

4.3 Evaluación.

Uno de los problemas que nos encontramos al definir un método de seguimiento visual es el de como evaluar su funcionamiento. Este hecho es debido a que la medida de evaluación más correcta es la diferencia entre la posición real del objeto y su posición estimada. Sin embargo, en la práctica, no es posible tener las medidas reales de la posición del objeto³.

En la teoría clásica de los métodos de seguimiento[10], la evaluación se realiza utilizando métodos de simulación. Al tener el modelo de movimiento del objeto bien definido y el estado estar compuesto por componentes de posición y velocidad, se pueden generar medidas con diferente cantidad de ruido para estudiar el comportamiento del filtro.

En el caso del seguimiento visual es más complicado realizar una simulación de las medidas debido a que éstas son imágenes. Una alternativa simple es la de anotar de forma manual una secuencia, es decir, marcar sobre cada imagen la posición del objeto. Esta forma de evaluación es costosa por la gran cantidad de imágenes que contiene una secuencia. Si se orienta la evaluación hacia una aplicación concreta del método, es posible definir medidas prácticas de rendimiento[71].

La evaluación del rendimiento de nuestro método la realizaremos en dos partes. Una primera evaluación que se muestra en esta sección, se realizará escogiendo una secuencia de imágenes donde el objeto de interés queda parcial y totalmente ocluido por la escena y por otro objeto, ver Fig. 4.5. Se ha anotado esta secuencia de forma manual para tener las posiciones “reales” del objeto, y se comparan los resultados obtenidos por el algoritmo *iTrack* con el Filtro de Kalman y el Filtro Bayesiano implementado con densidades muestrales. La segunda parte de la evaluación se realizará orientada a aplicación utilizando medidas prácticas de rendimiento y se mostrará en el capítulo siguiente.

Para poder comparar los tres métodos, se ha seguido el siguiente protocolo:

1. La componente de posición horizontal del centro de masas del objeto de interés, x_t , es la escogida para realizar la comparación debido a que sólo se tienen las anotaciones manuales de ésta.
2. El modelo de movimiento se ha aprendido previamente utilizando el método de *maximum likelihood* (ver sección 3.3) y se utiliza en los tres métodos. La decisión de utilizar el modelo dinámico en todos los métodos tiene la misión de intentar demostrar la mejora del rendimiento al utilizar la imagen en el proceso de corrección.
3. Todos los métodos utilizan el mismo proceso de extracción de características de la imagen y se inicializan en la misma posición.

³Podrían utilizarse dispositivos físicos para tener medidas precisas de su posición en el mundo, pero eso no implicaría que puedan ser utilizadas directamente en la imagen.

4. Se ha escogido un problema de seguimiento de un sólo objeto para evitar el proceso de asociación de datos y facilitar la visualización del resultado de los filtros de partículas, ya que la densidad posterior se representará por una función Gaussiana (unimodal).



(a) Escena.



(b) Objeto de interés.



(c) Oclusión parcial con la escena.



(d) Oclusión parcial con otro objeto.



(e) Oclusión completa.



(f) Fin de oclusión.

Figura 4.5: Secuencia de test para la evaluación puntual del método.

En el algoritmo *iTrack* se ha utilizado el método aleatorio de selección de muestras. También se han realizado experimentos con el algoritmo determinista, pero los resultados empeoran. Para la actualización del modelo de apariencia se ha utilizado como regla de ajuste del coeficiente de aprendizaje la Ec. (4.9).

Por último, comentar que para el Filtro Bayesiano y el algoritmo *iTrack* la estimación del estado del objeto se realiza calculando la media de la representación muestral. Para conocer el valor de incertidumbre de la estimación obtenida se calcula la varianza de las muestras a partir de la expresión:

$$\hat{\sigma}_t^2 = \frac{1}{N} \sum_{i=1}^N (s_t^i - \hat{s}_t)^2 . \quad (4.12)$$

En ambos filtros se han usado $N = 1000$ muestras para representar la densidad. Se han realizado también experimentos con mayor número de muestras, el resultado observado es la reducción de la varianza.

El resumen de los resultados obtenidos en los experimentos se muestran en las gráficas de las Figs. 4.6, 4.7 y 4.8. El eje de abscisas muestra el tiempo, y el de ordenadas la posición horizontal en píxeles. En estas gráficas, se puede observar el valor real, representado por círculos ('o'), y la medida obtenida por la función de extracción de características, representada por aspas ('x'). Para facilitar la visualización de las gráficas, el valor de la predicción se muestra previo al instante de tiempo, en forma de asterisco ('*') para la media y como barra la varianza, donde el largo de la barra es 3 veces la desviación estándar. Finalmente, en un instante de tiempo posterior, se muestra el resultado de la estimación en forma de media y varianza, donde la media se visualiza con un signo más ('+') y la varianza en forma de barra.

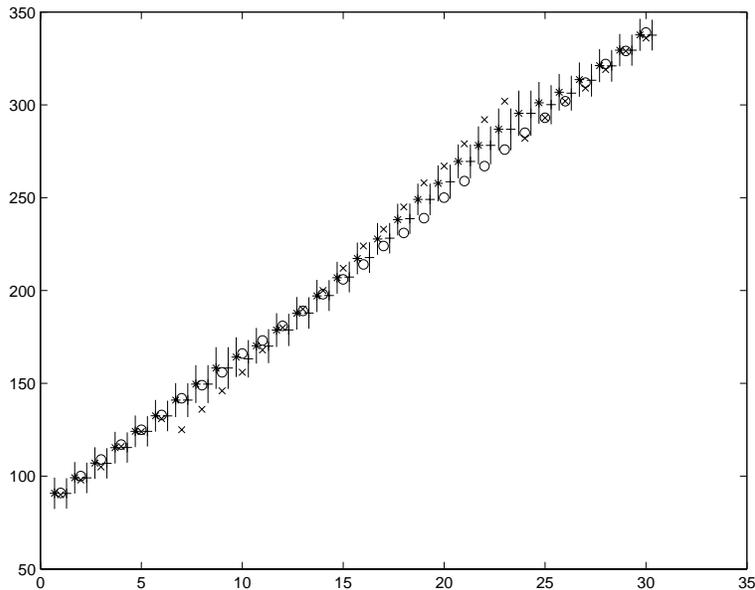


Figura 4.6: Resultados del Filtro de Kalman.

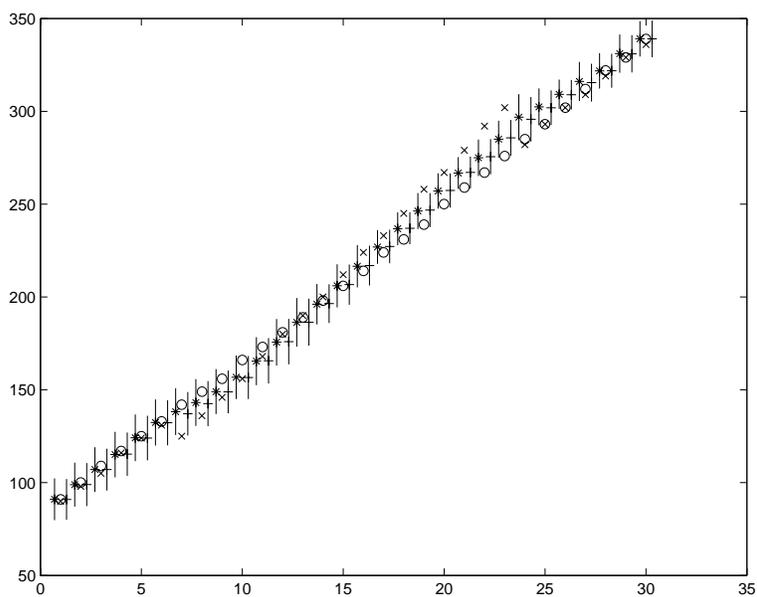


Figura 4.7: Resultados del Filtro Bayesiano implementado con muestras.

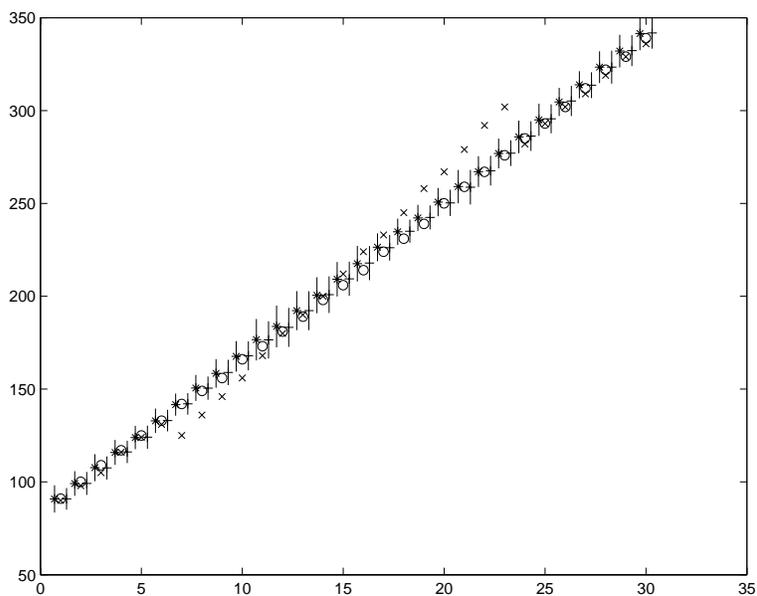


Figura 4.8: Resultados del algoritmo iTrack.

En la Tabla 4.1 se muestran los errores de los tres filtros respecto a la medida real. Se muestran la suma de del valor absoluto de la diferencia entre la posición real y la estimada de cada frame (SAE), y la media de esta diferencia (MAE). Se puede ver en esta Tabla como mejoran los resultados con la introducción del modelo de apariencia en el proceso de corrección del filtro. También se observa como es importante la utilización de un esquema de estimación si comparamos el error de localización con los errores que tienen los filtros.

	SAE	MAE
Localización	914.0000	7.2540
Kalman	680.1809	5.3983
Bayesiano	575.2219	4.5653
iTrack	247.2923	1.9626

Tabla 4.1: Comparación del error obtenido por los diferentes métodos de seguimiento visual en la secuencia de test.

Señalar el parecido que hay entre los resultados del Filtro de Kalman y el Filtro Bayesiano implementado con muestras, ya que, en este caso, ambos se basan en la misma función de corrección. Sin embargo, si comparamos el valor de incertidumbre media de ambos filtros, representado por la varianza, ver Tabla 4.2, se ve la diferencia que hay al utilizar una representación muestral. Como hemos comentado anteriormente, la forma más sencilla de disminuir la varianza de la representación muestral es aumentando el número de muestras. Pero si observamos el valor de incertidumbre que tiene el algoritmo *iTrack*, se observa que el efecto de utilizar una función de corrección con más información, que en nuestro caso es directamente los valores de apariencia de la imagen, también disminuye la incertidumbre de la estimación.

	Incertidumbre media
Kalman	8.1875
Bayesiano	10.5541
iTrack	7.5542

Tabla 4.2: Incertidumbre media de los diferentes estimadores.

Finalmente, el resultado visual de *iTrack* se muestra en la Fig. 4.9. El resultado del método de localización se muestra con una caja negra y el estado estimado como una caja roja, cuyo grosor es proporcional a la incerteza. Se observa el aumento de la incerteza cuando el objeto de interés está completamente ocluido.

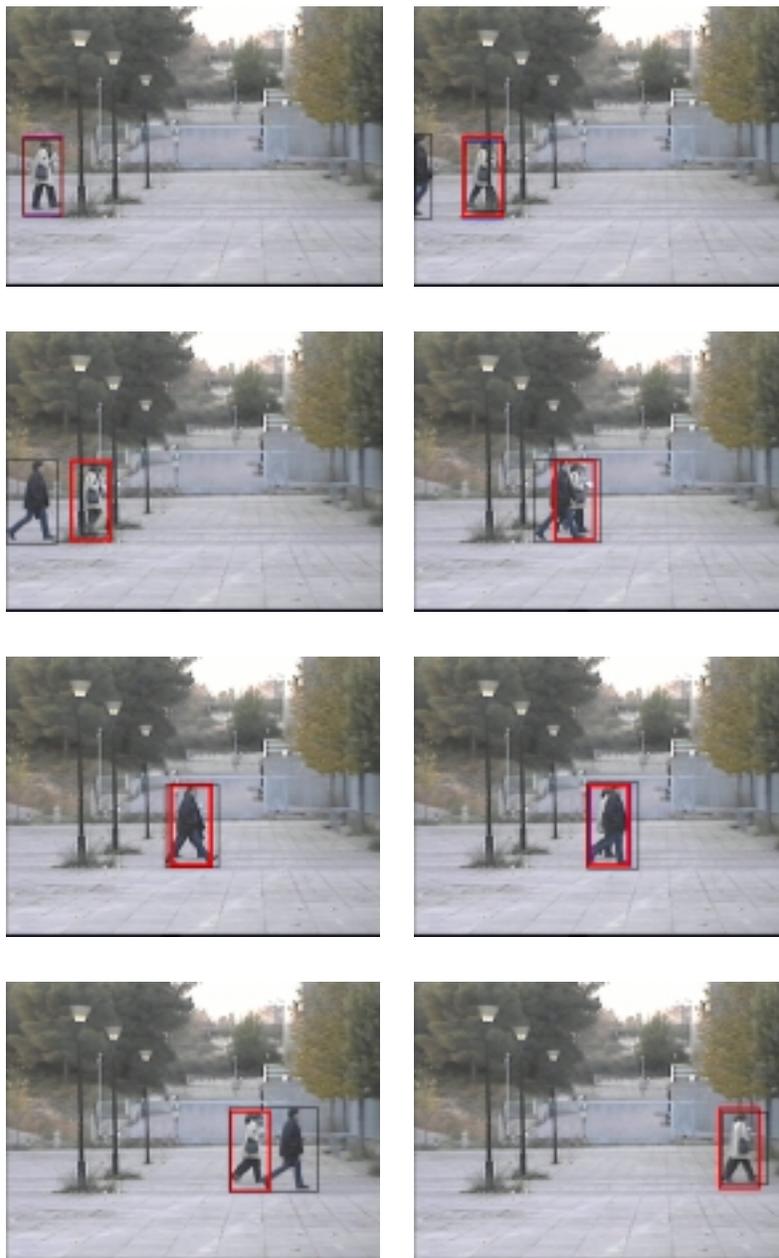


Figura 4.9: Resultados visuales del seguimiento con *iTrack* en la secuencia de test .

4.4 Ampliación para múltiples objetos.

El algoritmo *iTrack* permite una representación multimodal de la densidad posterior. Esto significa que es posible mantener el estado de múltiples objetos con un sólo estimador. A nivel de la representación muestral, supone que dentro del conjunto de muestras, existen subconjuntos pertenecientes a la estimación de un sólo objeto, ver Fig. 4.10. Además, no es necesario un proceso de asociación de datos previo para saber que medidas pertenecen a cada objeto.



Figura 4.10: Representación muestral multimodal para el seguimiento de múltiples objetos. En la figura se muestra la densidad muestral del marginal en x del estado.

Por otro lado, un algoritmo de seguimiento de múltiples objetos debe tener en cuenta dos aspectos importantes: la identificación de los objetos y los eventos. Como eventos, estudiaremos los casos particulares de la aparición, desaparición y oclusión de los objetos de interés. Los estudiaremos desde el punto de vista de la representación que utiliza el algoritmo *iTrack*, para ver cuales de estos casos particulares quedan incluidos en el algoritmo original, y lo extenderemos para tratar los casos no contemplados. El tema de la identificación lo trataremos de forma especial ya que es posible abordarlo de formas muy diferentes. Principalmente, desde el punto de vista del etiquetado individual de cada objeto. Esto nos permitirá, como veremos más adelante, la reconstrucción de la trayectoria seguida por cada objeto.

4.4.1 Identificación.

Para resolver el problema de la identificación, añadiremos a la definición del estado de un objeto una etiqueta única que distinguirá las muestras pertenecientes a los diferentes objetos presentes en la imagen:

$$\mathbf{s}_t^i = (\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t, l) ,$$

donde l es la etiqueta que indica que la posición, la velocidad y el tamaño estimados pertenecen a un objeto determinado. La importancia de la etiqueta es que asociará

el modelo de apariencia del objeto l a las muestras que contengan esta etiqueta para poder calcular la función de *likelihood*. No se incluye el subíndice de tiempo en la variable que representa a la etiqueta puesto que consideramos que no cambia su valor a lo largo del tiempo.

Otra utilidad que encontramos en el estado aumentado por la etiqueta es que nos permite estimar el nuevo estado de cada objeto de forma simple, calculando la media del marginal de cada etiqueta:

$$\hat{s}_{L,t} = \frac{1}{N_L} \sum_{i,l=L} s_t^i .$$

donde N_L es el número de muestras pertenecientes al objeto L . Si no tuvieramos la etiqueta, sería muy compleja la estimación del nuevo estado de cada objeto debido a que no podríamos saber a que objeto pertenece cada muestra. Podría resolverse el problema utilizando el algoritmo EM [6], pero ralentizaría en exceso el proceso de cálculo del posterior. Además, la clasificación no sería sencilla debido a la dimensión del vector de estado.

La inclusión de la etiqueta modifica el funcionamiento del algoritmo básico para identificar los nuevos objetos que aparecen en la imagen. Inicialmente, todas las muestras generadas por el *prior* se etiquetan a 0, y se les asigna un peso proporcional al número de muestras, $\pi = 1/N$. Cuando se calcula la nueva representación del posterior se asignan estas muestras a los objetos existentes en el instante de tiempo anterior, utilizando como técnica de asociación el vecino más cercano. Finalmente, si quedan muestras que no han sido relacionadas con ninguno de los objetos seguidos, se les asigna la etiqueta de un nuevo objeto.

Un problema que nos hemos encontrado en los experimentos realizados para comprobar el funcionamiento del algoritmo con la inclusión de la etiqueta, es el del caso de que la apariencia de uno de los objetos seguidos varíe menos en relación a la apariencia de los otros objetos. Es decir, que la representación muestral puede provocar una tendencia del posterior hacia el objeto con *likelihood* dominante. Para solucionar este problema, una vez calculados todos los pesos de las muestras se normalizan teniendo en cuenta su pertenencia a un objeto:

$$\pi_t^i = \frac{\pi_t^i}{\sum_{j,l=L} \pi_t^j} \cdot c, \quad \text{donde } c = \frac{N_L}{N} . \quad (4.13)$$

4.4.2 Eventos.

Definiremos como **evento** la relación del estado de un objeto con la escena o con el estado de otro objeto. Existen multitud de eventos en los que puede estar un objeto, pero la mayoría de ellos requieren un conocimiento previo de la escena donde ocurre el seguimiento[13]. Uno de los objetivos que nos planteamos en la definición del algoritmo *iTrack* era la no utilización de este conocimiento previo de la escena. Por ello, dentro del conjunto de eventos posibles en los que se puede encontrar cada objeto,

estudiaremos los que es posible identificar sin un modelo simbólico de la escena: la aparición, la desaparición, la oclusión con la escena, la oclusión con otro objeto y la reaparición de objetos. Para poder estudiar las causas que provocan estos eventos, introduciremos el término **grupo** para referirnos a un conjunto de objetos del mundo real que aparecen en la imagen como un sólo objeto, algunos ejemplos de grupos se muestran en la Fig. 4.11. Los eventos que estudiaremos y las causas que los provocan, se muestran en la Tabla 4.3.



Figura 4.11: Ejemplos visuales de grupo de objetos.

Evento	Causas
Aparición	Primera ocurrencia del objeto en la imagen o separación del objeto de un grupo.
Desaparición	Salida del objeto de la imagen.
Oclusión	Desaparición del objeto dentro de la imagen sin la presencia de otro objeto próximo, es decir, la oclusión del objeto con la escena.
Grupo	Oclusión por otro objeto, es decir, la formación de un grupo.
Reaparición	Fin de oclusión con la escena o separación de grupo cuando el objeto había sido seguido individualmente con anterioridad.

Tabla 4.3: Eventos de interés y causas que los provocan.

En la definición de los objetivos del algoritmo *iTrack*, nos planteamos la utilización de una representación muestral y la inclusión de una densidad *prior* de inicialización para poder aplicarlo en un problema de seguimiento de múltiples objetos. Repasemos como el algoritmo, en su definición original (ver Fig. 4.4) maneja los eventos definidos por la Tabla 4.3:

Aparición : la inclusión de un *prior* de inicialización en cada paso del algoritmo permite la generación de nuevas muestras debidas a la primera ocurrencia de un objeto en la escena.

Desaparición : se restringen los valores que pueden tomar las componentes de posición de las muestras a las dimensiones de la imagen. La forma de realizarlo es dando un peso 0 a estas muestras de forma que no aparecerán en la representación del posterior.

Oclusión : el problema que tiene una oclusión con la escena es que el valor del *likelihood* disminuye. Si hay otros objetos en la escena, esto provoca que se generen menos muestras de este objeto, llegando a provocar su desaparición si la oclusión es larga.

Grupo : El comportamiento del algoritmo es idéntico al de la oclusión con la escena. La diferencia fundamental entre ambos eventos es que es posible seguir al objeto ocluido porque se mantiene el seguimiento del grupo.

Reaparición : en el caso de oclusiones o entradas en grupo breves, es posible identificar la reaparición de forma sencilla. Sin embargo, en el caso de largas oclusiones es necesario un proceso de reconocimiento para distinguir este evento del de aparición.

Como es posible comprobar en el estudio particular de cada uno de los eventos, el mayor problema que nos encontramos es el de las oclusiones. Principalmente debido a la actualización del modelo de apariencia, ya que si no se detectan estos eventos, la actualización del modelo de apariencia es errónea, provocando la inestabilidad del algoritmo.

4.4.3 Resultados.

Para probar la extensión del algoritmo *iTrack* para el seguimiento de múltiples objetos se han utilizado 3 secuencias de test. La duración media de las secuencias es de 300 imágenes (12 segundos). En todas las secuencias el *prior*, $p(\mathbf{s}_t)$, se define de forma manual, es decir, generando las muestras de una densidad Gaussiana definida alrededor del estado inicial de cada objeto que aparece en la escena. En la primera secuencia, ver Fig. 4.12 se muestra una secuencia sencilla donde aparecen 2 objetos pero no existen cruces ni oclusiones entre ellos. Con esta secuencia se demuestra la habilidad del algoritmo para manejar una representación multimodal.

En la segunda secuencia, ver Fig. 4.13, se observa como un grupo de objetos se identifica como un sólo objeto ya que inicialmente aparecen juntos en la imagen. Sin embargo, en la última secuencia, ver Fig. 4.14, se muestra como el algoritmo detecta la entrada de un objeto en un grupo y es capaz de recuperarlo cuando termina la oclusión con el otro objeto. Finalmente en la Fig. 4.15 se muestra la evolución temporal del marginal en x de la densidad de probabilidad para la tercera secuencia de test.

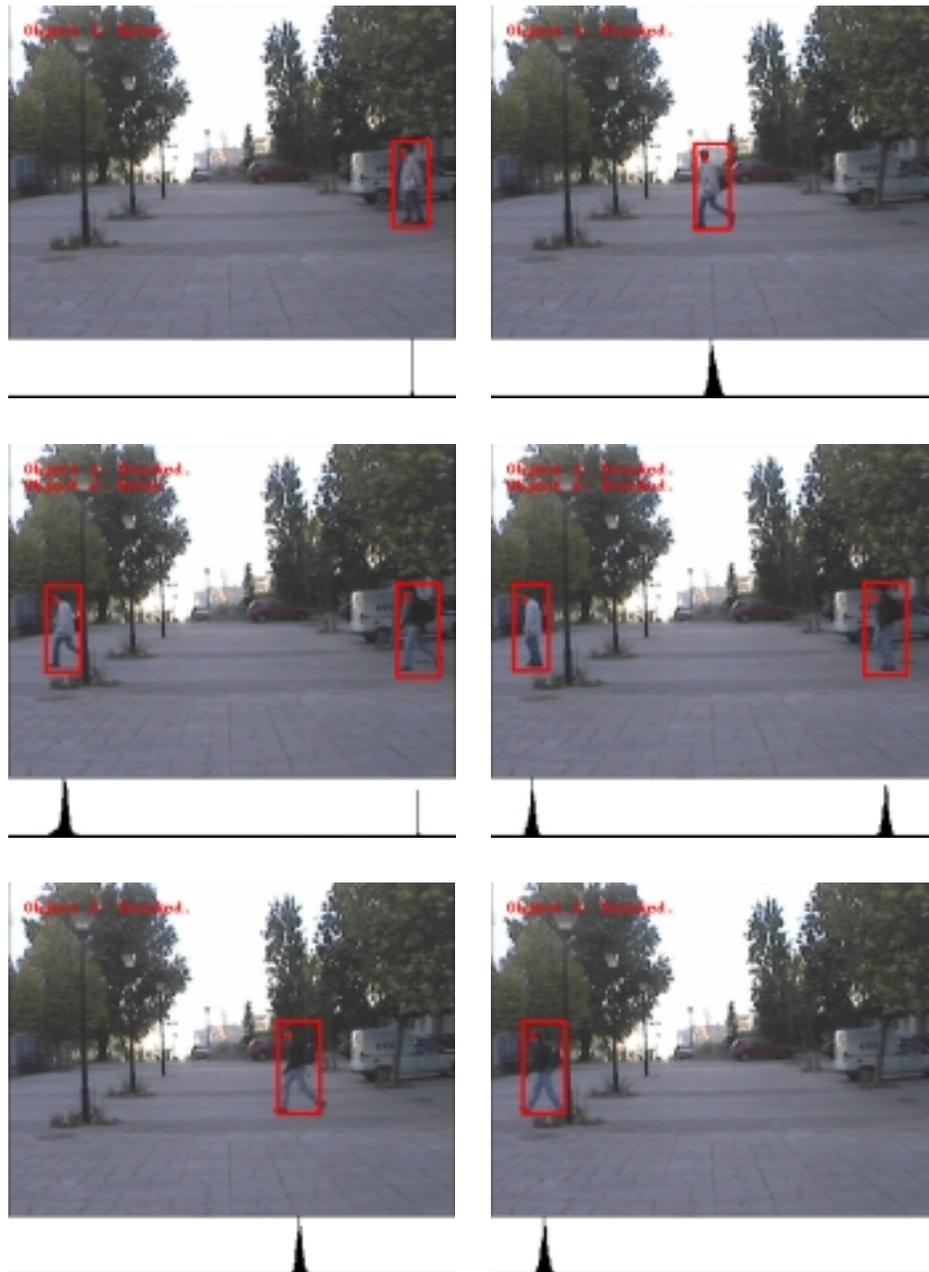


Figura 4.12: Resultados visuales de la extensión del algoritmo *iTrack* con la secuencia de test número 1. En cada imagen se muestra la densidad marginal en x para probar la propagación de densidades multimodales.

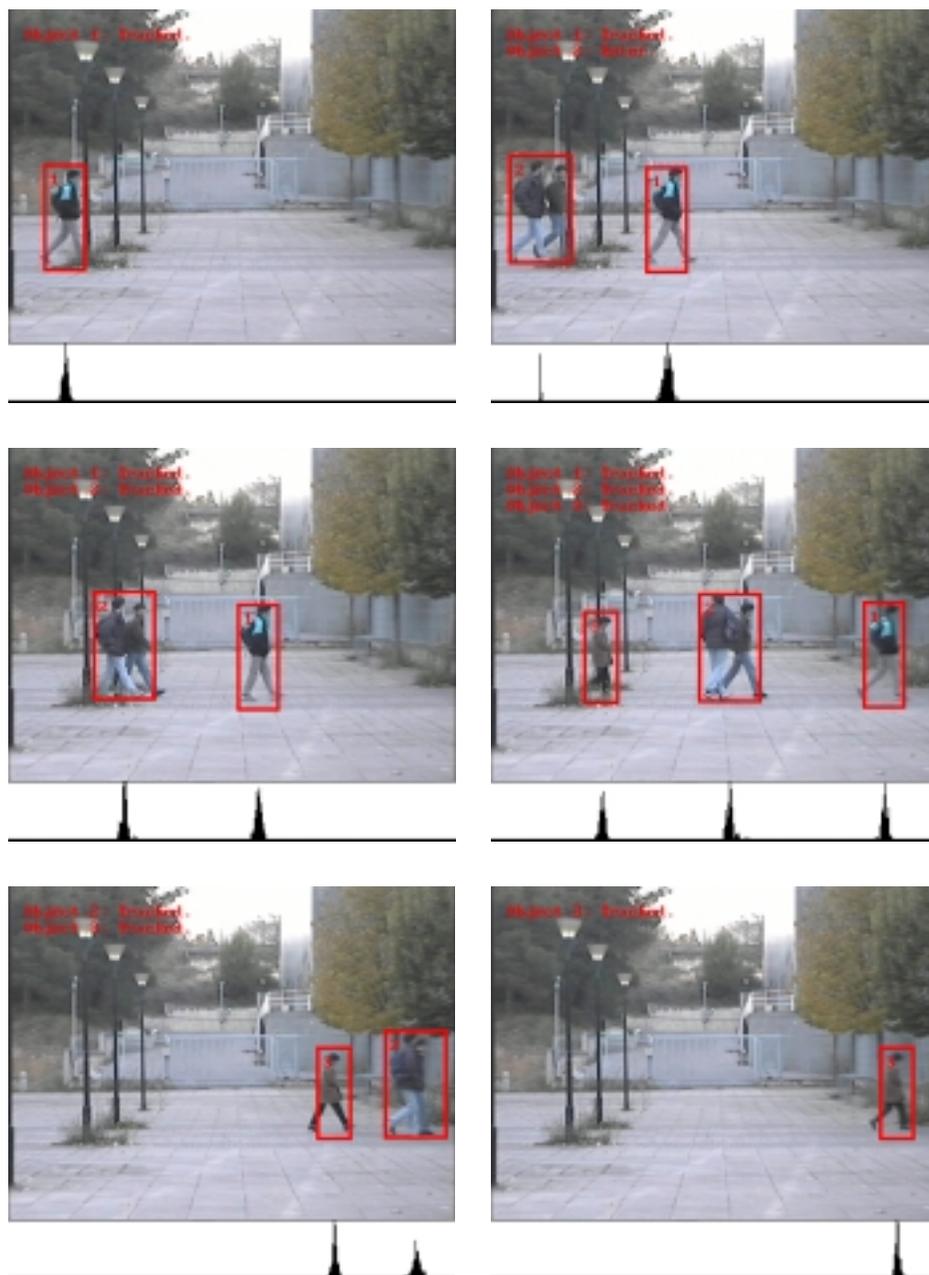


Figura 4.13: Resultados visuales para la secuencia de test número 2.

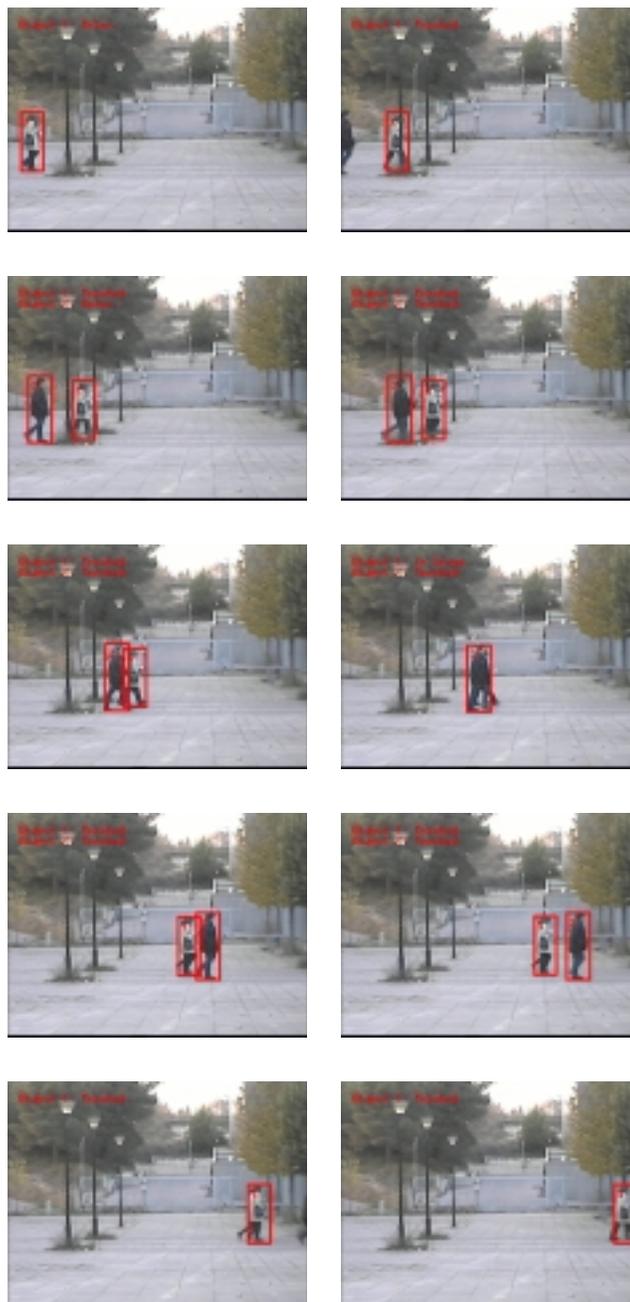


Figura 4.14: Resultados visuales del seguimiento con *iTrack* en la secuencia de test número 3.

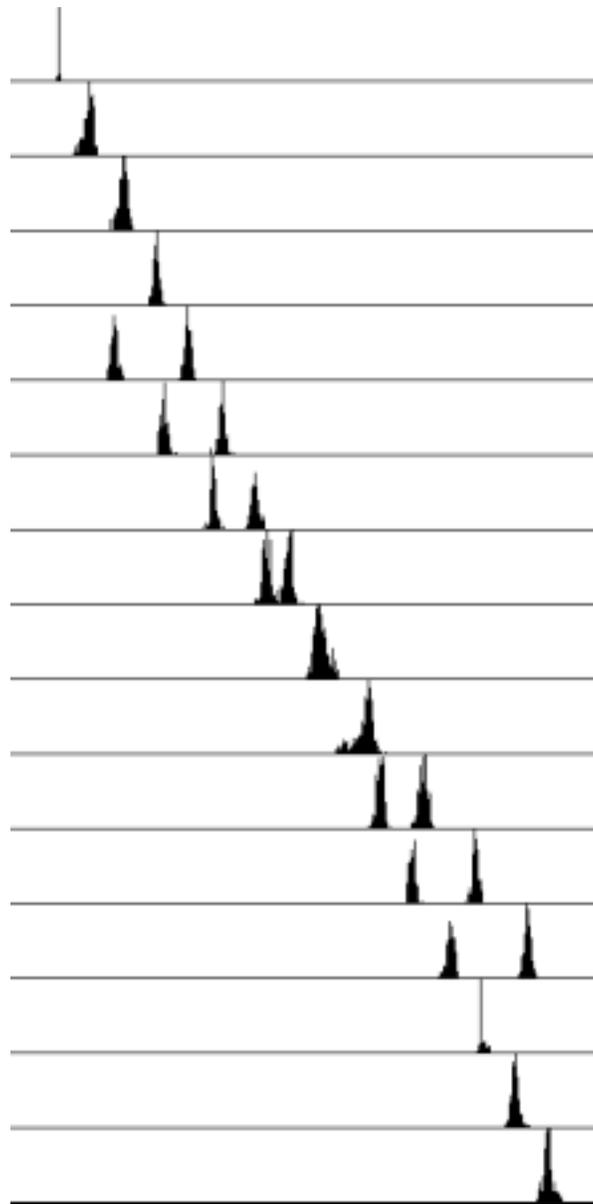


Figura 4.15: Evolución de la densidad marginal en x cada 10 imágenes (cada 0.4s) para la secuencia de test número 3, ver Fig. 4.14.