de asimilación de la superficie.

Podemos considerar los ejemplos de la Fig 3.13, representados en dos dimensiones, de posibles situaciones en las que una partícula puede estar en una superficie del volumen. Tanto en el caso a) como en el c) la partícula cae dentro de un voxel de “foreground”, pero las proporciones entre ambos tipos de voxel, para el entorno dado, son distintas. En el caso b) la partícula cae en un voxel de “background” sin embargo podemos considerar que presenta cierta verosimilitud. El último caso ilustra la posible falta de solidez en la reconstrucción. Las proporciones de voxels a 0 y a 1 (o lo que es lo mismo, de “background” y “foreground”) son, respectivamente, 3-6, 6-3, 5-4 y 4-5. De esta simple observación se extraen dos conclusiones. La primera es que la proporción que resume la semejanza a un entorno de superficie es aproximadamente la paridad entre ambos tipos de voxel, tal y como dejábamos ver antes del ejemplo. La segunda es que la medida es imprecisa, dado que no analiza las posiciones de dichos voxels en el entorno. El peso del

caso d) y el complementario de c) sería el mismo, pero la situación es totalmente distinta. A pesar de ello, los experimentos realizados con este criterio dan buenos resultados debido a la carencia de solidez en el interior de los volúmenes, lo cual permite repartir aún más las partículas.

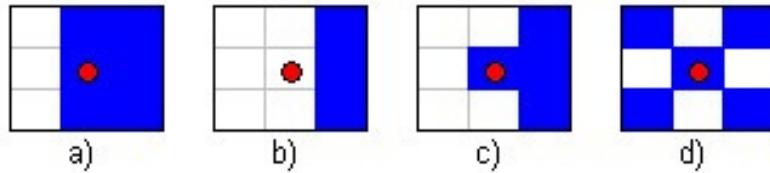


Fig. 3.13 Análisis de cuatro entornos de partícula para el cálculo de su peso. En azul se marcan los voxels de “foreground”.

La expresión 3.8 resume formalmente la estrategia de asimilación del entorno a la superficie sin tener en cuenta la distribución de los voxels. A parte de las expresiones introducidas en 3.6, se utiliza una función f cuya variable independiente es el número de voxels de “foreground” en la vecindad analizada. Esta función es simplemente una medida de conformación de la curva de peso, resultante de calcular el complementario de la diferencia normalizada entre el número de voxels a cero y a uno. f es 1 por defecto, de modo que la curva no se modifica. Si f es proporcional al número de voxels de “foreground” se favorecerá la mayor presencia de este tipo de elementos. Como ejemplo, con la función a 1 para cualquier valor, el caso a) y b) de la figura 3.13 tendrán el mismo peso (0,6667). Si empleamos una conformación lineal que vaya de 0 a 1, los nuevos pesos serían distintos ($a=0,44444$ y $b=0,22222$).

$$w_k^i \propto f(|C(x_k^i, q) \in \Phi|) \left[1 - \left[\frac{1}{|C(x_k^i, q)|} \left(\sum_{v \in C(x_k^i, q)} r(v) - \left(|C(x_k^i, q)| - \sum_{v \in C(x_k^i, q)} r(v) \right) \right) \right] \right] \quad (3.8)$$

Otro método de conformación que puede añadirse a la expresión en 3.8 es una función cuya variable independiente sea la distancia al centroide estimado. Del mismo modo que para la proporción de voxels de *foreground* y *background*, podemos definir funciones inversamente proporcionales a la distancia voxel-centroide. Estas funciones reducirían errores en situaciones que las partículas muy alejadas del centroide estimado pueden ser consideradas verosímiles debido a objetos espurios. De este modo, al llegar al paso de remuestreo, dichas partículas con tendencia a añadir error a la posición estimada se reproducirían en menor número o directamente no sobrevivirían.

Con estos criterios unidos a la previa propagación y al remuestreo se puede efectuar un seguimiento individual consistente y robusto a ruido de reconstrucción y a occlusiones causadas por objetos de fondo. Sin embargo el seguimiento de múltiples personas tiene otro nivel de complejidad. Los seres humanos se relacionan e interaccionan entre ellos de varias maneras. Al nivel en el que trabaja el seguimiento nos interesa un subconjunto de dichas interacciones, que son las que ocurren entre varios individuos cuando hay contacto físico o mucha proximidad. Esta cercanía física se traduce en la aparición de un único volumen conexo en intervalos determinados de tiempo, donde anteriormente había

múltiples objetivos. Obviamente, las herramientas presentadas hasta ahora no permiten combatir estos casos, ya que desde el punto de vista de la geometría de los volúmenes los múltiples objetivos fusionados son igualmente verosímiles en general, y puede ocurrir que alguna región de esta fusiónatraiga a varios filtros de partículas causando errores en el seguimiento y la pérdida de los objetivos. Por ello, existe otro nivel de verosimilitud dentro del cálculo de pesos.

Las variantes presentadas sólo estudian los volúmenes. Para el algoritmo se ha estudiado también el color como característica útil para el seguimiento, pero los resultados no han permitido incluirlo en el diseño debido a los problemas en la extracción de dicha característica. Sin embargo se ha sopesado la aportación del color en caso de poder extraerlo de forma fiable (ver Anexo).

3.3.3.3 Seguimiento de múltiples personas y bloqueo de partículas.

Tal y como se explica en 2.3.3, la aproximación más adecuada al filtro de partículas para múltiples objetivos en términos de coste computacional y complejidad es la utilización de un filtro de partículas para cada objetivo. En definitiva, tratamos un sistema de estados que caracteriza la dinámica de m personas como m sistemas modelando una única persona. Habíamos mencionado que el inconveniente de este tratamiento es que se asume independencia entre los distintos objetivos, lo cual, al existir interacciones, no es cierto, porque un filtro de partículas puede considerar voxels de otro objetivo distinto al que seguía en los instantes anteriores. Este problema se reduce al plano XY, dado que se considera que las personas que hay en la sala siempre están de pie o sentadas, no tumbadas ni unas encima de las otras. Esto permite asumir que dos objetivos no pueden ocupar el mismo espacio en este plano, gracias a lo cual se pueden diseñar métodos de penalización de partículas que pretendan ocupar el espacio de otro objetivo. Por otra parte, no podrá haber varios filtros de partículas ocupando el mismo espacio, por lo que se eliminan aquellos que tengan un cierto grado de solapamiento con otros.



Fig. 3.14 Zonas sobre las que se remuestrean las partículas de los filtros de partículas.

Considérese que se dispone de M filtros de partículas, donde M es además el número de personas existentes en la sala. Con el enfoque presentado hasta ahora, dos filtros próximos, m_1 y m_2 , tras el paso de remuestreo pueden considerar verosímiles partículas que caen en el volumen vecino. Como se puede apreciar en la imagen fig. 3.14, tras el paso de remuestreo las partículas se distribuyen en las zonas roja (m_1) y verde (m_2). Esto podría producir un intercambio de filtros de partículas. Para evitar esta situación se deben

bloquear aquellas partículas que caigan en la zona de influencia de un filtro cercano. Mediante este **bloqueo** modelamos las interacciones en la sala. Existen propuestas en la literatura de filtros de partículas [16, 2, 5], de modo que el presente trabajo extiende estas posibilidades a 3D. La implementación de este sistema de bloqueo se basa en asociar a cada filtro una elipsoide que parte del centroide X_{k-1} . El semieje Z se obtiene de la respectiva coordenada del centroide estimado mientras que los ejes X e Y son fijos para que no se propague el error de estimación. Cualquier partícula de otro filtro que caiga dentro de este elipsoide se penaliza mediante una función. Esta exclusión se refleja en el cálculo de pesos del siguiente modo:

$$w_k^i \propto p(z_k | x_k^i) \prod_{\substack{m=1 \\ m \neq j}}^M \eta(\hat{X}_{k-1}^m, x_k^i) \quad (3.9)$$

Donde el índice j indica el filtro de partículas al que pertenece la partícula x_k^i , η es la función de penalización y \hat{X}_{k-1}^m es el estado del filtro m -ésimo en el instante $k-1$, dado que el centroide no se acaba de calcular hasta tener todos los pesos de las partículas.

En las imágenes que se muestran a continuación (3.15) podemos observar la mejoría que presenta la estrategia de bloqueo. Al no utilizar bloqueo las partículas de ambos filtros ocupan posiciones en el volumen del otro. Esto hace que los centros de masas resultantes de la estimación se atraigan hasta que ambos ocupan el mismo lugar. El algoritmo elimina el segundo filtro porque considera que ambos siguen al mismo objetivo. Con el bloqueo ambos filtros se mantienen en sus posiciones.

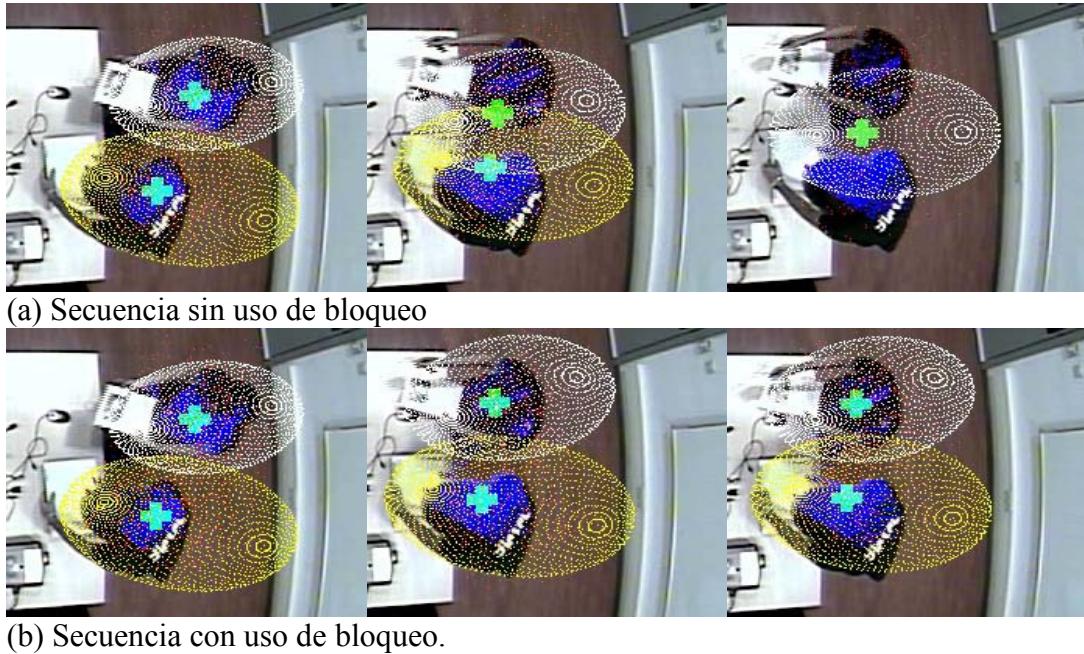


Fig. 3.15 Vista cenital de dos ejecuciones con 600 partículas y arista de voxel de 2 cm.

3.3.3.4 Remuestreo

Como ya introdujimos en 2.3.2, el remuestreo se efectúa en cada iteración para asegurar una cierta diversificación espacial de las partículas a lo largo del tiempo. Dado que no aplicamos ningún modelo de movimiento en la fase de propagación, el remuestreo ha de garantizar que el conjunto de partículas cubra entornos cuya probabilidad de futura ocupación por parte del objetivo sea alta, o dicho de otro modo, efectuar el remuestreo alrededor de las partículas con más peso. En primer lugar se calcula cuantas muestras nuevas generará cada partícula. Para ello simplemente se efectúa el siguiente cómputo:

$$\left| x_k^{(i)} \right| = w_k^i N_s \quad (3.10)$$

Donde $|x_k^{(i)}|$ denota el número de nuevas partículas asociadas a la actual partícula x_k^i , w_k^i son los pesos asociados a ésta y N_s el número total de muestras empleadas.

A continuación se generan ese número de partículas del siguiente modo:

$$x_k^j = x_k^i + U\left[-\frac{\Delta x}{2}, \frac{\Delta x}{2}\right]\vec{i} + U\left[-\frac{\Delta y}{2}, \frac{\Delta y}{2}\right]\vec{j} + U\left[-\frac{\Delta z}{2}, \frac{\Delta z}{2}\right]\vec{k} \quad (3.11)$$

Dado que cada partícula es una coordenada 3D, lo que se hace es generar una nueva muestra desplazando x_k^i una distancia aleatoria obtenida a partir de distribuciones uniformes en X, Y, Z. Sea r una coordenada cualquiera, N_s el número de partículas empleado y h_v el tamaño de la arista del voxel en cm. Un valor, en cm, para el incremento de distancia empleado es:

$$\Delta r = 3.9 \sqrt{h_v} \sqrt[3]{N_s} e^{-\frac{N_s h_v}{20000}} \quad (3.12)$$

El número de partículas que aplica en 3.12 es variable, y debe considerarse que el problema de seguimiento admite un límite en el número de partículas utilizadas por encima del cual hay un sobredimensionamiento. Por la sencillez del modelo de persona que emplea el algoritmo y el hecho de que los volúmenes en los que existe conectividad suelen tener entre 500 y 20.000 voxels (raramente sobrepasan los 10.000), puede determinarse que, en general, usar más de 2000 partículas supone tener más muestras de las necesarias. Los resultados experimentales corroboran este hecho en términos de efectividad estadística del seguimiento. Los gráficos que hay a continuación muestran la curva de valores para tamaños de voxel típicos de 2, 3 y 5 cm. Se puede observar como cuánto más grande es el voxel antes tiende a colapsarse la distancia. Asimismo, el número de partículas a emplear se reduce en los casos en que el voxel es mayor.

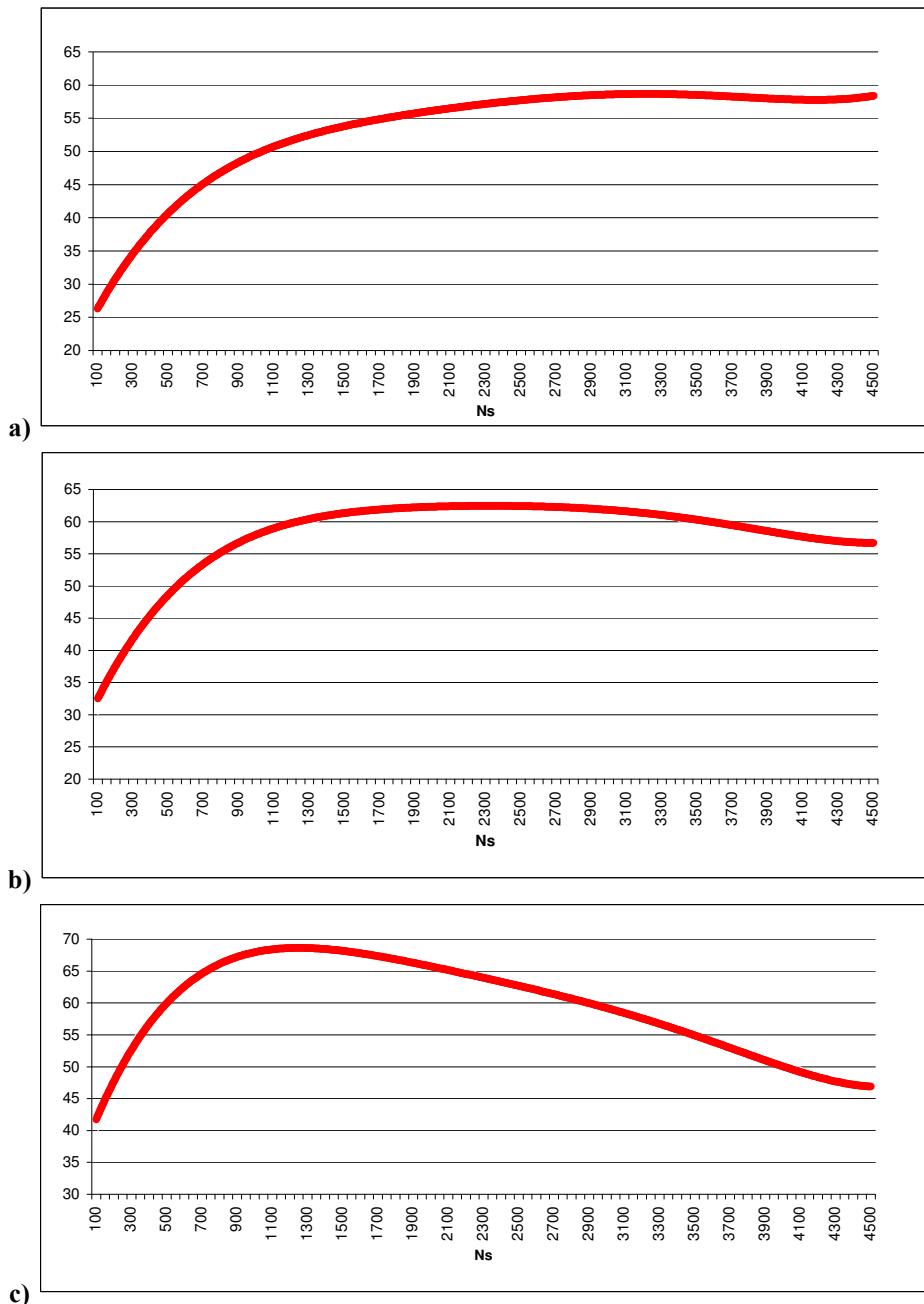


Fig. 3.16 Gráficos de distancia de remuestreo en función del número de partículas y el tamaño de voxel.
a) 2cm. b) 3cm c) 5cm.

Otra relación empleada tiende a distribuir las partículas con un criterio opuesto. A menor número de partículas mayor es la distancia de remuestreo, para que el espacio ocupado sea equivalente en todos los casos. Sea $\max(r)$ el valor máximo que puede alcanzar una distancia en un eje, dicho de otro modo, el tamaño de la sala en ese eje. El nuevo incremento de distancia es:

$$\Delta r = \frac{\max(r)}{\sqrt[3]{N_s}} \quad (3.13)$$

En este caso, tenemos una distancia por eje. Además, se permite el doble de distancia para el eje Z, ya que interesa distribuir las partículas en la zona donde se encuentra el volumen. En la fig. 3.17 podemos ver un ejemplo concreto de valores de distancia, para todos los tamaños de voxel .

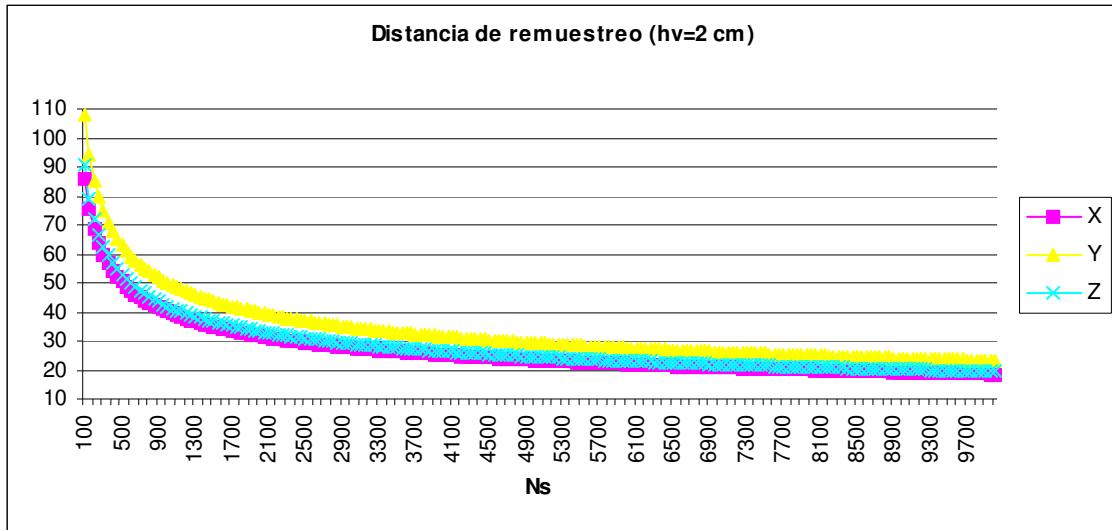


Fig. 3.17 Distancia de remuestreo en cm para los distintos ejes. Se toman como medidas de sala 400x500x210 cm.

Para los valores típicos en los que nos movemos a la hora de efectuar experimentos (200-1000 partículas), este criterio funciona notablemente bien. En los resultados experimentales que se presentan en el capítulo 4 se podrá ver una comparativa para ver que aportan ambos criterios.

4 Resultados y evaluación

El presente capítulo está dedicado a la evaluación de los resultados obtenidos por la implementación del algoritmo presentado en el capítulo 3.

Las ejecuciones del algoritmo con diferentes juegos de parámetros han sido llevadas a cabo sobre seminarios que han tenido lugar en la sala CHIL de la UPC. No se han efectuado ejecuciones en tiempo real si no a partir de los datos ya procesados obtenidos de la secuencia de video resultante del seminario. En primer lugar se emplean los fotogramas obtenidos de las 5 cámaras de la sala, las 4 situadas en las esquinas y la cenital (ver apartado 3.1), almacenados con formato JPEG de tamaño 720x576. También es necesario el uso del conjunto de volúmenes reconstruidos a partir de dichos fotogramas. Finalmente se emplean los ficheros de calibración de las cámaras para representar coordenadas 3D en los fotogramas, permitiendo tener un formato de evaluación puramente visual. Todos estos son los datos de entrada del algoritmo y es a partir de ellos que se estiman las posiciones de las múltiples personas en la sala.

Previamente a la presentación de los resultados, se introduce un conjunto de métricas para la evaluación estadística de un sistema de seguimiento de múltiples objetivos que permita comparar cuantitativamente la calidad que ofrece el algoritmo propuesto en sus distintas versiones y también frente a otros trabajos.

4.1 Métricas

Paralelamente a la proliferación de proyectos de interpretación de lenguaje humano que incorporan, entre otros, sistemas de seguimiento de múltiples personas, reconocimiento del habla o detección de gestos, ha crecido la necesidad de emplear un sistema que mida de modo cuantitativo la calidad de las estimaciones. Con este fin se diseñaron las evaluaciones CLEAR (*Classification of Events, Activities and Relationships*) [9, 10]. En ellas se definen un protocolo a seguir y unos parámetros a aplicar con el fin de llegar a una estandarización que permita tener resultados fácilmente comparables. Dentro de CLEAR se definen métricas que resumen el conjunto de errores que se dan típicamente en un sistema de seguimiento múltiple. Los requisitos para la definición de las métricas son básicamente las siguientes:

- 1) Capacidad para evaluar la precisión de la estimación de posición en términos de distancia.
- 2) Capacidad para evaluar la consistencia del seguimiento, definida como la medida en qué un algoritmo traza correctamente la trayectoria de cada objetivo.
- 3) Facilidad de uso y de interpretación de los resultados.
- 4) Uso del mínimo número posible de parámetros ajustables para poder obtener medidas comparables.
- 5) Capacidad de evaluar diferentes tipos de algoritmo (2D, 3D, puntuales, zonales, etc.)

A partir de estos criterios y teniendo en cuenta que en cada instante k el algoritmo produce una serie de hipótesis $\{h_1, \dots, h_M\}$ acerca de la ubicación de los objetivos presentes en la sala $\{o_1, \dots, o_N\}$, el sistema de evaluación efectúa los siguientes pasos:

- 1) Establece la mejor correspondencia entre las hipótesis y cada uno de los objetivos.
- 2) Para cada par h_m-o_n , calcula el error de estimación de la posición (o distancia entre h_m y o_n).
- 3) Establece los errores en las correspondencias entre hipótesis y objetivos, tales como desigualdad entre M y N o intercambio de pares.

Dada la distinta naturaleza de los dos últimos pasos, se definieron dos métricas, MOTP (*Multiple Object Tracking Precision*) y MOTA (*Multiple Object Tracking Accuracy*) [12]. El MOTP da el valor en media del error calculado en el paso 2) y el MOTA calcula la consistencia en porcentaje a partir de la contabilización de errores del paso 3).

Previamente a analizar en detalle ambas métricas por separado, es necesario definir una serie de eventos de los que derivará el valor de las métricas: intercambio, falsa detección y pérdida. El intercambio (*missmatch*) consiste en el cruce de hipótesis entre objetivos. La falsa detección (*false positive*) tiene lugar cuando hay una hipótesis que no se corresponde con ningún objetivo. La pérdida (*miss*) es el evento complementario a la falsa detección y tiene lugar cuando un objetivo no tiene hipótesis asignada.

El procedimiento de evaluación se inicia estableciendo la mejor correspondencia entre pares de objetivos e hipótesis. Para ello se define previamente un umbral de distancia, D_{th} , por encima del cuál no se establece correspondencia y se clasifica a la hipótesis como no válida. Este umbral determina una distancia para la cual no podemos referirnos a error de precisión sino a que el algoritmo está siguiendo otra cosa. Para cada instante k se empieza asignando el par o_n-h_m con mínima distancia mutua, y así sucesivamente para el resto de objetivos e hipótesis. Si el conjunto de asignaciones contradice al existente en $k-1$, entonces se contabiliza un intercambio. A continuación se calcula la distancia $d_{nm,k}$ de cada correspondencia existente. El número de hipótesis restantes son contabilizadas como falsas detecciones y el número de objetivos sin hipótesis como pérdidas. Este procedimiento se repite fotograma tras fotograma. En el instante inicial no hay un conjunto de correspondencias previo, por lo que se asignan pares pero no hay posibilidad de que haya intercambio. La fig. 4.1 muestra ejemplos de eventos en el seguimiento y su evaluación.

Las métricas MOTP y MOTA recogen de modo formal los eventos y magnitudes citadas en el procedimiento de evaluación. El MOTP (*Multiple Object Track Precision*) evalúa el algoritmo en términos de precisión de la posición estimada independientemente de la consistencia temporal de la asignación de identidades. La expresión que la define es la siguiente:

$$MOTP = \frac{\sum_k \left(\sum_n^N d_{nm,k} \right)}{\sum_k c_k} \quad (4.1)$$

donde c_k es el número de correspondencias halladas en el instante k y $d_{nm,k}$ es la distancia entre el objetivo n y su hipótesis m asignada en el instante k -ésimo.

El MOTA (*Multiple Object Track Accuracy*) mide la calidad estadística de las asignaciones de identidad efectuadas por el algoritmo o lo que es lo mismo, la consistencia de la estimación. Sea mme_k , fp_k y m_k intercambios, falsas detecciones y pérdidas respectivamente. Definiremos T_k como el número de objetivos presentes en el instante k . La expresión formal del MOTA es:

$$MOTA = \frac{1 - \sum_k (mme_k + fp_k + m_k)}{\sum_k T_k} \quad (4.2)$$

El MOTA puede ser desglosado en las tres ratios de error de intercambios, falsas detecciones y pérdidas:

$$\frac{mme_k}{mme_k} = \frac{\sum_k (mme_k)}{\sum_k T_k} \quad (4.3)$$

$$\frac{fp_k}{fp_k} = \frac{\sum_k (fp_k)}{\sum_k T_k} \quad (4.4)$$

$$\frac{m_k}{m_k} = \frac{\sum_k (m_k)}{\sum_k T_k} \quad (4.5)$$

Este desglose será útil a la hora de determinar qué problemas o errores típicos presenta un algoritmo de seguimiento.

La razón por la que no se hace una media de error independiente por instante de tiempo en vez de una media global es que de este modo la evaluación mide de forma más intuitiva el comportamiento. En la fig. 4.2 se ha ilustrado un ejemplo hipotético que demuestra esta afirmación.

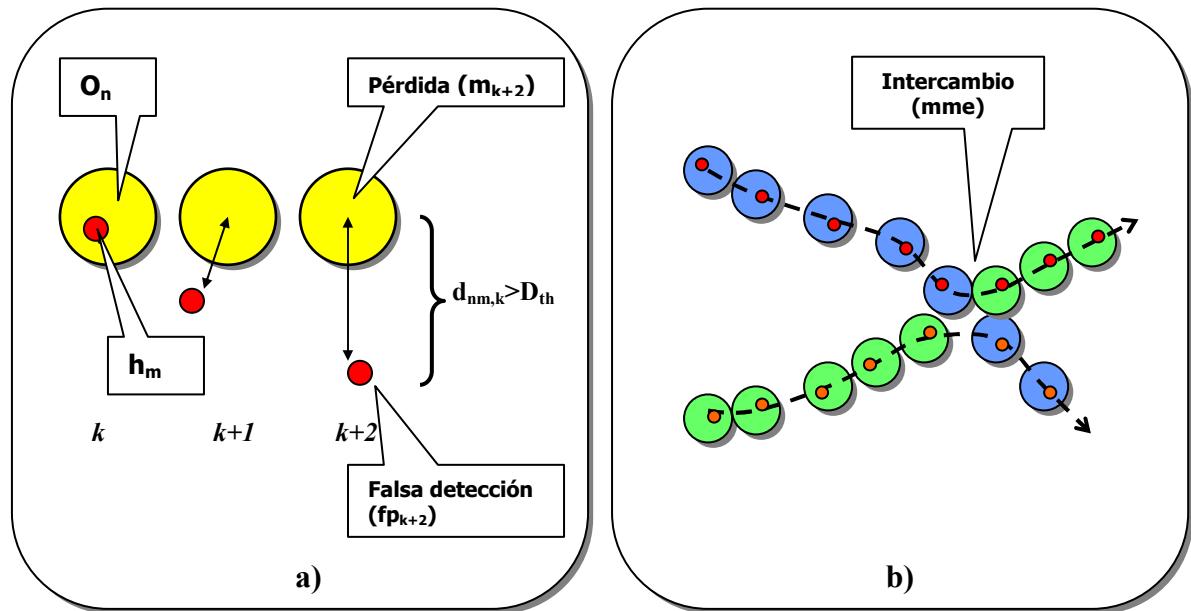


Fig. 4.1 Casuísticas de seguimiento y evaluación de las mismas. a) La hipótesis generada se aleja del objeto al que sigue. En el fotograma $k+1$ supone una penalización en el MOTP, mientras que en el $k+2$ se produce una pérdida asociada al objeto y una falsa detección asociada a la hipótesis, penalizando el MOTA.. b) Intercambio entre hipótesis y objetivos. La línea discontinua muestra la trayectoria definida por cada una de las hipótesis (círculos rojo y naranja). Nótese que al hacer las asignaciones objetivo-hipótesis según un criterio de mínima distancia, sólo se producen dos intercambios (uno por cada hipótesis) mientras que si las correspondencias se efectuasen teniendo en cuenta los pares que se han dado más veces contaría 5 intercambios (3 asociadas a la hipótesis roja y 2 a la naranja).

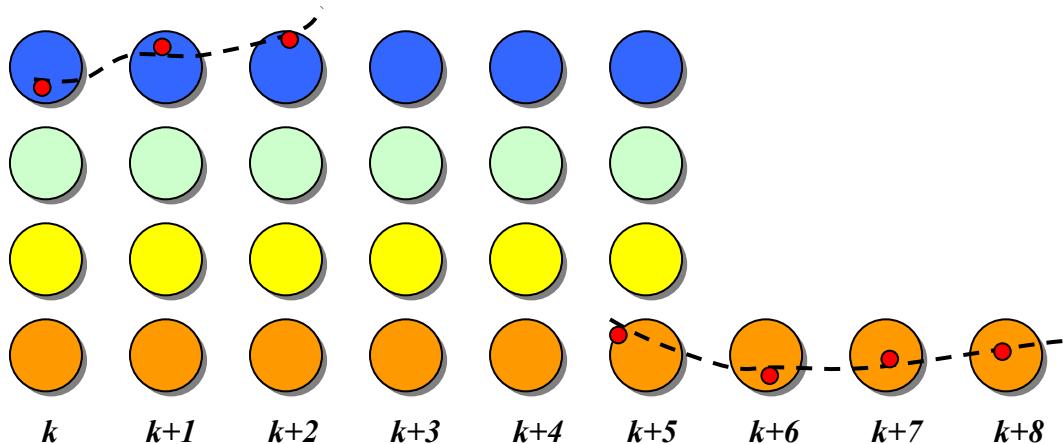


Fig. 4.2 Si contabilizamos pérdidas fotograma a fotograma obtenemos 75% en los tres primeros y en $k+5$, 100% en $k+3$ y $k+4$ y 0% en los tres últimos. La tasa de error cometido por pérdidas para la secuencia, según este cómputo, es 56%. Si hacemos el cálculo como se propone, tenemos 20 pérdidas sobre 27 objetivos presentes, lo que da una tasa de error por pérdidas de 74,1%. Una tasa de error de 56% no refleja el mal comportamiento que ofrece visualmente el seguimiento ilustrado.

4.2 Algoritmo de seguimiento basado en Kalman

Antes de obtener resultados cuantificados para el diseño propuesto en este trabajo, se ha evaluado un algoritmo diseñado en la UPC basado en el filtro de Kalman [15].

Este algoritmo parte del mismo punto que el propuesto en este proyecto. Efectúa una reconstrucción de la escena 3D mediante la redundancia de las imágenes ofrecidas por las múltiples cámaras de la sal CHIL. A continuación se efectúa un análisis de conectividad 26 (se considera cualquier contacto entre aristas, caras o vértices) entre voxels y se considera que modelan personas si la cardinalidad de los volúmenes conexos supera un cierto umbral.

A partir de la información volumétrica, se obtiene caracterización mediante color, basada, por ejemplo, en tono dominante, histograma o histograma en función de la altura. Se emplea para ello una sencilla técnica de coloreado de voxels de reducida carga computacional. Cada uno de los volúmenes conexos que modelan personas se identifica por su centroide. Para cada voxel del volumen se analiza si su distancia a una cámara determinada es menor que la del centroide de ese volumen a la misma cámara. Si es así, se examina que no exista otro centroide distancia menor a esa cámara. Esta medida sirve para descartar que haya otro posible objeto que ocluya al voxel bajo estudio. Si se determina que no lo hay entonces se colorea el voxel proyectándolo sobre la cámara. Si hay un centroide con distancia menor a la cámara entonces se analiza la distancia entre el segmento Voxel-cámara y el centroide ajeno comparándola con un umbral que sirve para determinar si se colorea el voxel o no.

Tras la extracción de las características mencionadas, se emplea un conjunto de parámetros para efectuar el seguimiento: la velocidad del centroide, el volumen y el histograma de color. Así, en el instante k se dispone de un volumen centrado en (x_c, y_c, z_c) con velocidad v_c e histograma H_c . Para el fotograma $k+1$ se efectúa la extracción de características presentada, obteniendo un conjunto de posibles volúmenes candidatos. Se emplea el filtro de Kalman sobre cada volumen en k para actualizar la posición y la velocidad en su centroide. De esto modo obtenemos un conjunto de hipótesis actualizadas con las que hacer una asignación con los posibles candidatos. Para dicha asignación puede ser empleada la distancia Euclídea entre candidatos, lo que constituye un recurso geométrico, y una distancia de Mahanalobis o de Bhattacharyya, para los histogramas hallados. Dado que la medida de color es altamente vulnerable a ruido y que el cómputo en espacios de dimensiones elevadas se complica mucho, el algoritmo emplea únicamente una versión ponderada de la distancia Euclídea.

Los resultados obtenidos con este algoritmo en seguimiento visual multipersona son los siguientes:

MOTP	\bar{m}	\bar{fp}	\bar{mme}	MOTA
195 mm	21.24%	46.16%	4.22%	28.35%

Tabla 4.1. Resultados de la evaluación del algoritmo basado en filtro de Kalman.

La tasa de error de falsas detecciones es muy alta debido, principalmente, a la detección de numerosos objetos espurios. El aporte de error debido al uso del filtro de Kalman y las reglas de asociación como método de predicción del movimiento de personas es más generalizado en términos de tasa de error.

4.3 Algoritmo basado en filtro de partículas

En el capítulo 3 se han detallado una serie de parámetros mediante el ajuste de los cuales se puede conseguir mejorar la calidad final del algoritmo diseñado. Para nuestros experimentos hemos empleado la parametrización siguiente:

- Se emplea apertura morfológica con un elemento estructurante de 1x1x5 voxels.
- El método de clasificación de volúmenes conexos que modelan personas se hace según lo indicado en 3.3.1.
- Para reducir el coste computacional el algoritmo de asignación de partículas a voxels se efectúa una búsqueda en el entorno de la partícula y se asigna al primer voxel vacío, en vez de efectuar un ranking de voxels en función de si estos están a 1 o a 0 y de la distancia que presentan a la partícula.
- Empleo del método de asimilación de superficie para el cálculo de pesos sin ninguna función de conformación.

- $\eta(\hat{X}_{k-1}^m, x_k^i) = 0$ si $\frac{[(X_{k-1}^m - x_k^i)\vec{x}]^2}{(30)^2} + \frac{[(X_{k-1}^m - x_k^i)\vec{y}]^2}{(40)^2} + \frac{[(X_{k-1}^m - x_k^i)\vec{z}]^2}{(X_{k-1}^m \vec{z})^2} \leq 1, \forall m \neq i$.

Esto es, si la partícula cae en una elipsoide de semiejes $a=30$ cm, $b=40$ cm y c la coordenada z que devuelve el filtro entonces el peso de dicha partícula es 0.

- Para el remuestreo, $\Delta r = \frac{\max(r)}{\sqrt[3]{N_s}}$. En z se dobla esta distancia.

MOTP	Tamaño de voxel			
Partículas	2 cm	3 cm	4 cm	5 cm
50	222	226	223	231
100	206	210	210	212
150	193	197	200	209
300	187	191	201	208
600	185	190	194	206
1000	188	193	195	208

Tabla 4.2 MOTP medio obtenido en los seminarios evaluados.

MOTA	Tamaño de voxel			
Partículas	2 cm	3 cm	4 cm	5 cm
50	9,94%	18,47%	15,23%	17,31%
100	64,99%	55,23%	56,62%	47,1%
150	74,95%	71,57%	65,73%	52,16%
300	81,4%	77,68%	66,51%	87,5%
600	81,19%	75,61%	71,25%	57,38%
1000	79,82%	71,27%	72,5%	57%

Tabla 4.3 MOTA medio obtenido en los seminarios evaluados.

El algoritmo se ha ejecutado con unos seminarios de unos 7500 fotogramas cada uno llevados a cabo en la sala CHIL de la UPC durante el 6-7-2005, el 20-7-2005 y el 22-7-

2005. En dichos seminarios hay hasta 6 personas dentro de la sala, así como elementos de mobiliario tales como sillas y mesas. Las posiciones de las personas han sido etiquetadas manualmente para disponer de datos de *groundtruth* con los que aplicar las métricas. Los resultados que se muestran a continuación son una media de todas las evaluaciones llevadas a cabo.

Estos mismos resultados en forma de gráficas se muestran en la fig. 4.3 y 4.4, comparando con los resultados obtenidos por el algoritmo basado en el filtro de Kalman.

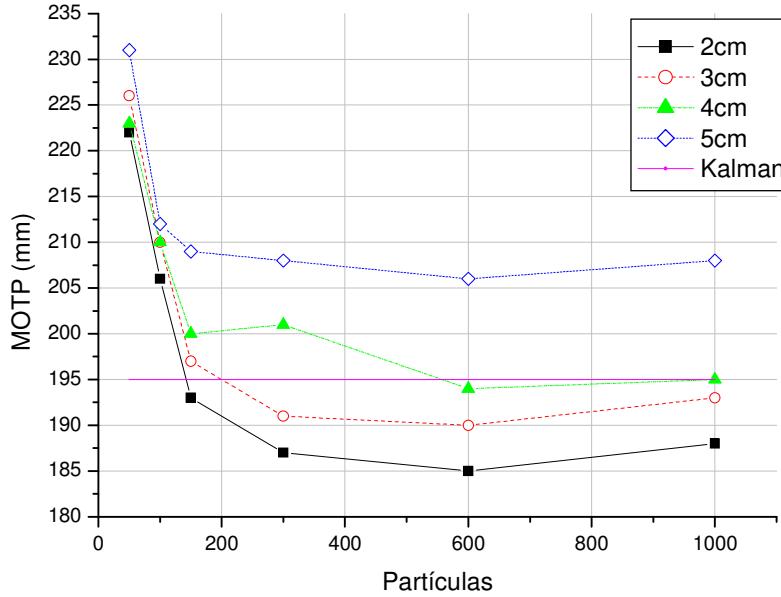


Fig. 4.3 MOTP medio obtenido en todos los seminarios procesados para distinto número de partículas y tamaño de voxel. En magenta se muestran los resultados obtenidos mediante el algoritmo basado en Kalman.

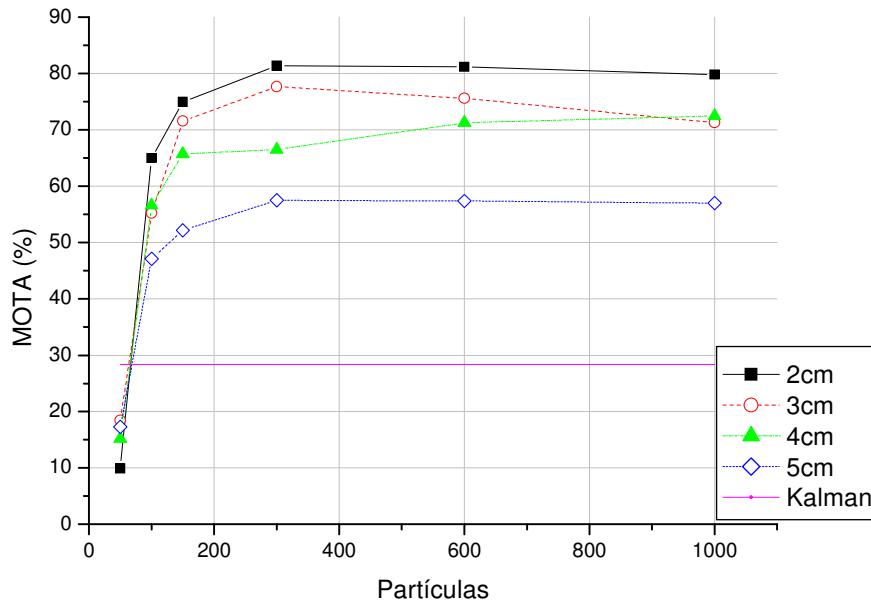


Fig. 4.4 MOTA medio obtenido en todos los seminarios procesados para distinto número de partículas y tamaño de voxel. En magenta se muestran los resultados obtenidos mediante el algoritmo basado en Kalman

Como era de esperar, los mejores resultados se obtienen con el menor tamaño de voxel. Para este caso las ratios obtenidas se pueden ver en la tabla 4.3.

Partículas	MOTP	\bar{m}	\bar{fp}	\bar{mme}	MOTA
50	222	22,5%	18,3%	49,3%	9,9%
100	206	10,0%	16,9%	8,1%	65,0%
150	193	19,3%	4,4%	1,4%	74,9%
300	187	4,6%	12,9%	1,1%	81,4%
600	185	3,3%	14,6%	0,9%	81,2%
1000	188	4,3%	14,7%	1,0%	80,0%

Tabla 4.4 Resultados cuantitativos para un tamaño de voxel de 2cm.

De la evolución de los resultados se desprende que no hay mejora sustancial empleando más de 600 partículas y que el comportamiento optimizado se halla alrededor de 500 partículas por filtro. Las tasas de error mostradas en la tabla 4.3 muestran que el algoritmo resuelve de forma notable los posibles cruces entre personas (fig. 4.4) y que el error más frecuente sigue siendo la falsa detección. Esto suele ocurrir debido a los espurios que supone el movimiento de mobiliario de la sala (sillas, proyectores, portátiles, etc.) que reducen la precisión del seguimiento hasta sobrepasar el umbral D_{th} .

Uno de los casos más frecuentes de falsa detección detectados ocurre cuando alguno de los asistentes cambia un objeto de lugar, creando un volumen. Cuando esta persona se aleja de dicho objeto, puede suceder que la reconstrucción de su volumen no resulte tan verosímil y sólida como la del volumen indeseado, de modo que el centroide estimado suele quedar en una posición intermedia entre el sujeto y el objeto espurio (ver fig 4.7). Tras esta situación el algoritmo puede volver a detectar a la persona, o bien corrigiendo la posición gracias a una mejor reconstrucción del volumen en los fotogramas siguientes o bien creando un nuevo filtro de partículas, pudiendo reducir así el error por pérdida, aunque el filtro que queda siguiendo a la silla contabiliza falsas detecciones. Este tipo de error ocurre frecuentemente con las sillas de la sala. Los asistentes, al tomar asiento, crean un volumen asociado a la silla. Al alejarse de la silla el algoritmo puede considerar que ésta forma parte de la persona o que es otra persona distinta (como la elipsoide verde de la Fig. 4.8)

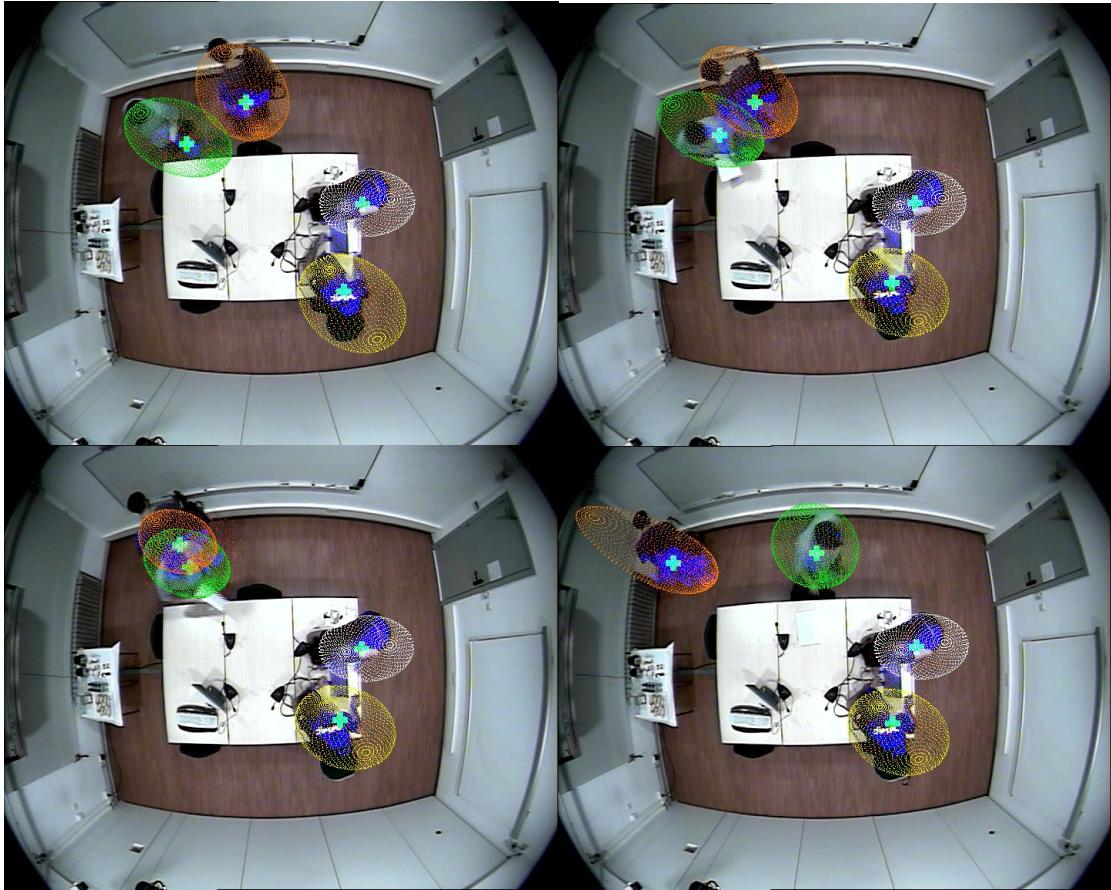


Fig. 4.4. Ejecución con 600 partículas y tamaño de voxel de 2cm. El algoritmo resuelve con éxito el cruce de dos asistentes al seminario del 06-07-2005.

El problema se observa en las ejecuciones y que explica de modo global el 20% de fallos de seguimiento está fuertemente asociado a los volúmenes reconstruidos (fig. 4.6). En muchos fotogramas resulta difícil decir si los volúmenes observados se corresponden a personas, incluso manualmente. Si los volúmenes que modelan a las personas son suficientemente sólidos y conservan la forma y la altura, no sólo el filtro se comporta mejor en términos de precisión y eficacia, sino que es posible añadir reglas de decisión relativamente sencillas y con alta probabilidad de validez que permitirían reducir los efectos negativos de objetos espurios. Esto puede ser visto en la detección de personas e inicialización de filtros. Al tener volúmenes poco sólidos es más difícil hallar un conjunto conexo que cumpla las condiciones expuestas en 3.3.1. Esto redunda en que durante un número determinado de fotogramas se tiene un objetivo sin hipótesis, lo cual es, por definición, una pérdida.

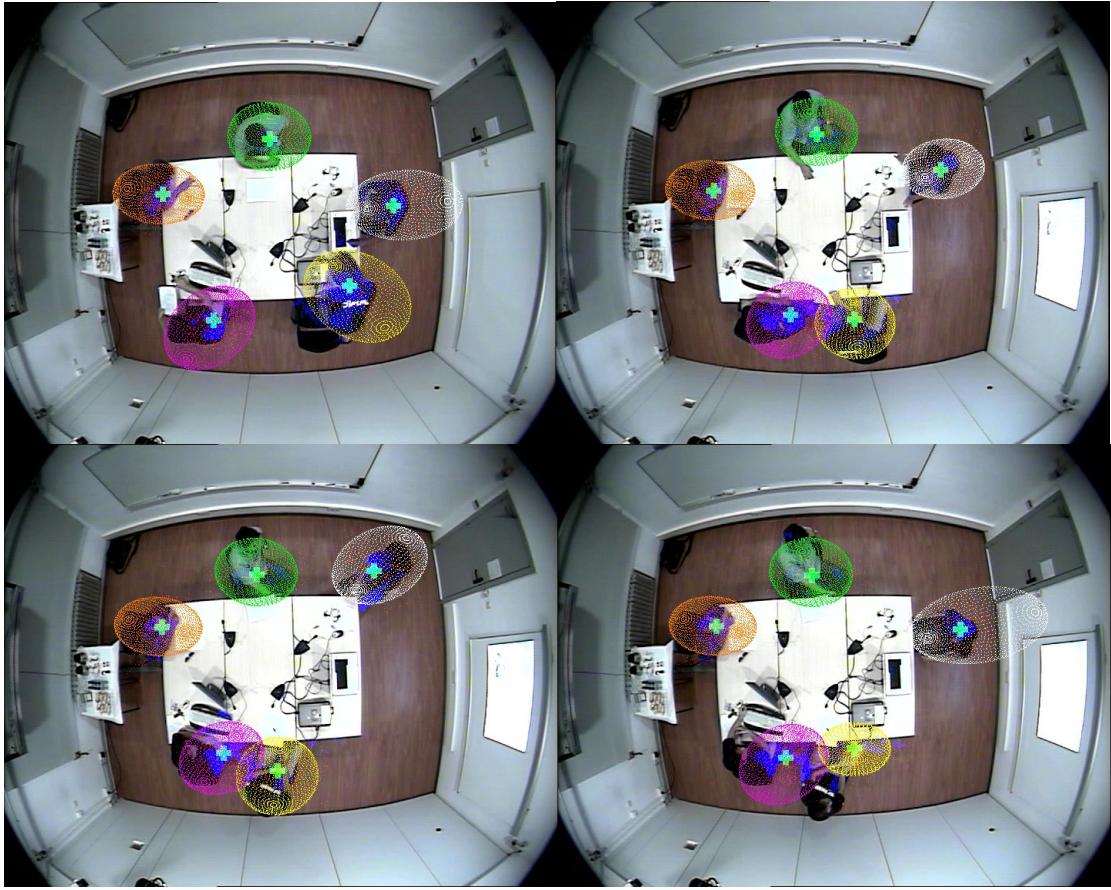


Fig. 4.6 Ejecución del seminario del 07-06-2005 con 600 partículas y voxel de 2cm. Puede observarse como la reconstrucción de un volumen espurio cerca del proyector, en al parte inferior derecha de cada imagen, atrae al filtro amarillo inutilizando el bloqueo que mantenía al filtro magenta estimando la posición correcta.

Al examinar las secuencias de vídeo resultantes, puede observarse, especialmente con cantidades reducidas de partículas, que existe un cierto nerviosismo del filtro. Este efecto es notorio cuando la persona está quieta. Las sucesivas hipótesis del filtro oscilan alrededor del centroide objetivo, siendo la amplitud de dichas oscilaciones erróneas mayor cuanto menor es el número de partículas.

Aún así, los resultados expuestos son un 50% superiores en términos de MOTA y unos 10 mm más precisos (MOTP) que los del algoritmo basado en Kalman, lo cual demuestra la eficacia de un algoritmo sencillo pero enfocado a distribuciones multimodales frente a otro basado en Kalman.

Asimismo, los resultados expuestos, a pesar de que hay que tener en cuenta que únicamente evalúan seminarios UPC, son superiores a los resultados de otros algoritmos presentados en las evaluaciones CLEAR'06 [9,13].

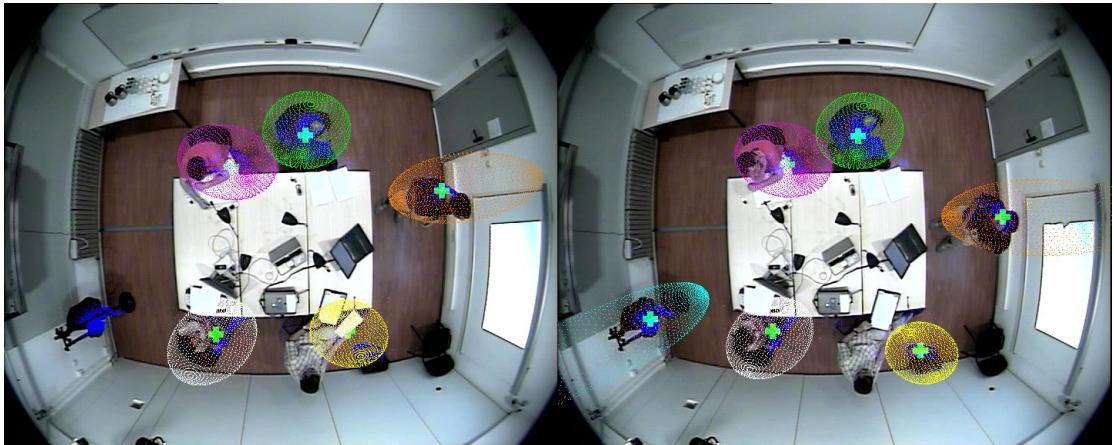


Fig. 4.7 Ejemplos de falsa detección en una ejecución con 1000 partículas y voxels de 2 cm sobre el seminario del 20-07-2005. La chaqueta colgada por uno de los asistentes es detectada en un instante concreto como un objetivo. En la parte inferior derecha, un objeto depositado constituye un conjunto conexo de voxels que atrae al filtro de partículas amarillo. Este último caso está causado además por la falta de un volumen conexo en la posición en que se sienta la persona objetivo.

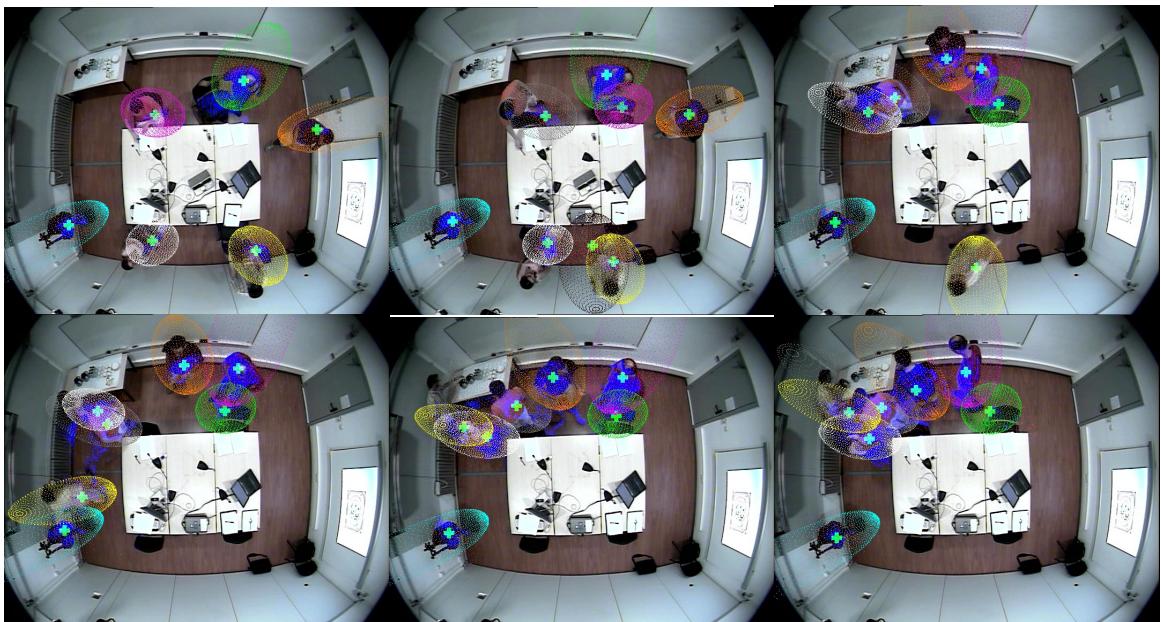


Fig. 4.8 Ejecución con 1000 partículas y voxels de 2 cm sobre el seminario del 20-07-2005. Los asistentes se concentran alrededor de la mesa ubicada en la parte superior izquierda. A pesar de las falsas detecciones, el algoritmo mantiene una consistencia notable con las hipótesis acerca de la posición de cada persona.

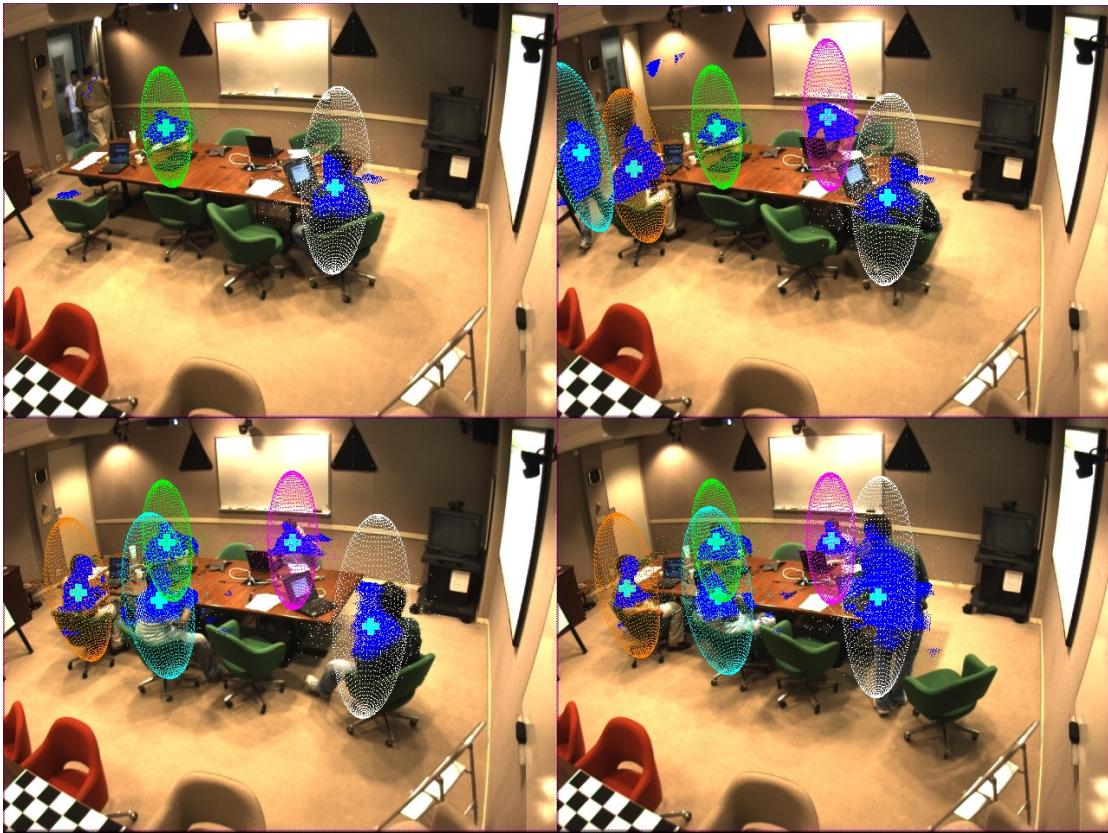


Fig. 4.9. Fotogramas 800, 900, 1000 y 1100 de una secuencia de IBM no evaluada con las métricas CLEAR. Los únicos intercambios que se producen son debidos a que los asistentes salen de la zona en que es posible reconstruir volúmenes. Si se garantiza que la reconstrucción es suficientemente sólida y guarda una cierta correlación con la forma a la que representa, puede plantearse un filtrado de los voxels para una altura inferior a un cierto umbral, eliminando así gran parte de los volúmenes de las sillas.

4.4 Comparativa de remuestreo y asignación de partículas a voxels.

Otras alternativas de parametrización han sido presentadas, por lo que se han probado algunas de ellas en un solo seminario a modo de comparativa. Dicho seminario tuvo lugar el 6-7-2005. Consta de 7500 fotogramas de 720x576 a 25 fps.

El objetivo es determinar los resultados para los dos tipos de remuestreo propuestos en 3.3.3.3. También se ha probado el algoritmo con un sistema de asignación de partículas a voxels que emplea un *ranking* de candidatos dentro de un entorno de la partícula, priorizando los voxels de *foreground* más cercanos.

La tabla 4.4 resume los resultados obtenidos para todo el seminario con el detalle de las ratios de error. Llamamos “Normal” a la versión del algoritmo empleada en el apartado 4.3, con la que hemos evaluado todos los seminarios disponibles. “Ranking” es una versión que tiene la misma distancia de remuestreo pero que emplea un criterio de distribución de partículas más complejo, tal y como hemos explicado anteriormente, mientras “Remuestreo Expandido” emplea las distancias de (3.12). “Remuestreo Expandido+Ranking” combina las dos estrategias.

Versión	Partículas	MOTP	<i>m</i>	<i>fp</i>	<i>mme</i>	MOTA
Normal	50	244mm	18,7%	15,3%	41,0%	25,1%
	100	227mm	7,4%	12,5%	6,0%	74,0%
	150	211mm	22,7%	5,2%	1,6%	70,5%
	300	197mm	3,9%	11,0%	0,9%	84,2%
	600	195mm	2,8%	12,2%	0,8%	84,3%
	1000	198mm	3,5%	12,0%	0,8%	83,7%
Ranking	50	228mm	56,6%	2,0%	12,3%	29,0%
	100	226mm	10,8%	8,9%	1,4%	78,9%
	150	220mm	8,7%	6,9%	1,1%	83,4%
	300	210mm	5,6%	11,8%	0,7%	81,9%
	600	208mm	7,1%	11,0%	0,5%	81,4%
	1000	227mm	9,2%	7,4%	1,0%	82,4%
Remuestreo Expandido	50	216mm	56,9%	2,3%	11,7%	29,1%
	100	208mm	18,3%	9,8%	3,5%	68,4%
	150	187mm	10,6%	13,4%	1,4%	74,5%
	300	189mm	5,3%	17,4%	0,8%	76,4%
	600	189mm	5,2%	17,4%	1,0%	76,4%
	1000	194mm	3,2%	11,3%	0,6%	84,8%
R. Expandido+Ranking	50	228mm	56,6%	2,0%	12,3%	29,0%
	100	215mm	7,2%	10,1%	1,1%	81,6%
	150	201mm	6,6%	11,8%	0,9%	80,7%
	300	206mm	6,3%	11,3%	0,8%	81,6%
	600	208mm	6,4%	10,6%	0,7%	82,3%
	1000	210mm	4,4%	10,4%	0,8%	84,4%

Tabla 4.5. Resultados comparativos entre propuestas de distancia de remuestreo y de redistribución de partículas.

Los resultados muestran que la combinación escogida para las evaluaciones completas es la más eficiente, dado que consigue mejores resultados con menos partículas y sin emplear ningún ranking, método que aumenta la complejidad computacional del algoritmo. Por otro lado, podemos observar como las estrategias basadas en muestreo expandido alcanzan su mayor MOTA con el máximo de partículas. En términos de precisión (MOTP) hay que destacar el efecto negativo del empleo del ranking de voxels priorizando los de *foreground* a menor distancia. Este resultado revela que obtener nuevas muestras en la iteración k con la observación z_k , a nivel de voxel, sin analizar el entorno, no mejora el resultado, más bien todo lo contrario. Esto es debido a que el volumen reconstruido, que constituye la observación disponible, es altamente ruidosa, especialmente si tomamos cada voxel por separado. Esto, sumado al hecho de que un filtro de partículas SIR es muy sensible a “*outliers*” causados por ruido de reconstrucción, hace que la precisión decaiga si consideramos que un voxel es una muestra más verosímil por ser de *foreground*. Aunque en general parece que se reducen las falsas detecciones, las alternativas basadas en este ranking presentan, en general, tasas de pérdidas mayores. Esto puede observarse especialmente cuando no se emplea muestreo expandido. En el caso en que sí se utiliza el remuestreo expandido, se consigue mejorar el MOTA para un número inferior a 1000

partículas pero se paga en términos de precisión y, como ya hemos dicho, complejidad computacional.

5. Conclusiones

En este último apartado resumimos brevemente los resultados obtenidos en función de los objetivos marcados al inicio del proyecto. Esto supone revisar la contribución tanto de la solución propuesta al problema de seguimiento como de este trabajo a la participación de la UPC en el proyecto CHIL. Finalmente, revisaremos los puntos a desarrollar en un futuro.

5.1 Cumplimiento de objetivos

El objetivo básico de este proyecto era desarrollar un algoritmo de seguimiento multipersona en entornos de sala inteligente con múltiples cámaras. El algoritmo tenía que estar basado en la tecnología de filtros de partículas.

Inicialmente se definió un diseño que utilizase no sólo los volúmenes reconstruidos sino también la información de color. Los resultados experimentales (ver Anexos) han demostrado que la información de color que podemos extraer hoy por hoy no tiene la suficiente consistencia para elaborar o robustecer un algoritmo que trabaja con una voxelización del mundo tridimensional.

Este trabajo no supone la definición de un problema nuevo ni propone una solución novedosa. Sin embargo, el enfoque orientado a voxel y a escenarios de sala inteligente sí que es una característica diferencial con respecto al gran número de trabajos de algoritmos de seguimiento basados en filtros de partículas que se han publicado hasta la fecha.

Este sistema de seguimiento supone un avance cualitativo en términos de eficacia respecto al algoritmo diseñado en la UPC basado en filtro de Kalman, demostrando así que el filtro de partículas es la mejor estrategia actual para un algoritmo de seguimiento. Desde el punto de vista experimental, las razones que llevan a tal conclusión son las siguientes:

- Los resultados obtenidos con las métricas CLEAR
- Sencillez de implementación
- La escalabilidad del filtro de partículas permite añadir de modo sencillo espacios de características distintos para el cálculo de pesos.
- La capacidad del filtro de partículas de hacer seguimiento sobre una región del espacio y no sobre un único punto.

A las ventajas intrínsecas de esta tecnología se añaden las que conlleva el hecho de trabajar con una voxelización del escenario tridimensional.

- El algoritmo no tiene que modelar cambios de escala.
- La discretización en voxels adecuada permite emplear un número reducido de partículas

Sin embargo, son no menos importantes las contraprestaciones que presenta la reconstrucción 3D de la escena:

- Gran cantidad de errores y ruido debido a cambios de iluminación y objetos de mobiliario de la sala que constituyen espurios.
- Incremento de la complejidad computacional

El algoritmo presentado, aunque resuelve muchos intercambios correctamente y pierde pocos objetivos, es bastante sensible al ruido de reconstrucción, algo que se traduce especialmente en el incremento de falsas detecciones. Es por ello que el primero de los inconvenientes mostrados supone la fuente de definición de un subconjunto de problemas para los que hallar métodos y reglas de decisión para robustecer el filtro de partículas.

En el marco del proyecto CHIL y dentro de la campaña de evaluaciones CLEAR '06 [9], los resultados obtenidos son prometedores, ya que, en comparación con el mejor resultado, nuestro algoritmo supera en casi un 20% la estadística de MOTA y el MOTP obtenido es casi 20 mm inferior.

Es por todo ello que el trabajo desarrollado constituye no sólo una solución más al problema de seguimiento multipersona 3D en entornos de sala inteligente con múltiples vistas, sino que sienta unas bases a desarrollar.

5.2 Trabajo Futuro.

Las conclusiones a las que se ha llegado con este trabajo dejan abiertas vías de investigación entorno al seguimiento multi-persona en salas inteligentes basado en filtros de partículas. El trabajo a desarrollar empieza por una revisión de los juegos de parámetros que admite el diseño actual y su completa evaluación visual y según el formato estándar mostrado en 4.1. Asimismo adquiere especial importancia la mejora de reglas de decisión y conformación de pesos que permitan minimizar el efecto negativo del ruido de reconstrucción y de los objetos molestos.

El diseño propuesto emplea esquemas de remuestreo y especialmente de propagación de partículas muy sencillos. Es por ello que el estudio de métodos alternativos de obtención de muestras a partir de distribuciones que permitan mejorar la estimación también resultaría adecuado.

Vista la capacidad de los filtros de partículas de efectuar estimaciones robustas empleando varios espacios de características como funciones de verosimilitud queda estudiar la posibilidad de añadir otras modalidades, como por ejemplo la orientación o el color.

Finalmente, restaría mejorar la eficiencia del algoritmo, en términos de programación, para su implementación en tiempo real.

Anexo – Estudio del espacio de color en el seguimiento basado en filtro de partículas

Este apéndice muestra el trabajo realizado alrededor de la información de color disponible a través de las múltiples cámaras de la sala, como incorporarla a los voxels y por qué no se ha añadido esta modalidad en el seguimiento diseñado.

A.1 Obtención de la Información de color

El algoritmo de reconstrucción de la escena presentado no explota la información de color, simplemente se usa para discernir los objetos del fondo. Sin embargo, se podría poner foco en el color, de modo que la reconstrucción de la escena fuese una asignación de un color a cada voxel. Esto permitiría extraer otra característica útil para el seguimiento en determinadas situaciones tales como fusión de volúmenes de personas distintas.

Dado que cada voxel puede ser visto desde varios puntos, este planteamiento exige hallar voxels que tengan un único color para cualquier posible interpretación de la escena o, lo que es lo mismo, hallar voxels fotoconsistentes [23, 24].

Las principales dificultades que hallamos para conseguir una asignación que permita trabajar con el color en 3D son la ambigüedad espacial y cromática de ciertas regiones de las escena, tal y como puede verse en las ilustraciones que se muestran a continuación.

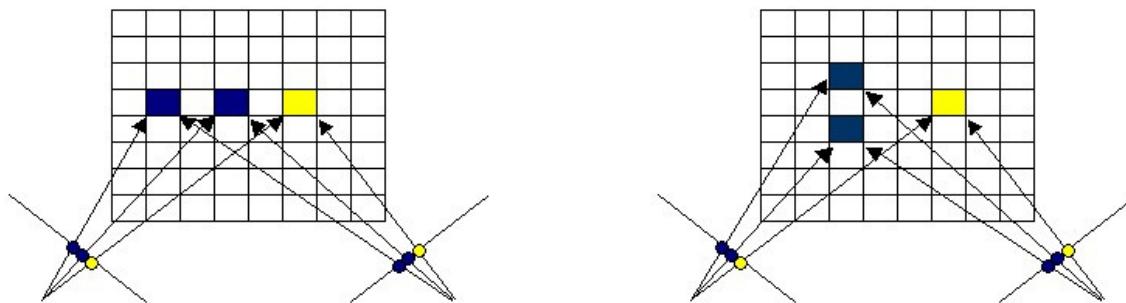


Fig. A.1. Ambigüedad espacial. Dos posibles asignaciones de color en la escena.

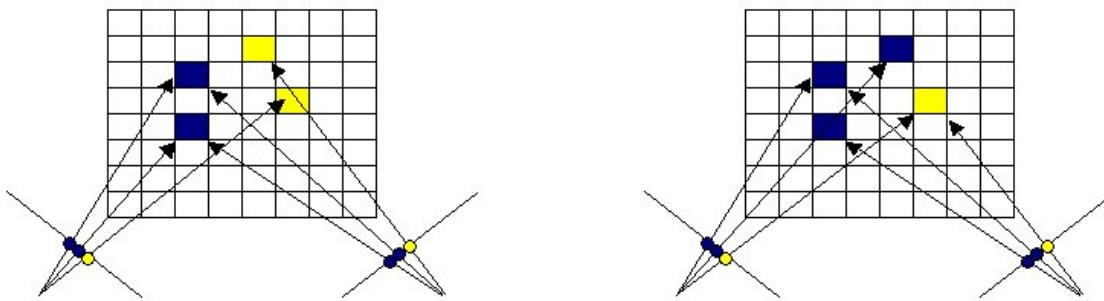


Fig. A.2. Ambigüedad cromática. El voxel superior tiene asignaciones distintas en las dos escenas debido a las occlusiones del mismo. Podría considerarse que el voxel no es fotoconsistente.

La primera consideración que podemos tomar para seleccionar voxels fotoconsistentes es construir una estadística a partir del conjunto de colores que resultan de proyectar cada uno de los voxels en las cinco cámaras de la sala. Una vez determinada la estadística se rechazan aquellos tonos que difieren más de los parámetros establecidos hasta hallar aquél o aquellos que mejor se ajustan a la estadística. Este tipo de métodos de estimación de estadísticas y rechazo de *outliers* no emplea información geométrica de la sala y por lo tanto no es eficiente, porque no considera occlusiones a priori (la fotoconsistencia las determinaría a posteriori) y porque aplica criterios para voxels internos cuyo color es indeterminado. Además presenta un problema en cuánto a criterios de ordenación y métricas en el espacio de color (RGB, CbCr, etc).

Parece evidente que el enfoque óptimo para el entorno multi-vista en una sala inteligente es hallar el conjunto de cámaras que ven un voxel determinado sin occlusiones y determinar entonces la fotoconsistencia para el conjunto de colores que resultan de proyectar ese voxel en las cámaras con visión directa.

Para determinar si hay visión directa entre un voxel y una cámara determinada, simplemente habría que hallar una recta entre ambos elementos que no pasase por ningún voxel perteneciente a objetos activos. Sin embargo, el problema real es trazar una recta en un espacio 3D con granularidad de voxel. Para solventar esta dificultad se ha empleado una adaptación del algoritmo diseñado por Bresenham [31, 32] en 1962, cuya finalidad era aproximar una recta en un espacio 2D discreto.

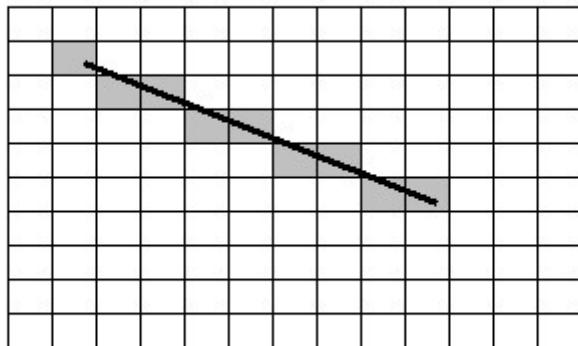


Fig. A.3 Rayo de Bresenham 2D.

La adaptación tridimensional consiste en hallar el conjunto de voxels que atraviesa la recta que va de (x_0, y_0, z_0) a (x_1, y_1, z_1) , siendo en nuestro caso (x_0, y_0, z_0) las coordenadas de voxel y (x_1, y_1, z_1) las coordenadas del centro de cámara. Con el algoritmo se obtiene un array de voxels que son analizados. Si alguno es detectado como perteneciente a un objeto activo, es decir, forma parte de algún volumen reconstruido en ese instante, se considera que hay occlusión y que, por lo tanto, no hay visión directa entre esa cámara y el voxel bajo estudio. Este cálculo debe excluir a los voxels del propio volumen para colorear el interior o propagar posteriormente el color de la superficie hacia el centroide con la misma técnica de rayo de Bresenham, sólo que ahora se colorean todos los voxels del array obtenido.

Finalmente, para el conjunto de cámaras que presentan visibilidad directa se promedia el color obtenido en cada píxel proyectado.

A.2 Filtro de partículas con información de color

El objetivo de la definición consistente de un color por voxel es robustecer el algoritmo de seguimiento. Existen trabajos que combinan los filtros de partículas con la información de color o que simplemente usan el color para efectuar un seguimiento robusto. Muchos han sido desarrollados con éxito notable en 2D [7, 17, 28, 30]. En nuestro caso, el uso de la información geométrica de los volúmenes puede ser insuficiente, especialmente en casos de fusión volumétrica. Por ello, ha de modificarse el algoritmo de modo que la información de color sea empleada en los instantes donde pueda aumentar la fiabilidad de la estimación. Con este fin hay que determinar en qué grado el histograma de color de un objetivo lo discrimina del resto o como mínimo de aquellos con los cuales puede interaccionar estableciendo contacto. El histograma del objetivo lo definimos con el conjunto de voxels coloreados dentro de la región elipsoidal que el algoritmo emplea para el modelado de interacciones (3.3.3.3). Para validar dicho histograma éste ha de presentar una cierta autocorrelación a lo largo del tiempo. Para ello empleamos la métrica de Bhattacharyya [22] entre distintos instantes de tiempo.

$$\rho_k = \sum_j^{N_b} \sqrt{h_k(j)h_{k-n}(j)} \quad (\text{A.1})$$

donde $h(j)_k$ es el mencionado histograma en el instante k , N_b el número de niveles que tiene dicho histograma. Respecto al histograma, podemos trabajar en el espacio RGB aunque el más indicado sería el CbCr para eliminar errores causados por cambios de iluminación. Si la métrica de Bhattacharyya obtenida instante a instante supera un cierto umbral de fiabilidad ρ_{Th} , podemos emplear la información de color, además de los volúmenes reconstruidos, para mejorar la estimación de posición.

La propuesta de diseño que contempla el uso de color introduce esta característica en el cálculo de pesos. Sea $w_k^{i(g)}$ el peso de una partícula calculado tal y como hemos mostrado en el apartado 3.3.3.2, es decir, vía el análisis geométrico de un entorno de la partícula. Sea asimismo $w_k^{i(c)}$ el peso hipotético de un partícula calculado mediante un análisis del color que puede tener asignado. El peso final de la partícula será:

$$w_k^i = \alpha w_k^{i(g)} + \beta w_k^{i(c)} \quad (\text{A.2})$$

Las constantes α y β variarán proporcionalmente al grado de verosimilitud que puedan tener una u otra características del espacio estudiado. Así, dos personas muy próximas en el espacio pero con vestimentas de colores distintos tendrán como resultado α pequeña y β grande. También puede considerarse para determinar α y β la desviación de las partículas con peso en x e y respecto al centroide. Una desviación grande haría crecer β , puesto que ciertas partículas podrían estar considerando la verosimilitud geométrica de objetos espurios. Si es alta se pueden emplear constantes complementarias para el cálculo de pesos, esto es, el caso particular en que β es $1-\alpha$, sin embargo este método tiende a polarizar fuertemente el cálculo de pesos hacia uno u otro método y haciendo que un error del elemento que determina α se propague más fácilmente al seguimiento. Para determinar

α la información más relevante de que disponemos es la posición estimada de los objetivos. La distancia entre centroides será proporcional a α, indicando que elementos muy próximos tienden a fusionarse, haciendo que las partículas de varios filtros se entremezclen elevando la probabilidad de error. Para β la métrica de Bhattacharyya es la herramienta básica que determina qué grado de verosimilitud tiene un color de pertenecer a una persona. En este caso en particular empleamos una métrica en forma de distancia, que no es más que el valor complementario que obtenemos del cálculo del coeficiente de Bhattacharyya, es decir, $d=1-\rho$. Una distancia grande entre filtros de partículas indica que los histogramas de sus respectivos objetivos permiten distinguir entre las personas a seguir.

Determinar el nuevo peso $w_k^{i(c)}$ es sencillamente asignar un color a cada partícula y ver que probabilidad tiene dicho tono en el histograma del filtro. Puesto que cada partícula ha sido asignada a un voxel, se trata de determinar nuevamente el color de un voxel. Formalmente, si $h_k(j)$ es el histograma asociado un filtro de partículas en el instante k , x_k^i la partícula i -ésima del filtro mencionado y $C_{rgb}(\cdot)$ una función que asigna un color RGB a la coordenada representada por x_k^i , el peso se define como la siguiente expresión:

$$w_k^{i(c)} = h_k(C_{RGB}(x_k^i)) \quad (\text{A.3})$$

A.3 Dificultades

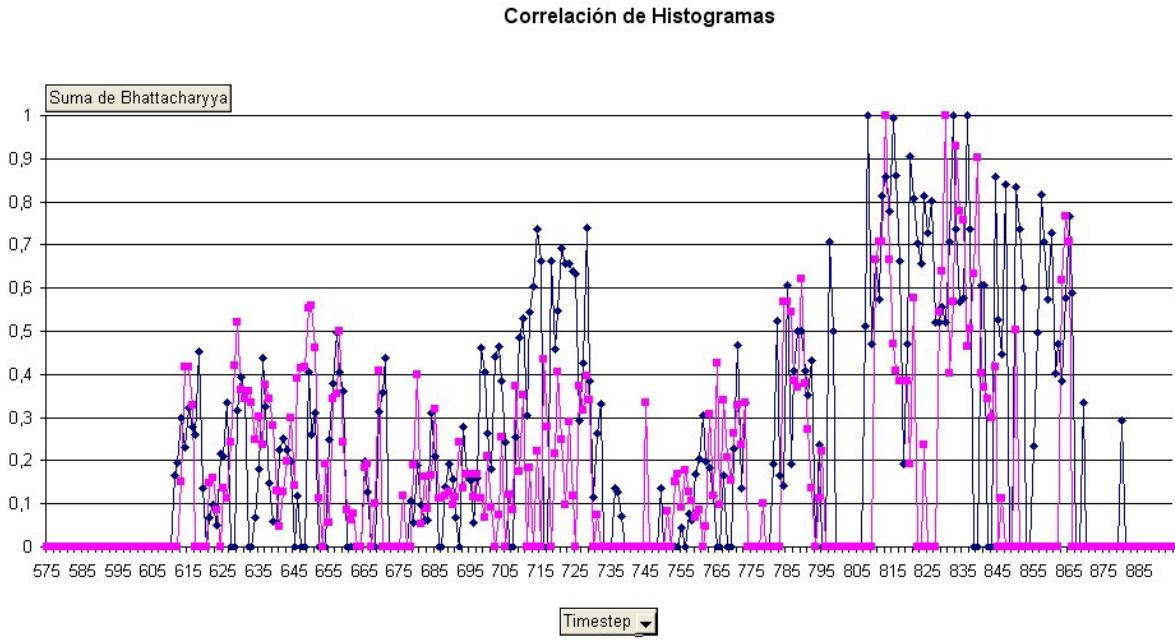
Al implementar el coloreado superficial de los volúmenes se observa a simple vista que el conjunto de colores obtenidos no será suficiente para diferenciar dos objetivos. Más allá de la percepción visual (fig A.4), la correlación entre los histogramas instante tras instante valida esta aseveración (fig. A.5 y A.6).

Esta limitación se debe, por una parte, a la sencillez del planteamiento empleado. Sin embargo, la problemática real deriva de los errores de calibración y definición de las cámaras de la sala y constituye la razón por la que no se considera el color en el algoritmo definitivo



Fig. A.4. Fotograma de seminario y su correspondiente coloreado de voxel.

a)



b)

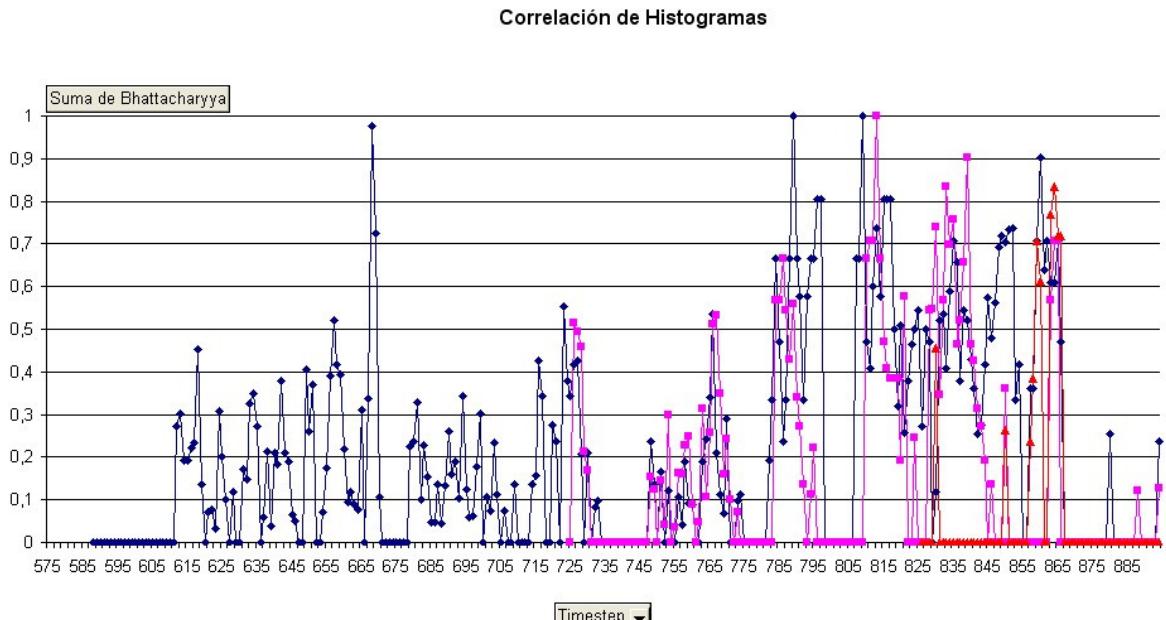


Fig. A.5. Métrica de Bhattacharyya entre un histograma en el instante k y $k-1$. a) Calculado para los filtros 1 y 2 (las dos primeras personas en entrar a la sala). b) Calculado para los filtros 3, 4 y 5. La métrica rara vez supera el 0.75 por lo que los histogramas no se pueden validar. Los ceros suelen ocurrir porque no se han podido extraer ningún histograma.

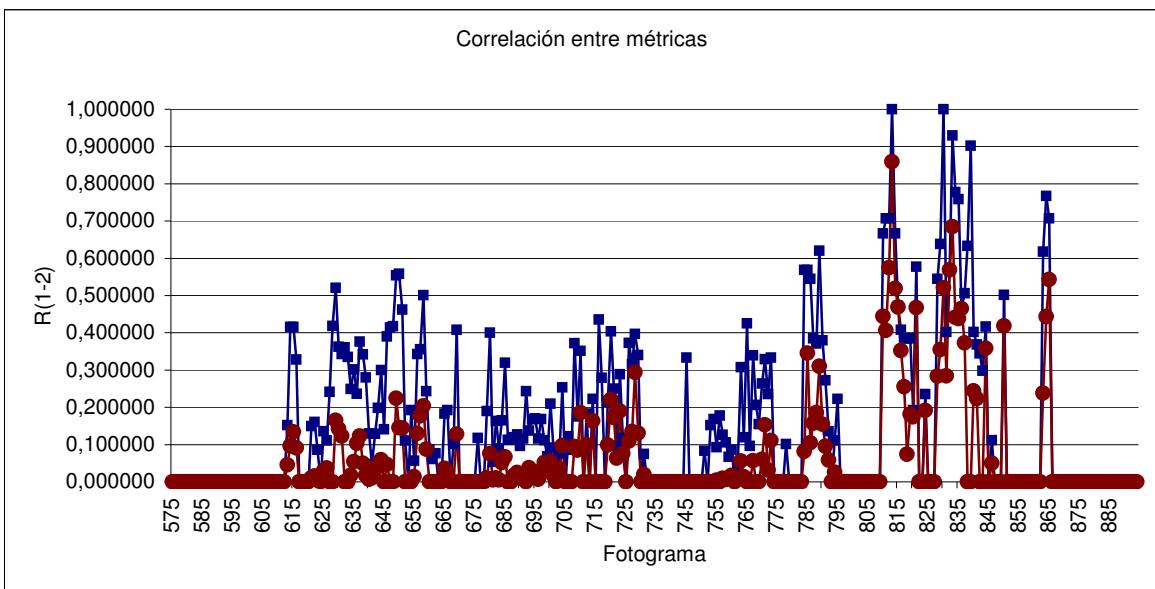


Fig. A.6. El gráfico muestra la métrica de Bhattacharyya para el filtro 2 y el producto de las métricas 1 y 2, a modo de medida de correlación. El resultado indica que los factores que hacen fallar al recurso de color son mayormente externos al algoritmo y afectan prácticamente por igual a todos los filtros.

En resumen, no se ha añadido el color como característica para calcular la verosimilitud de las partículas debido a la poca calidad que ofrecen las imágenes de las cámaras de la sala para efectuar el coloreado de voxels.

Bibliografía

- [1] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, “*A Tutorial on Particle Filters for On-Line Non-Linear/Non-Gaussian Bayesian Tracking*”, IEEE Trans. Signal Processing, vol. 50, no. 2, pp. 174-189, 2002.
- [2] Z. Khan T. Balch and F. Dellaert, “*Efficient Particle Filter-Based Tracking of Multiple Interacting Targets Using an MRF-Based Motion Model*” Proc. IEEE/RSJ Conf. Intelligent Robots and Systems, 2003.
- [3] K. Bernardin, T. Gehrig, R. Stiefelhagen. ”*Multi- and Single View Multiperson Tracking for Smart Room Environments*”, CLEAR Evaluation Workshop, Southampton, UK, April 2006
- [4] M. Isard and A. Blake, “*Condensation - Conditional density propagation for visual tracking*,” Intl. J. of Computer Vision, vol. 29, no. 1, 1998.
- [5] Zia Khan, Tucker Balch, and Frank Dellaert, “An MCMC-based Particle Filter for Tracking Multiple Interacting Targets”, Technical Report number GIT-GVU-03-35 October 2003
- [6] J. Sabourne, F. Charpillet, “*Using Interval Particle Filter for Marker less 3D Human Motion Capture*”, Tools with Artificial Intelligence, 2005. ICTAI 05. 17th IEEE International Conference on Volume , Issue , 14-16 Nov. 2005 Page(s): 7 pp.
- [7] E. Maggio and A. Cavallaro, ”*Hybrid particle filter and Mean Shift tracker with adaptive transition model*”, Proc. Int. Conf. Acoustics, Speech, and Signal Processing 2005.
- [8] Neal Checka, Kevin Wilson, Michael Siracusa, Trevor Darrell, “*Multiple Person and Speaker Activity Tracking with a Particle Filter*”, International Conference on Acoustics, Speech, and Signal Processing, 2004.
- [9] R. Stiefelhagen, K. Bernardin, R. Bowers, J. Garofolo, D. Mostefa, P. Soundararajan, ”*The CLEAR 2006 Evaluation*”, Proceedings of the first International CLEAR evaluation workshop, CLEAR 2006, Springer Lecture Notes in Computer Science, No. 4122., pp 1-45.
- [10] ”*CLEAR evaluation webpage*”, <http://www.clear-evaluation.org>.
- [11] ”*CHIL - Computers In the Human Interaction Loop*”, <http://chil.server.de>.
- [12] Keni Bernardin, Alexander Elbs, Rainer Stiefelhagen, ”*Multiple Object Tracking Performance Metrics and Evaluation in a Smart Room Environment*”, The Sixth IEEE International Workshop on Visual Surveillance (in conjunction with ECCV), Graz, Austria, May 2006.
- [13] A. López, C. Canton-Ferrer, J. R. Casas. ”*Multi-Person 3D Tracking with Particle Filters on Voxels*”. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Honolulu (USA), April 16-20, 2007.
- [14] C. Canton-Ferrer, J.R. Casas, M. Pardàs. (2005). ”*Towards a bayesian approach to robust finding correspondances in multiple view geometry environments*”. Published in Lecture notes in computer science , 3515 () : 281-289.
- [15] A. Abad, C. Canton-Ferrer, C. Segura, J.L. Landabaso, D. Macho, J.R. Casas, J. Hernando, M. Pardàs, C. Nadeu. ”*UPC Audio, Video and Multimodal Person Tracking Systems in the CLEAR Evaluation Campaign*”. CLEAR Evaluation Workshop, Southampton (UK), April 6-7 2006. Published in Lecture Notes on Computer Science, vol.4122, pp. 93-104, Springer-Verlag.
- [16] J. MacCormick and A. Blake, ”*A probabilistic exclusion principle for tracking multiple objects*,” in Proc. Int. Conf. Comput. Vision, 1999, pp. 572–578.

- [17] K. Nummiaro, E. Koller-Meier, and L. Van Gool, “*A color-based particle filter*,” in Proc.of the 1st Workshop on Generative-Model-Based Vision, June 2002, pp. 53–60.
- [18] Kai Nickel, Tobias Gehrig, Rainer Stiefelhagen, John McDonough, ”*A Joint Particle Filter for Audio-visual Speaker Tracking*”, International Conference on Multimodal Interfaces ICMI 05, Trento, Italy, October 2005
- [19] D. Reid, “*An algorithm for tracking multiple targets*,” *IEEE Trans. Automatic Control*, vol. AC-24, pp. 843–854, 1979.
- [20] C. Stauffer and W. Grimson. “*Adaptive background mixture models for realtime tracking*”. In Proc. of CVPR, 1998, 333–339, 1998.
- [21] O. Garcia Codina “*Mapping 2D images and 3D world objects in a multicamera system*”, MS Thesis, 2004.
- [22] F. Aherne, N. Thacker, and P. Rockett, “*The Bhattacharyya Metric as an Absolute Similarity Measure for Frequency Coded Data*,” *Kybernetika*, vol. 34, no. 4, pp. 363-368, 1998.
- [23] Seitz, S.M., Dyer, C.R., “*Photorealistic scene reconstruction by voxel coloring*”, CVPR. (1997) 1067–1073
- [24] Kutulakos, K.N., Seitz, S.M, “*A theory of shape by space carving*”. *International Journal of Computer Vision* 38 (2000) 199–218
- [25] Y. Rui and Y. Chen, “*Better Proposal Distributions: Object Tracking Using Unscented Particle Filter*” Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. II, pp. 786-793, 2001.
- [26] N. Bergman, “*Recursive Bayesian estimation: Navigation and tracking applications*”, PhD. Dissertation, Linköping Univ., Linköping, Sweeden, 1999.
- [27] D. Focken and R. Stiefelhagen, “*Towards vision-based 3D people tracking in a smart room.*,” in IEEE Int. Conf. on Multimodal Interfaces, 2002, pp. 400–405.
- [28] D. Comaniciu, V. Ramesh, and P. Meer, “*Kernel-based object tracking*,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 564–577, May 2003.
- [29] “*VACE-Video Analysis and Context Extraction*,” <http://www.ic-arda.org>.
- [30] K. Nummiaro, E. Koller-Meier, and L. Van Gool, “*Object Tracking with an Adaptative Color-Based Particle Filter*”, Symposium for Pattern Recognition of the DAGM (2002) 353-360.
- [31] Jack E. Bresenham, “*Algorithm for Computer Control of a Digital Plotter*”, IBM Systems Journal, 4(1):25-30, 1965.
- [32] “*The Bresenham Line-Drawing Algorithm*”, <http://www.cs.helsinki.fi/group/goa/mallinnus/lines/bresenh.html>
- [33] “*Smart Rooms*” <http://vismod.media.mit.edu/vismod/demos/smartroom/>
- [34] J.L. Landabaso, M.Pardàs, J.R.Casas, “*Reconstruction of 3D shapes considering inconsistent 2D silhouettes*”, IEEE International Conference on Image Processing (ICIP2006), Atlanta, GA, October 8-11, 2006