

Seguimiento visual 3D de múltiples objetos combinando filtros de partículas

Pablo Barrera, José María Cañas y Carlos Agüero

Abstract—En el presente artículo describimos un nuevo algoritmo para la estimación muestreada de distribuciones de probabilidad multimodales, como las que aparecen en el problema de seguimiento de múltiples objetos. El algoritmo emplea la combinación de dos filtros de partículas con el fin de eliminar tendencia monomodal que presentan filtros como el de condensación. Los dos filtros se encargan de tareas diferentes: uno mantendrá los objetos conocidos y otro localizará nuevos objetos y evitará la pérdida de los ya conocidos. El sistema ha sido probado con varios experimentos con objetos reales para comprobar su capacidad de seguimiento frente a varios objetos, haciendo énfasis en la influencia de los diferentes parámetros del algoritmo.

Index Terms—computer vision, multimodal estimation, multi-object tracking, 3D, localisation, multi-camera

I. INTRODUCCIÓN

EL seguimiento visual de múltiples objetos es un problema complejo que actualmente se encuentra abierto. Su importancia viene tanto de su utilidad en aplicaciones de videovigilancia, interfaces hombre-máquina o como parte de sistemas de visión más complejos, por ejemplo robots, que emplean el seguimiento para focalizar la atención del sistema sobre uno o varios objetos en concreto.

Entendemos por seguimiento visual la capacidad de localizar un objeto y mantener dicha localización a lo largo del tiempo. Cuando empleamos para ello imágenes, nos podemos enfrentar a dos tipos de seguimiento principalmente, el 2D o el 3D. El seguimiento 3D resulta más complejo dado que es necesario estimar cuál es la profundidad a la que se encuentra el objeto.

Para resolver el problema de localización 3D visual es necesario emplear cámaras. Tradicionalmente se han empleado técnicas de triangulación, que son sencillas y rápidas obteniendo estimaciones precisas de la posición 3D de un objeto. Su principal limitación es que resulta necesario emparejar todos los puntos entre todas las imágenes. El emparejamiento es, en general, un problema complejo y aún más si la separación entre las cámaras es grande.

En nuestro caso, para extraer la información 3D de varias cámaras, proponemos la utilización de un filtro de partículas combinando condensación con muestreo enfatizado.

Todos los autores están en la Universidad Rey Juan Carlos.
E-mail: pablo.barrera.jmplaza@gsyc.es

Este trabajo ha sido parcialmente financiado por el Ministerio de educación (DPI2007-66556-C03) y por la Comunidad de Madrid (S-0505/DPI/0176 y CCG07-URJC/DPI-1694).

El primero estará encargado de del seguimiento de los objetos y el otro de la localización de nuevos objetos. Además, este filtro de partículas no necesita una fase de emparejamiento de puntos. Muestra el espacio de estados y evalúa qué posiciones se corresponden con objetos de verdad según la información visual de las cámaras.

La sección II se centra en trabajos relacionados sobre el seguimiento de varios objetos. La sección III describe el algoritmo de seguimiento propuesto así como sus bases teóricas. La sección IV presenta varios experimentos sobre seguimiento visual de varios objetos que hemos realizado empleando en algoritmo propuesto. Finalmente, recogemos las conclusiones más relevantes de este trabajo y posibles líneas futuras.

II. TRABAJOS RELACIONADOS

El problema del seguimiento visual de un único objeto ha sido resuelto satisfactoriamente por varios trabajos anteriormente, siendo el filtro de condensación [IB98a] una de las aportaciones más relevantes en este campo.

Sin embargo, el seguimiento de varios objetos al mismo tiempo todavía se encuentra abierto. Quizá el cambio más significativo al extender el seguimiento de un objeto al de varios aparece al considerar múltiples soluciones igualmente válidas.

La aproximación más directa a la hora de buscar varios objetos consiste en segmentar los datos de entrada, separando las observaciones que provienen de cada uno de los objetos. Entonces sería posible emplear técnicas de seguimiento de un único objeto. Esta familia de soluciones fue la primera en explorarse con trabajos como los presentados en [Rei79], donde se empleaban varios filtros de Kalman en paralelo. Trabajos más modernos emplean la continuidad del movimiento para realizar una segmentación correcta, [NJS06], o métodos probabilísticos para estimar tanto el número de objetos que hay en la escena como para segmentar las observaciones [ORS04]. El coste computacional de estas aproximaciones crece de forma lineal con el número de objetos, siempre y cuando seamos capaces de segmentar la información de entrada.

Por otro lado, es posible aumentar el tamaño de espacio de estados con el fin de expresar todos los objetos en un único punto de dicho espacio. La distribución de probabilidad usada se conoce con el nombre de JMPD (*Joint Multitarget Probability Density*) [Kas97], [KKH03]. La ventaja de usar un JMPD es que los métodos de búsqueda de soluciones para un único objeto pueden funcionar sin grandes cambios. Su principales limitaciones son que la dimensionalidad del espacio de búsqueda crece enormemente, dificultando búsqueda, y que el

número de objetos a seguir debe conocerse *a priori*, o por lo menos se debe tener una cota superior.

Otras aproximaciones no reducen el problema a un planteamiento monomodal, sino que emplean una distribución de probabilidad multimodal para la localización de los objetos, donde cada modo pertenece a un objeto diferente. La solución no será un único punto de la función de distribución sino varios, uno por objeto. Estas aproximaciones reconducen la tendencia monomodal que presentan los filtros de partículas [KKH03].

Por ejemplo, existen métodos que utilizan la repulsión entre partículas [TC02], [MB00] para distribuir las partículas de forma multimodal sobre el espacio de estados. Aunque pueden funcionar correctamente no se ajustan al marco teórico de Monte Carlo.

Otras aproximaciones respetan los formalismos de ese marco y expresan de forma analítica la distribución de probabilidad objetivo, generalmente como combinación de gaussianas, extrayendo la información necesaria del conjunto de partículas. En lugar de dejar que unas influyan en otras, modifican las técnicas de colocación de las nuevas partículas para conseguir un reparto multimodal algo más equitativo. Por ejemplo, Vermaak [VDP03] propone modelar la distribución objetivo como una mezcla no paramétrica de filtros de condensación. Este método recibe el nombre de *mixture of particle filters*. La distribución combinada tendrá de manera explícita un carácter multimodal, necesario para el seguimiento de múltiples objetos.

En esta línea, la propuesta de Koller-Meier [KMA01] consiste en dividir el problema en dos partes: el seguimiento de objetos y el manejo de la aparición y desaparición de objetos. Empleando una combinación de filtros es capaz de seguir varios objetos. Posteriormente modifica esta combinación atendiendo a las nuevas apariciones y desapariciones de objetos. Para ello introduce nuevas partículas en zonas de inicialización y mantiene las partículas ya conocidas en las zonas de interés.

De manera similar el método propuesto en [MSG⁺05], se centra en mejorar el remuestreo de las partículas. En su caso el paso de remuestreo para cada objeto está aislado del resto de los objetos, con el fin de no eliminar partículas de cada objeto. [OTF⁺04], por su parte, mantiene una estructura similar a la que añade la capacidad de adaptar los modelos de observación según van apareciendo nuevos objetos en la escena, mejorando la capacidad a la hora de colocar nuevas partículas.

III. ALGORITMO PROPUESTO

Nuestro problema es seguir varios objetos en 3D a partir de información visual. El espacio de estados es tridimensional y sobre él definimos la función de densidad de probabilidad (fdp) de ocupación. Idealmente los máximos de esa fdp marcarán la posición de los múltiples objetos que puedan existir. Esa fdp es dinámica ya que los objetos pueden moverse.

El algoritmo de seguimiento que proponemos trata de estimar esa fdp multimodal continuamente, partiendo del flujo de imágenes de la escena. Para ello combina dos técnicas

de Monte Carlo, el filtro de condensación y el muestreo enfatizado, complementando las carencias de cada una con la otra. Ambas son técnicas muestreadas. La idea es mantener muestras en las zonas de interés de la función de distribución *a posteriori* estimada. El muestreo enfatizado proporciona información acerca de dónde se encuentran los objetos, añadiendo las nuevas muestras tan pronto como aparecen. El algoritmo de condensación desecha los errores del muestreo enfatizado, descartando las muestras mal situadas y resituando más afinadamente las muestras cerca de la posición de los objetos. Como intuición podemos pensar que el muestreo enfatizado realiza una búsqueda global y el algoritmo de condensación una búsqueda local sobre las zonas de interés.

La idea de combinar varios filtros de partículas no es nueva, como muestran los trabajos de [IB98b], [KMA01] y [OTF⁺04] entre otros.

A. Fundamentos matemáticos de los filtros de partículas

Los filtros de partículas emplean las aproximaciones numéricas de Monte Carlo para resolver de manera aproximada el problema de filtrado bayesiano a través del uso de números aleatorios.

Supongamos que tenemos una serie de puntos de soporte $\{x_n^i, w_n^i\}_{i=1}^{N_p}$, también llamados partículas, y sus pesos asociados w_n^i , de tal forma que seamos capaces de realizar la aproximación de la siguiente distribución de probabilidad:

$$p(x_n | z_{1:k}) \approx \sum_{i=1}^{N_p} w_n^i \delta(x_n - x_{1:n}^i) \quad (1)$$

Para obtener una buena representación incluso con un número limitado de muestras se deben resolver dos problemas: elegir las partículas y calcular sus pesos asociados. Algoritmos como el de condensación y el de muestreo enfatizado dan respuesta a estas preguntas. Para más información se puede consultar [AMGC02].

B. Combinación de condensación y muestreo enfatizado

El muestreo enfatizado y el de condensación pretenden aproximar, en las zonas de interés, la fdp $p(x_k | z_k)$. Nos referiremos como $\hat{p}_{NS}(x_k | z_k)$ a la aproximación muestral que nos da el muestreo enfatizado y $\hat{p}_S(x_k | z_k)$ a la aproximación obtenida con condensación. Esta última, a diferencia de la primera, tiene la característica de ser secuencial, es decir, la solución en un instante dado se calcula empleando el instante anterior.

Usando estas dos aproximaciones podemos construir una tercera aproximación a la verdadera distribución de probabilidad $p(x_k | z_k)$ como una combinación lineal de la siguiente forma:

$$\hat{p}(x_k | z_k) = (1 - \alpha)\hat{p}_S(x_k | z_k) + \alpha\hat{p}_{NS}(x_k | z_k) \quad (2)$$

$\hat{p}(x_k | z_k)$ estará formada por partículas que proceden de cada uno de los filtros. El parámetro α , $0 \leq \alpha \leq 1$, sirve para ajustar la influencia de cada una de las dos aproximaciones, convirtiendo a \hat{p} en una verdadera distribución de probabilidad.

Los casos en los que α sea igual a 0 o a 1 son los del filtro de condensación y el muestreo enfatizado respectivamente.

La parte no secuencial de la distribución *a posteriori*, \hat{p}_{NS} , puede aportar partículas a \hat{p} sin impedimento alguno. En cada iteración no necesita información sobre la iteración anterior, sino que las coloca únicamente en función de la observación en ese instante de tiempo, z_k . Esto nos permite emplear información, que el algoritmo de condensación no está considerando [RC01], para proporcionar nuevas zonas de interés a \hat{p} independientemente de la evolución anterior del filtro.

La parte secuencial de la distribución *a posteriori*, p_S , es decir, el filtro de condensación, realiza un remuestreo de los pesos desde su estimación de p . Esto permite que los pesos se normalicen todos al mismo valor, evitando el problema de su degeneración. En nuestro caso podemos tomar la estimación de p que hemos creado, \hat{p} . Si se trata de una aproximación válida de p es posible tomar partículas de ella sin interferir en el comportamiento del filtro de condensación. De esta forma se consigue que este filtro emplee muestras que proceden tanto de la iteración pasada (optimización local) como partículas propuestas por la *distribución de propuesta* (búsqueda global).

Para muestrear de \hat{p} usaremos el método de la ruleta. En el funcionamiento habitual los pesos del filtro de condensación y del muestreo enfatizado suman uno. Ahora el factor de corrección α se introduce en los sumatorios. De esta forma obtendríamos la combinación:

$$\hat{p} = \sum w_S^i \delta(x - x_S^i) + \sum w_{NS}^i \delta(x - x_{NS}^i) \quad (3)$$

dónde w_S sumaría $(1 - \alpha)$ y w_{NS} sumaría α . Aplicando la ruleta sobre el conjunto de partículas para obtener la nueva población la probabilidad de tomar una partículas de cada parte es $(1 - \alpha)$ y α respectivamente. Este parámetro nos permite modular el algoritmo entre sus dos extremos: optimización local y obtención de nuevos objetos.

El cálculo de los pesos de cada componente es análogo al empleado en el filtro de condensación y en el de muestreo enfatizado, siguiendo el formalismo de Monte Carlo, dado que dicho formalismo no se ha modificado. El único cambio será añadir el factor de corrección $(1 - \alpha)$ y α en cada caso.

Los pesos de la parte secuencial se calculan así:

$$w_k^i = \frac{p(x_k^i | z_k)}{q_S(x_k^i | x_{k-1}^i, z_k)} = \frac{p(z_k | x_k^i)}{p(z_k)} \cdot w_{k-1}^i \propto p(z_k | x_k^i) \cdot w_{k-1}^i \quad (4)$$

La distribución de muestreo, $q_S(x_k^i | x_{k-1}^i, z_k)$, para este caso es igual al modelo de movimiento, $p(x_k | x_{k-1})$. Para evitar la dependencia de w_{k-1}^i se remuestrea en todos los instantes de tiempo. Esto hace que los pesos finales se simplifiquen a

$$w_k^i = (1 - \alpha) \frac{p(z_k | x_k^i)}{\sum_j^M w_k^j} \quad (5)$$

normalizados a $1 - \alpha$, la parte de probabilidad que les corresponde según la expresión (2).

Para el cálculo de los pesos de la parte no secuencial partiremos del cálculo de los pesos según Monte Carlo:

$$w_k^i = \frac{p(x_k^i | z_k)}{q_{NS}(x_k^i | z_k)} = \frac{p(z_k | x_k^i) p(x_k)}{p(z_k) q_{NS}(z_k | x_k)} \quad (6)$$

Supondremos que $p(x_k)$ es uniforme dado que no tenemos información *a priori* sobre ella. Esto nos permite simplificar la anterior expresión de la siguiente forma:

$$w_k^i = \frac{p(z_k | x_k^i)}{q_{NS}(z_k | x_k)} \quad (7)$$

Según veremos en la sección III-E, nuestra elección de la *distribución de propuesta*, q_{NS} , es tal que todas las partículas que proporciona tienen la misma probabilidad de ser elegidas. De esta manera podemos simplificar el cálculo de los pesos a exactamente la misma expresión que usamos para la parte secuencial, ponderada esta vez por α :

$$w_k^i = \alpha \frac{p(z_k | x_k^i)}{\sum_j^{M^*} w_k^j} \quad (8)$$

$[\{x_k^i, w_k^i\}_{i=1}^N] =$ <i>ESTIMACION_MULTIMODAL</i> ($[\{x_k^i, w_k^i\}_{i=1}^N]$) <ul style="list-style-type: none"> • Remuestrea con ruleta $(1 - \alpha)N$ veces. • Parte secuencial: <ul style="list-style-type: none"> - Repetir con i desde 1 hasta $(1 - \alpha)N$ <ul style="list-style-type: none"> * $x_S^i \sim p(x_k x_{k-1})$ * $w_S^i \propto p(z_k x_k)$ - Normaliza pesos para que $\sum_{i=1}^{(1-\alpha)N} w_S^i = 1 - \alpha$ • Parte no secuencial: <ul style="list-style-type: none"> - Repetir con i desde $(1 - \alpha)N + 1$ hasta N <ul style="list-style-type: none"> * $x_{NS}^i \sim q(x_k x_{k-1}, z_k)$ * $w_{NS}^i = 1$ - Normaliza pesos para que $\sum_{i=(1-\alpha)N+1}^N w_{NS}^i = \alpha$
--

TABLE I
ALGORITMO DE ESTIMACIÓN DE FDP MULTIMODAL.

El algoritmo presentado respeta los formalismos de Monte Carlo y de filtrado bayesiano. Por ello la combinación de las dos partes, la secuencial y la no secuencial, en una única representación también respeta las suposiciones de convergencia de las técnicas de Monte Carlo, esto es:

$$\lim_{N \rightarrow \infty} |I - \hat{I}| = 0 \quad (9)$$

siendo I de la esperanza de una función $f(x)$ sobre la distribución de probabilidad de x e \hat{I} la esperanza calculada desde las muestras. La bondad de las aproximaciones de Monte Carlo se miden en función del error de la aproximación al valor de I . Si usamos el estimador propuesto en la expresión (3) podemos calcular el error de aproximación de la siguiente forma:

$$\hat{I} = (1 - \alpha) \sum w_S^i f(x_S^i) + \alpha \sum w_{NS}^i f(x_{NS}^i) \quad (10)$$

$$|I - \hat{I}| = |I - (1 - \alpha) \sum w_S^i f(x_S^i) + \alpha \sum w_{NS}^i f(x_{NS}^i)| \quad (11)$$

$$= |(1 - \alpha) \left(I - \sum w_S^i f(x_S^i) \right) + \alpha \left(I - \sum w_{NS}^i f(x_{NS}^i) \right)| \quad (12)$$

$$= |(1 - \alpha)(I - \hat{I}_S) + \alpha(I - \hat{I}_{NS})| \quad (13)$$

El límite de este error será la combinación lineal de los límites de $|I - \hat{I}_S|$ e $|I - \hat{I}_{NS}|$ que tienden a cero cuando el número de partículas tiende a infinito. Por lo tanto, el estimador \hat{I} , combinación de otros dos estimadores muestrales, es también un estimador muestral válido según los criterios de Monte Carlo.

C. Modelo de movimiento

Dentro de condensación, el modelo de movimiento que empleamos es el más genérico posible, con el fin de poder adaptarnos sin cambios a cualquier posible movimiento. Supondremos únicamente continuidad en el movimiento, esto es, las zonas cercanas a la posición actual serán más probables que las zonas lejanas. En particular un modelo gaussiano centrado en la posición actual, con una matriz de covarianza diagonal y varianzas iguales en todas las direcciones. La varianza, σ , será el único parámetro libre. Ésta dependerá de la velocidad del movimiento con respecto a la velocidad con la que tomamos nuevas observaciones de la escena.

D. Modelo de observación

Para los experimentos hemos fijado que los objetos a seguir sean pelotas de un determinado color llamativo. Cada partícula representa una posición 3D que podemos proyectar en las cámaras. Si la partícula pertenece al objeto, la zona de la imagen en la que proyecta debería ser del color de dicho objeto. Para nuestro modelo de observación sólo resulta necesario filtrar esta pequeña zona de la imagen para comprobar si la hipótesis resulta verosímil o no. Para ello tomamos una ventana de tamaño W (5×5 píxeles como valor típico) en torno al punto de proyección de esa partícula en la imagen y realizaremos un filtrado de color en esa zona (figura 1).

El valor de probabilidad obtenido proporciona una única medida de lo buena o mala que es la partícula y se calcula como:

$$p(x_k | z_k) = \prod_c \frac{\max(1, r)}{W^2} \quad (14)$$

donde r es el número de píxeles que pasa el filtro en la vecindad y c indica la cámara empleada.

Puede darse el caso que el número de píxeles que pasen el filtro en la ventana de vecindad sea cero. Para evitar esos pozos de probabilidad indeseables daremos un valor mínimo de probabilidad incluso si ningún píxel pasa el filtro de color.

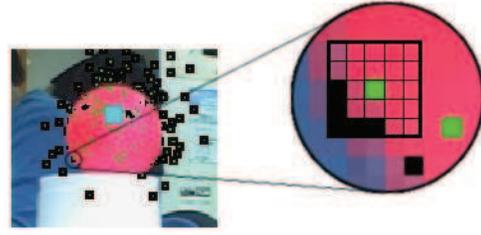


Fig. 1. Modelo de observación con filtro de color.

Para combinar la información entre todas las cámaras que tenemos asumimos independencia entre imágenes.

E. Muestreo enfatizado abductivo

Para colocar nuevas muestras empleando el muestreo enfatizado hemos desarrollado una distribución de propuesta $q_{ns}(x_k | z_k)$ abductiva. Su objetivo es compensar la búsqueda ciega del filtro de condensación. Tomando las imágenes como indicios sobre la posición de los objetos dirigirá la búsqueda hacia esas zonas del espacio. La correcta elección de esta función puede mejorar los resultados dramáticamente sin aumentar, necesariamente, la complejidad del sistema.

En el caso del seguimiento visual en 3D sabemos que los objetos estarán colocados sobre la intersección de las líneas de visión de las cámaras. Podemos tomar esta información a la hora de crear nuevas muestras. Si colocamos todas las muestras sobre la línea de visión de una cámara, todas las partículas proyectarán correctamente, por lo menos, en una cámara. Eventualmente alguna se situará bien en todas las cámaras, indicándonos la verdadera posición del objeto.

En primer lugar se filtra completamente la imagen por color. A continuación se realiza un histograma de color sobre cada eje de la imagen filtrada (figura 2). Obteniendo los máximos locales de los histogramas podemos conseguir las coordenadas x, y de la proyección del objeto. Tras obtener las coordenadas, resulta directo trazar una línea 3D desde el foco de la cámara que pase por ese punto en la imagen. Las nuevas hipótesis se colocarán sobre dicha línea de forma uniforme, entre una distancia mínima y una distancia máxima, es decir en zonas 3D prometedoras. Procediendo de esta forma no perdemos recursos computacionales en zonas sin ningún interés.

Al añadir una función de importancia que tiene en cuenta las observaciones directamente, la convergencia es mucho más rápida. Tan pronto como un objeto sea captado por la cámara el filtro no secuencial comenzará a colocar hipótesis en zonas del espacio compatibles con su proyección en las imágenes.

F. Estimación de la posición del objeto

El filtro de partículas como estimador de fdps no proporciona una estimación de la posición de cada objeto de manera directa. El algoritmo representa los modos de dichas distribuciones agrupando partículas en las zonas con alta probabilidad de contener objetos. Para conseguir la posición concreta de cada objeto se segmenta la distribución de esas

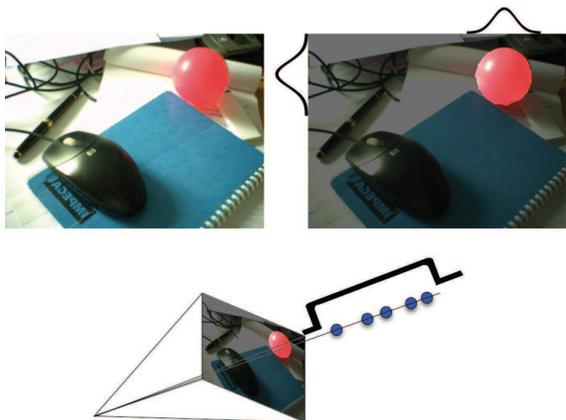


Fig. 2. Extracción de las líneas de visión para la abducción.

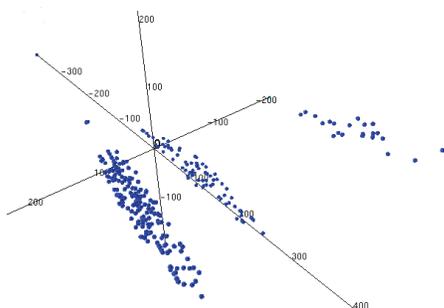


Fig. 3. Representación muestreada de la fdp.

partículas y se calcula la media ponderada de cada grupo. Por ejemplo, en la figura 3 podemos ver cómo se distribuyen las partículas en el espacio de estados, aglutinándose en tres zonas, correspondientes a tres objetos. El algoritmo de segmentación deberá separar los tres grupos que existen. Debe notarse que esta segmentación no es de los datos de entrada, las imágenes, sino de las partículas en el espacio de estados.

El algoritmo que hemos empleado es el EM, o de maximización de esperanza (*expectation maximization*) [Mac03]. Además, usamos el algoritmo iterativo ISODATA (del inglés *Iterative Self-Organizing Data Analysis Techniques*) [BH65], con alguna modificación menor de optimización, para ajustar el número de grupos en el algoritmo EM.

IV. EXPERIMENTOS

Con el fin de validar el algoritmo propuesto hemos realizado numerosas pruebas empíricas sobre una implementación real.

El prototipo emplea dos cámaras calibradas para seguir varias pelotas rosas como las de la figura 4. Las dos cámaras son del modelo iSight con conexión Firewire, con una resolución de 640x480 píxeles a 15 fps. Este prototipo se ejecuta en un ordenador de sobremesa convencional.

Describimos primero la ejecución típica del algoritmo y analizamos experimentalmente características como el reparto de partículas entre objetos, la precisión, el coste computacional o la vivacidad en la incorporación de nuevos objetos al

seguimiento. También estudiamos la influencia del parámetro α , que es central en nuestra propuesta al balancear la combinación entre condensación y el muestreo enfatizado abductivo.



Fig. 4. Localización de cuatro objetos en un determinado instante.

A. Ejecución típica

En el primer experimento el par de cámaras recibe imágenes como las de la figura 4, con cuatro pelotas rosas colgadas de diferentes soportes. Las pelotas se mueven siguiendo una trayectoria pendular sobre el plano XY, cada una de ellas a una velocidad y amplitud diferentes. Cada pelota se encuentra a una distancia diferente de las cámaras. El experimento se ha realizado con 1.000 partículas y el valor de $\alpha = 0,2$.

El sistema es capaz de localizar los cuatro objetos simultáneamente y de obtener vivazmente su posición 3D a lo largo del tiempo. Hemos proyectado en las imágenes de la figura 4 las estimaciones tridimensionales calculadas, mostrándolas como recuadros verdes. Se comprueba que caen correctamente dentro de los cuatro objetos. Las trayectorias seguidas en el eje horizontal pueden verse en la figura 5. Se aprecia que el sistema de seguimiento registra un movimiento ondulatorio que sigue fielmente la trayectoria real con cierto ruido.

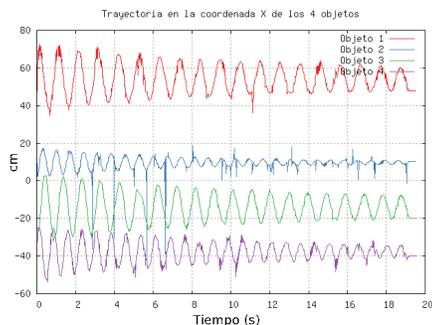


Fig. 5. Seguimiento de cuatro objetos para un movimiento oscilatorio.

B. Reparto de partículas

En el segundo experimento hemos situado dos pelotas estáticas delante del par de cámaras (figura 6) y analizamos el reparto de partículas entre los dos objetos para diferentes valores de α . El mecanismo fundamental para que el sistema sea capaz de estimar dos objetos al mismo tiempo es el reparto de las partículas. Debe asignar un número suficiente de

α	1.000 partículas			2.000 partículas		
	Objeto 1	Objeto 2	Total	Objeto 1	Objeto 2	Total
0,0	77,0%	0%	77,7%	76,5%	0%	76,5%
0,1	8,0%	77,1%	85,1%	7,1%	69,8%	76,8%
0,2	7,9%	71,4%	79,3%	6,6%	67,9%	74,4%
0,4	12,4%	60,9%	73,3%	11,4%	58,6%	69,9%
0,6	15,6%	56,2%	71,8%	16,1%	50,5%	66,6%
0,8	20,0%	51,1%	71,1%	18,0%	46,9%	64,8%
0,9	19,0%	50,1%	69,1%	18,5%	46,0%	64,5%
1,0	20,3%	49,7%	70,0%	18,9%	44,4%	63,3%

TABLE II
REPARTO DEL NÚMERO DE PARTÍCULAS ENTRE DOS OBJETOS ESTÁTICOS
EN FUNCIÓN DEL PARÁMETRO α .

partículas a cada objeto. En el cuadro II se muestra el reparto de partículas entre los dos objetos.

Cuando el parámetro α es cero el sistema únicamente converge a uno de los objetos, el primero que encuentra, situando todas las partículas en él. Al incrementar α , y con ello el número de partículas procedentes del muestreo enfatizado, hay un reparto más equitativo entre los objetos, pues la abducción se encarga de colocar partículas por igual en todos los objetos existentes.

Condensación pura ($\alpha = 0$ en el sistema híbrido) no reparte bien las partículas entre los objetos. Puede representar temporalmente ambigüedad entre varias soluciones posibles pero es incapaz de mantener varios modos en el largo plazo. La nube de partículas se abre hasta que alguna proyecta bien en una de las cámaras. En ese momento el proceso de remuestreo descarta todas las otras partículas para centrarse en esa única posición, hace que se converja únicamente a uno de los dos objetos, abandonando al otro, incluso si se parte de dos poblaciones bien situadas de partículas. La inicialización y la evolución aleatoria fijan qué pelota concreta se localiza. El sistema no saldrá de esa situación degenerada, dado que la solución es estable de por sí, compatible con las imágenes. En esa situación, si aparecen nuevos objetos el sistema los ignorará. En la figura 6 podemos ver la estimación del algoritmo de condensación de los objetos de la escena, representada por el cuadrado verde que aparece sólo sobre la pelota de la izquierda.

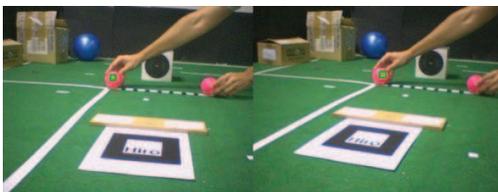


Fig. 6. Resultados de localización de dos objetos.

Aun en el caso que todas las partículas las coloque la función de abducción ($\alpha = 1$), existe una diferencia entre el número de partículas situadas sobre cada objeto. Este reparto asimétrico está causado por el tamaño diferente, entre los objetos, de las zonas que proyectan bien en las imágenes. El tamaño de esa *zona de incertidumbre* asociada a cada objeto, depende del tamaño real del objeto y de su posición relativa con respecto a las cámaras, como muestra la figura 7.

Cuanto mayor sea este área mayor será el número de partículas asignadas al objeto mediante el muestreo enfatizado abductivo. En el caso de la tabla II el “objeto 2” es el que tiene asociado un área de incertidumbre mayor.

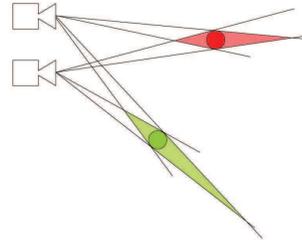


Fig. 7. Áreas de incertidumbre para dos objetos.

Las partículas que caen fuera de las áreas marcadas en la figura 7 obtendrán valores de probabilidad bajos en el modelo de observación, serán descartadas por la segmentación y no influyen en la estimación de los objetos. La tabla II muestra que un 30% de las partículas caen en zonas de poco interés (aunque eso se sabe sólo después de crearlas) y aproximadamente el 70% en zonas buenas.

Como puede apreciarse en la tabla II, la cantidad global de partículas de la población no influye en el reparto entre objetos de las mismas, posiblemente porque la proporción entre las áreas de incertidumbre no varía. Sin embargo, las variaciones en el parámetro α sí modifican el número de partículas sobre cada objeto, aunque nunca baja de cierto mínimo indispensable para cada objeto.

Una vez que mantenemos una población mínima en un objeto seremos capaces de proporcionar una estimación para su posición. El algoritmo de estimación de la posición tan solo necesita unas pocas partículas bien colocadas para proporcionar estimaciones suficientemente correctas. Colocar un número suficiente de partículas sobre cada objeto resulta más relevante que realizar un reparto lo más simétrico posible. Será la parte no secuencial del algoritmo propuesto la que se encargue de mantener un mínimo de partículas en cada objeto. De esa forma se consiguen los dos objetivos buscados: localizar todos los objetos presentes en la escena, reaccionando rápidamente a las nuevas incorporaciones, y mantener una población estable en cada uno de ellos.

C. Incorporación de nuevos objetos

Para estudiar el funcionamiento del sistema frente a la aparición de nuevos objetos se tienen los cinco objetos de la figura 8. Primero aparecen escalonados en el tiempo (en el orden especificado en la figura), y en un segundo experimento aparecen todos ellos a la vez.

En la figura 9 podemos ver el reparto de partículas según van apareciendo los objetos para un valor de α de 0,4. Aun partiendo de una situación ventajosa para uno de los objetos, el algoritmo reparte las partículas entre todos los objetos a medida que van apareciendo. Al ir incrementando el número de objetos, la función de abducción añade nuevas zonas en las que colocar las partículas disponibles. Las partículas se

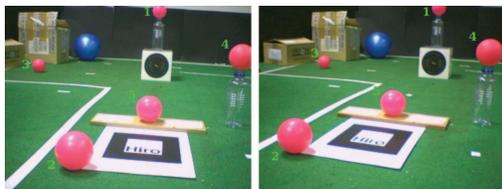


Fig. 8. Colocación inicial de cinco objetos y orden de aparición.

van distribuyendo sobre cada objeto hasta cubrir las cinco posiciones. Al encontrar un objeto nuevo el sistema coloca unas pocas partículas en él. El paso de remuestreo irá colocando más partículas en cada objeto localizado, aumentando el número de partículas rápidamente.

A lo largo del tiempo hay diferencias en el número de partículas asociadas a un mismo objeto. Esto se debe a la naturaleza probabilística del proceso de remuestreo y de la colocación de nuevas partículas, ya sean situadas por el modelo de movimiento o por la función de abducción. Lo importante es que el número de partículas para un objeto no desaparezca.

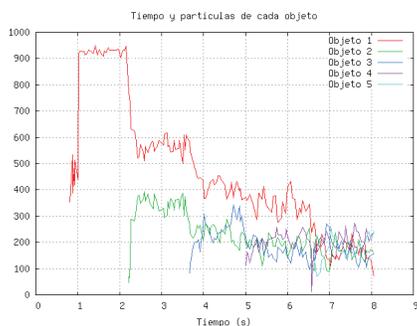


Fig. 9. Reparto de partículas a lo largo del tiempo para cinco objetos.

Cuando partimos de un único objeto y los otros cuatro aparecen a la vez, la aparición simultánea obliga al algoritmo a colocar partículas en varios lugares nuevos a la vez. En la figura 10 podemos ver el reparto de las partículas para $\alpha = 0,4$. Podemos ver cómo las partículas se colocan en todos los objetos tan pronto como aparecen, gracias a la función de propuesta abductiva que detecta los objetos tan pronto como aparecen en las observaciones.

Para ambos experimentos los tiempos de respuesta en incorporar los nuevos objetos son más o menos independientes de α , siempre que nos alejemos del valor 0, que marca la condensación pura y con ello la monomodalidad en el filtro de partículas.

D. Precisión

Los experimentos realizados dentro de una habitación de unos 50 metros cuadrados muestran errores centimétricos en la estimación de posición tridimensional. Este error incluye el debido al algoritmo de seguimiento y el introducido por una calibración inexacta de las cámaras reales. Esta exactitud es suficiente para aplicaciones de videovigilancia o asistenciales.

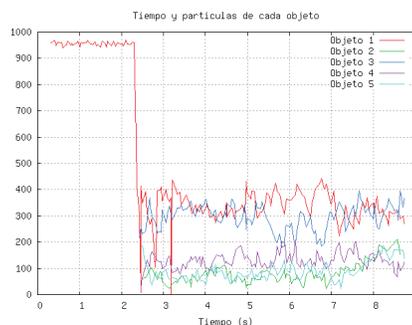


Fig. 10. Reparto de partículas para cinco objetos que aparecen a la vez.

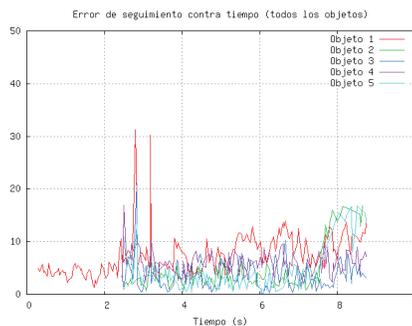


Fig. 11. Error de seguimiento con cinco objetos que aparecen a la vez.

En el experimento con cinco objetos con aparición simultánea los errores de seguimiento siguen la evolución de la figura 11, manteniéndose alrededor de 10 cm.

En un nuevo experimento colocamos sendos objetos a unos 60 y 120 centímetros de las cámaras respectivamente. En el cuadro III se muestran los errores de estimación para los dos objetos, con diferentes valores de α . Se puede apreciar que el error de estimación es sensiblemente mayor para el objeto más lejano, acorde con una calibración imperfecta de las cámaras. También que la precisión no depende significativamente de α . No obstante, hemos observado que la desviación típica del error es mayor para valores de α crecientes. Ésto aconseja elegir valores de α suficientemente lejos de 0 como para no incurrir en monomodalidad y lejos de 1 para no incurrir en errores de estimación altamente variables, con desviaciones típicas altas.

α	Objeto 1 Error (cm)	Objeto 2 Error (cm)	Iteraciones/segundo
0,0	1,97	—	143
0,1	1,92	15,04	32
0,2	2,09	14,01	29
0,4	2,04	16,17	28
0,6	1,98	15,74	24
0,8	1,97	16,75	21
0,9	1,97	16,93	20
1,0	1,95	17,11	20

TABLE III
ERRORES EN CENTÍMETROS PARA CADA OBJETO PARA DIFERENTES VALORES DE α E ITERACIONES POR SEGUNDO.

E. Coste computacional

En cuanto a la velocidad del algoritmo, el sistema funciona en tiempo real puesto que su coste computacional es liviano. Hemos medido las iteraciones por segundo que ejecutaba el algoritmo con diferentes valores α para evaluar su influencia. Los resultados en el cuadro III, muestran que el salto en coste computacional se produce al pasar de $\alpha = 0$ a $\alpha = 0,1$. Esto se debe a que la parte más costosa del algoritmo de seguimiento multiobjeto corresponde a la función de abducción, que requiere un filtrado completo de la imagen. Con $\alpha = 0$ no hay abducción y el algoritmo ejecuta ligero. Para valores de α mayores el incremento del número de partículas abductivas, más costosas que las partículas de condensación, conlleva un descenso paulatino en la velocidad, bajando hasta unas 20 iteraciones por segundo.

V. CONCLUSIONES

En este artículo hemos presentado un método para combinar dos filtros de partículas, uno secuencial (condensación) y otro no secuencial (muestreo enfatizado abductivo), con el fin de seguir en 3D visualmente varios objetos, aunque es generalizable a otros escenarios.

La base del sistema consiste en emplear partículas procedentes de dos filtros de partículas diferentes al mismo tiempo. Las partículas del muestreo enfatizado proporcionan un mínimo de partículas en cada objeto, y el algoritmo de condensación es capaz de seguirlos. Además, la abducción será capaz de localizar nuevos objetos según aparecen en la escena. Por otro lado, la naturaleza secuencial del algoritmo de condensación buscará localmente las mejores estimaciones posibles para cada objeto.

Hemos mostrado, tanto de manera teórica cómo con una serie de experimentos, cómo el sistema puede emplearse para seguir varios objetos al mismo tiempo. El algoritmo presentado no pierde objetos y es capaz de detectarlos tan rápido como aparecen en la escena. Esta es una gran diferencia con respecto a los filtros de partículas convencionales, cuya tendencia monomodal les hace inadecuados para seguir a varios objetos simultáneamente.

Otra ventaja del algoritmo propuesto es que se encuentra dentro del marco teórico de Monte Carlo. Esto permite aplicar todas las garantías sobre convergencia y margen de error de estos métodos a los resultados que obtengamos aplicando nuestro algoritmo. Otras soluciones que se han propuesto con anterioridad no respetaban dicho marco.

La combinación de dos filtros de partículas nos permite tratar la aparición y desaparición de objetos de forma natural y no como un hecho especial. El sistema es capaz de detectar los nuevos objetos igual que detecta los objetos presentes en la escena, por medio de la función de abducción. Tan pronto como aparece un nuevo objeto, la abducción nos permitirá localizarlo rápidamente.

Por último, cabe señalar que la velocidad que obtenemos es suficiente para el funcionamiento del sistema en tiempo real en un ordenador personal corriente.

Como líneas futuras de trabajo podemos indicar la mejora del modelo de observación, con el fin de resolver ciertas

limitaciones que hemos encontrado con el modelo actual y aumentar la precisión tridimensional. Este paso también nos permitirá abordar problemas más complejos visualmente, como puede ser el seguimiento de múltiples personas en un entorno concurrido.

También estamos trabajando en el seguimiento en espacios de estados más complejos, con el fin de representar objetos articulados. Una de las ventajas de los filtro de partículas frente a sistemas con triangulación tradicional es que permiten abordar el problema de los objetos articulados de manera mucho más sencilla.

REFERENCES

- [AMGC02] Sanjeev Arulampalam, Simon Maskell, Neil Gordon, and Tim Clapp. A tutorial on particle filters for on-line non-linear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174 – 188, Febrero 2002.
- [BH65] Geoffrey H. Ball and David J. Hall. Isodata, a novel method of data analysis and pattern classification. Technical Report AD-699616, Standord Research Institute, 1965.
- [IB98a] Michael Isard and Andrew Blake. Condensation – conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):5 – 28, 1998.
- [IB98b] Michael Isard and Andrew Blake. A mixed-state condensation tracker with automatic model-switching. In *Sixth International Conference on Computer Vision*, pages 107 – 112, 1998.
- [Kas97] Keith D. Kastella. Joint multitarget probabilities for detection and tracking. In *SPIE Proceedings, Acquisition, Tracking and Pointing*, Abril 1997.
- [KKH03] Chris M. Kreucher, Keith D. Kastella, and Alfred O. Hero. Tracking multiple targets using a particle filter representation of the joint multitarget probability density. In *Proceedings of SPIE-The International Society for Optical Engineering*, pages 258 – 269. International Society for Optical Engineering, 2003.
- [KMA01] Esther B. Koller-Meier and Frank Ade. Tracking multiple objects using the condensation algorithm. *Robotics and Autonomous Systems*, 34(2-3):93 – 105, 2001.
- [Mac03] David J.C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.
- [MB00] John MacCormick and Andrew Blake. A probabilistic exclusion principle for tracking multiple objects. *International Journal of Computer Vision*, 39(1):57 – 71, 2000.
- [MSG⁺05] Marta Marrón, Miguel A. Sotelo, Juan C. García, David Fernández, and Daniel Pizarro. "xpfc": an extended particle filter for tracking multiple and dynamic objects in complex environments. In *International Conference on Intelligent Robots and Systems*, pages 2474 – 2479 Agosto 2005.
- [NJS06] Hieu T. Nguyen, Qiang Ji, and Arnold W. M. Smeulders. Robust multi-target tracking using spatio-temporal context. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 578 – 585, 2006.
- [ORS04] Songhwai Oh, Stuart Russell, and Shankar Sastry. Markov chain monte carlo data association for general multiple-target tracking problems. In *43rd Conference on Decision and Control*, volume 1, pages 735 – 742, Diciembre 2004.
- [OTF⁺04] Kenji Okuma, Ali Taleghani, Nando De Freitas, James Little, and David Lowe. A boosted particle filter: Multitarget detection and tracking. In *European Conference in Computer Vision*, 2004.
- [RC01] Yong Rui and Yunqiang Chen. Better proposal distributions: Object tracking using unscented particle filter. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Kauai, Hawaii*, volume II, pages 786 – 793, 2001.
- [Rei79] Donald B. Reid. An algorithm for tracking multiple targets. *IEEE Tran. on Automatic Control*, 24(6), 1979.
- [TC02] David Tweed and Andrew Calway. Tracking many objects using subordinated condensation. In *British Machine Vision Conference*, pages 283 – 292, 2002.
- [VDP03] Jaco Vermaak, Arnaud Doucet, and Patrick Pérez. Maintaining multi-modality through mixture tracking. In *International Conference on Computer Vision*, page 1110, 2003.