Capítulo 4

iTrack

En este capítulo se presenta la definición de un algoritmo de seguimiento visual que utiliza como observaciones directamente los valores de intensidad o color de los píxels de la imagen. Este algoritmo, que denominaremos *iTrack*, se basará en la construcción y ajuste de un modelo estadístico de la apariencia del objeto que se desea seguir. El modelo de apariencia permite la definición de una función de *likelihood* robusta a oclusiones parciales o totales del objeto. Se utilizará el esquema de estimación del Filtraje Bayesiano en la definición teórica del método, que se implementará por medio de una representación muestral. Los resultados del algoritmo se mostrarán en comparación con los filtros de estimación más utilizados. Se comprobará como la introducción de los valores de la imagen en el proceso de corrección del algoritmo proporciona un mejor resultado. Finalmente, se ampliará la definición original del algoritmo para poder realizar el seguimiento de múltiples objetos. Esta ampliación es necesaria para resolver los dos problemas básicos de los algoritmos de seguimiento de múltiples objetos.

4.1 Objetivos.

En visión por computador, la imagen es nuestra ventana al mundo real. Dependiendo de los parámetros del sistema de adquisición veremos la proyección de una región 3D del mundo, **escena**, en una retícula 2D, **imagen**. Si se dispone de información adicional de la escena, como por ejemplo la distancia focal del objetivo de la cámara, es posible intentar su reconstrucción 3D a partir de la información extraída de la imagen. Sin embargo, en muchas aplicaciones prácticas, su reconstrucción es compleja porque no se dispone de información adicional o ésta puede contener errores. En estos casos sólo se tiene la imagen como única fuente de información fiable para la reconstrucción de la escena real.

Desde el punto de vista del seguimiento visual, si la secuencia de imágenes es la única fuente de información de que disponemos, el principal objetivo que nos planteamos es la definición de un algoritmo basado directamente en los valores de

intensidad o color de los píxels de cada imagen de la secuencia. Es decir, tomar estos valores como las observaciones que proporciona el sensor. Esto no excluye la posibilidad de que si se dispone de información adicional sea posible utilizarla en un proceso posterior al seguimento para corregir los resultados obtenidos. Por ejemplo, si se calibra la escena y se conoce la correspondencia entre la posición en la imagen y la posición en la escena, es posible traducir los resultados obtenidos del seguimiento en la imagen al mundo real.

Por otro lado, los métodos de estimación se basan en el conocimiento del movimiento del objeto a seguir. En este caso, es necesario un proceso previo de aprendizaje de este movimiento. El segundo objetivo que nos planteamos es que nuestro método sea posible utilizarlo de forma correcta sin realizar este paso previo de aprendizaje de movimiento. Asumiendo un modelo genérico de movimiento suave que ya funciona. Para poder cumplir este segundo objetivo, la función de corrección del estimador ha de utilizar la información contenida en la imagen. La mayoría de metodos basados en estimación, utilizan un proceso previo de extracción de las características de seguimiento del objeto y después usan funciones de corrección basadas en estas características. En este caso, los errores cometidos por el proceso de extracción de características se compensan con un buen modelo de movimiento del objeto.

Finalmente, se escogerá el esquema Bayesiano como esqueleto básico de nuestro método debido a que nos plantearemos el seguimiento de múltiples objetos sin un proceso previo de asociación de datos. Por tanto, la densidad de probabilidad utilizada para representar el estado de los objetos seguidos ha de ser multimodal. Esto implicará la utilización de la representación muestral de densidades de probabilidad.

Resumiendo, nuestro objetivo es la definición de un algoritmo de seguimento visual que utilice un método de estimación que lo haga robusto a errores a pesar de no hacer un aprendizaje previo del movimiento del objeto, que pueda representar el estado de múltiples objetos, y cuyas observaciones son directamente los valores de intensidad o color de cada imagen de la secuencia. Debido a esta última razón, que consideramos la más importante, lo denominaremos *iTrack* (**i**mage-based **Track**ing).

4.2 Definición de *iTrack*.

Consideraremos que el problema del seguimiento visual tiene lugar en una región del mundo y que las observaciones que podemos obtener son imágenes. Por tanto, la imagen nos define la región de interés para nuestro problema. Esto implica en que no estamos interesados en lo que ocurre fuera de la imagen, y que el sistema de coordenadas de nuestro espacio de estados es el sistema de coordenadas de la imagen.

La localización del objeto en la imagen vendrá determinada por una región de interés rectángular que contendrá a todo el objeto, ver Fig. 4.1. Por tanto, para definir el estado del objeto en tiempo t, se utilizará el centro de masas y el tamaño de la región de interés, $\boldsymbol{x}_t = (x_t, y_t)$ y $\boldsymbol{w}_t = (w_t, h_t)$. Inicialmente, asumiremos un

modelo de movimiento simple, de velocidad constante, por tanto, la estimación de la velocidad, $\boldsymbol{u}_t = (u_t, v_t)$, también formará parte del vector de estado. Las unidades de estos componentes son píxels y píxels/frame debido a que sus valores utilizan la imagen como espacio de estados. Para caracterizar el aspecto visual del objeto de interés podría utilizarse el color, la forma o la apariencia. En nuestro caso utilizaremos la apariencia (valores de intensidad o color de los píxels de la región de interés), ya contiene toda la información de la forma visual del objeto[1] disponible en la imagen. Notaremos el componente de apariencia como $\mathbf{M}_t$. Así, el estado del objeto de interés se define por $\boldsymbol{s}_t = (\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{w}_t, \mathbf{M}_t)$, ver Fig. 4.1.

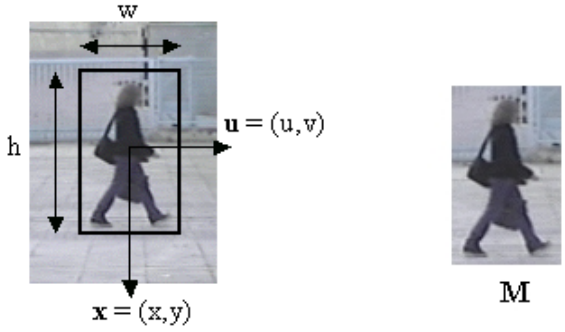

Figura 4.1: Componentes del estado de un objeto.

Como se ha visto en capítulos anteriores, el objetivo de un esquema de seguimiento visual es la estimación del estado del objeto en tiempo t, dados los estados anteriores hasta tiempo $t-1$ y las observaciones hasta tiempo t. Desde un punto de vista probabilístico el problema se plantea como el cálculo de la densidad de probabilidad del posterior sobre los parámetros del estado del objeto en tiempo t dadas las observaciones hasta tiempo t. En nuestro esquema definiremos las observaciones directamente como las imágenes. Por tanto, consideramos la historia de observaciones como la secuencia de imágenes hasta tiempo t, $\mathcal{I}_t = (\mathbf{I}_1, \ldots, \mathbf{I}_t)$. Esto nos lleva al cálculo de $p(\boldsymbol{s}_t|\mathcal{I}_t)$.

4.2.1 Formulación Bayesiana.

En primer lugar desarrollaremos el proceso de cálculo del posterior, $p(\boldsymbol{s}_t|\mathcal{I}_t)$, teniendo en cuenta la descripción de estado realizada, $\boldsymbol{s}_t = (\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{w}_t, \mathbf{M}_t)$, como el cálculo del marginal de la densidad conjunta para toda la historia de estados anteriores:

$$p(\boldsymbol{s}_t|\mathcal{I}_t) = \int p(\mathcal{S}_t|\mathcal{I}_t) d\mathcal{S}_{t-1} \ , \tag{4.1}$$

[1] En el caso de modelar la apariencia directamente a partir de los valores de una imagen color, se podría considerar que este modelo contiene toda la información de color y forma del objeto.

donde $\mathcal{S}_t$, es la historia de estados, $\mathcal{S}_t = (\boldsymbol{s}_1, \ldots, \boldsymbol{s}_t)$. Utilizando la regla de Bayes y la condición de Markov podemos eliminar la dependencia del posterior sobre los estados anteriores a $t-1$:

$$p(\boldsymbol{s}_t|\mathcal{I}_t) \propto p(\mathbf{I}_t|\boldsymbol{s}_t) \int p(\boldsymbol{s}_t|\boldsymbol{s}_{t-1}) p(\boldsymbol{s}_{t-1}|\mathcal{I}_{t-1}) d\boldsymbol{s}_{t-1} \ . \tag{4.2}$$

El desarrollo completo se puede ver en la sección 3.2.2. En este caso, $p(\mathbf{I}_t|\boldsymbol{s}_t)$ es la función de *likelihood*, o corrección; y la integral es el *prior* temporal, o predicción. Es importante observar como el contenido de la integral está formado por dos términos: el proceso dinámico que propaga la densidad del instante de tiempo $t-1$ al instante de tiempo t; el segundo término es el posterior en el instante de tiempo anterior.

4.2.2 Modelo dinámico.

La densidad que debemos modelar para definir el modelo dinámico es:

$$p(\boldsymbol{s}_t|\boldsymbol{s}_{t-1}) = p(\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{w}_t, \mathbf{M}_t|\boldsymbol{x}_{t-1}, \boldsymbol{u}_{t-1}, \boldsymbol{w}_{t-1}, \mathbf{M}_{t-1}) \ . \tag{4.3}$$

Para simplificar esta expresión, asumiremos las siguientes relaciones de independencia entre los componentes del vector de estado:

$$\begin{aligned} &p(\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{w}_t, \mathbf{M}_t|\boldsymbol{x}_{t-1}, \boldsymbol{u}_{t-1}, \boldsymbol{w}_{t-1}, \mathbf{M}_{t-1}) = \\ &\quad p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{u}_{t-1}) p(\boldsymbol{u}_t|\boldsymbol{u}_{t-1}) p(\boldsymbol{w}_t|\boldsymbol{w}_{t-1}) p(\mathbf{M}_t|\mathbf{M}_{t-1}) \ . \end{aligned}$$

Para poder definir un modelo dinámico ad-hoc para las componentes de posición y velocidad es necesario un proceso de aprendizaje previo al seguimiento tal y como se comento con anterioridad. Nuestro método se basa sólo en características que pueden ser encontradas en la imagen. Debido a que normalmente la frecuencia de muestreo de las imágenes es suficientemente alta el objeto no sufrirá un desplazamiento grande en la imagen, por ello se asumirá un modelo de movimiento de velocidad constante:

$$\begin{aligned} p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{u}_{t-1}) &= \eta(\boldsymbol{x}_t - (\boldsymbol{x}_{t-1} + \boldsymbol{u}_{t-1}), \sigma^{\boldsymbol{x}}) \\ p(\boldsymbol{u}_t|\boldsymbol{u}_{t-1}) &= \eta(\boldsymbol{u}_t - \boldsymbol{u}_{t-1}, \sigma^{\boldsymbol{u}}) \ , \end{aligned}$$

donde $\eta(\boldsymbol{\mu}, \sigma)$ denota una densidad Gaussiana de media $\boldsymbol{\mu}$ y desviación estándar σ. Las desviaciones $\sigma^{\boldsymbol{x}}$ y $\sigma^{\boldsymbol{u}}$ se definen de forma empírica dependiendo de las características del sistema de adquisición.

Para modelar los cambios de tamaño del objeto también se utilizará un modelo de cambio suave:

$$p(\boldsymbol{w}_t|\boldsymbol{w}_{t-1}) = \eta(\boldsymbol{w}_t - \boldsymbol{w}_{t-1}, \sigma^w) \ .$$

Para completar el modelo dinámico del vector de estados sólo queda la definición de la evolución del modelo de apariencia, $p(\mathbf{M}_t|\mathbf{M}_{t-1})$. Es posible que el objeto a seguir cambie su apariencia, sin embargo, este cambio se producirá normalmente de forma suave. La aproximación escogida se basa en que la apariencia del objeto no cambia de un frame al siguiente, es decir, la asunción de intensidad constante utilizada por la mayoría de métodos de cálculo de flujo óptico[37]. En términos probabilísticos, la densidad de la evolución del modelo de apariencia viene dada por:

$$p(\mathbf{M}_t|\mathbf{M}_{t-1}) = \delta(\mathbf{M}_t - \mathbf{M}_{t-1}) \ , \tag{4.4}$$

donde $\delta(\cdot)$ es la función delta de Dirac.

4.2.3 Función de *likelihood*.

La función de *likelihood*, $p(\mathbf{I}_t|\boldsymbol{s}_t)$, se utiliza para corregir la predicción realizada por el modelo dinámico. Tal y como se comentó en los objetivos del algoritmo, esta función dependerá de la imagen, es decir, que esta función evalúa el nuevo estado usando directamente los valores de intensidad o color de la imagen en tiempo t. El planteamiento es el siguiente: los valores del modelo de apariencia se compararan con los valores de la imagen en las posiciones definidas por la predicción del estado del objeto.

En términos probabilísticos, debemos definir $p(\mathbf{I}_t|\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{w}_t, \mathbf{M}_t)$. Si seguimos la idea propuesta, lo primero que observamos es que la función de *likelihood* es independiente a la componente de velocidad. De esta forma, la función de *likelihood* nos queda $p(\mathbf{I}_t|\boldsymbol{x}_t, \boldsymbol{w}_t, \mathbf{M}_t)$. El estado del objeto está representado por $\boldsymbol{x}_t$ y $\boldsymbol{w}_t$, esto nos define una región sobre la imagen $\mathbf{I}_t$ que denotaremos como $\mathbf{I}^p$. Para poder comparar el modelo de apariencia $\mathbf{M}_t$ con la región definida por la predicción, $\mathbf{I}^p$, es necesario realizar una transformación afín de la región de la imagen para que tenga las mismas dimensiones que el modelo de apariencia:

$$\mathbf{R} = \mathbf{A}\mathbf{I}^p \ ,$$

donde $\mathbf{A}$ es una matriz de transformación que es posible calcular utilizando los valores del estado del objeto. En nuestro caso, como no se ha incluído un componente de rotación, la matriz $\mathbf{A}$ estará formada sólo por los componentes de escalado. La función de *likelihood* queda:

$$p(\mathbf{I}_t|\boldsymbol{x}_t, \boldsymbol{w}_t, \mathbf{M}_t) = p(\mathbf{R}|\mathbf{M}_t) \ .$$

El método más sencillo de comparación, es la suma de diferencias al cuadrado (SSD)[54]. Sin embargo, éste método no es robusto a la presencia de *outliers*, debido a la forma cuadrática de la función de error[8]. Este efecto es más fácil de explicar en términos de seguimiento visual, en el caso de oclusiones parciales del objeto que provocarían un aumento del error aunque el estado del objeto fuera correcto.

Para evitar este problema utilizaremos una medida en la cual no influyan los errores provocados por posibles oclusiones o cambios bruscos en la forma del objeto. En primer lugar, asumiremos que la apariencia de un píxel en la posición (i, j) sólo depende de su valor de aparencia en la imagen anterior, esto se puede expresar como:

$$p(\mathbf{R}|\mathbf{M}_t) = \frac{1}{N} \sum_{i,j \in \mathbf{R}} p_{ij}(R_{ij}|M_{ij,t}) \ , \tag{4.5}$$

donde N es el número de píxels del modelo de apariencia, y p_{ij} es la probabilidad de que la apariencia del píxel (i, j) de la región $\mathbf{R}$ pertenezca a la distribución del píxel en la posición (i, j) del modelo de apariencia:

$$p_{ij}(R_{ij}|M_{ij,t}) = \eta(R_{ij} - M_{ij,t}, \sigma^M) \ , \tag{4.6}$$

donde $\eta(\cdot)$ es una densidad Gaussiana cuya desviación estándar, σ^M, se utiliza para modelar el ruido del dispositivo de adquisición.

La ventaja de la definición del *likelihood* de esta forma es que es robusto a la presencia de *outliers*, debido a que estos no penalizan la función. Así, permite que el objeto de interés pueda quedar parcial o totalmente ocluído, sin que por ello disminuya de forma drástica su probabilidad.

4.2.4 Modelado estadístico de la apariencia de un objeto.

Como se ha mostrado en la sección anterior, el modelo escogido para la apariencia de un objeto es el de una distribución Normal para cada píxel perteneciente al modelo:

$$p(M_{ij,t}) = N(\boldsymbol{\mu}_{ij,t}, \Sigma_M) \ , \tag{4.7}$$

donde

$$\Sigma_M = (\mathrm{Id})\sigma^M \ .$$

Es posible ver el valor de la desviación estándar, σ^M, como el modelo del ruido del dispositivo de adquisición de la secuencia. Para establecer el valor de la media, $\boldsymbol{\mu}_{ij,t}$, podríamos realizar un aprendizaje previo al proceso de seguimiento. Sin embargo, en la mayoría de casos prácticos no es posible saber la apariencia que tendrán los objetos a seguir. Por otro lado, el significado de la expresión de la evolución de la apariencia

de un objeto, Ec. (4.4), es de que una vez inicializado en la primera aparición del objeto, no cambia de forma brusca durante todo el proceso de seguimiento. Si esto ocurriera así, el sistema no sería robusto a cambios en la apariencia del objeto. Para poder mantener un modelo de apariencia correcto es necesario actualizarlo después de la estimación del estado del objeto.

Una vez estimado el estado del objeto a seguir, $p(\boldsymbol{s}_t|\mathcal{I}_t)$, se actualizará el modelo de apariencia del objeto utilizando un modelo adaptativo de ajuste de la media de la distribución de cada píxel del modelo:

$$\boldsymbol{\mu}_{ij,t} = \boldsymbol{\mu}_{ij,t-1} + \alpha(R_{ij,t} - \boldsymbol{\mu}_{ij,t-1}) \ , \tag{4.8}$$

donde $R_{i,j,t}$ es el valor de apariencia del píxel (i, j) de la región transformada de la imagen según los componentes de posición y tamaño del estado estimado en tiempo t, es decir, de $\mathbf{R}$.

Si analizamos la expresión de ajuste del modelo de apariencia, Ec. (4.8), más concretamente la función del parámetro de ajuste[2], vemos que para un valor de $\alpha = 0$, el modelo no cambia desde su inicialización, mientras que para un valor de $\alpha = 1$, el modelo siempre sería la apariencia del objeto en el instante de tiempo anterior. La forma escogida para establecer este valor ha sido dependiente del tiempo de seguimiento del objeto, es decir, del número de frames que ha sido seguido. Es lógico pensar que aumenta la incertidumbre de la apariencia del objeto cuanto mayor sea el tiempo de seguimiento debido a los posibles errores cometidos por la estimación de su estado. Por tanto, se ha escogido un coeficiente de ajuste variable que va disminuyendo mientras aumenta el tiempo de seguimiento del objeto. La primera posibilidad es utilizar como regla de ajuste:

$$\alpha_t = \frac{1}{t} \ , \tag{4.9}$$

donde t es el tiempo que ha sido seguido el objeto. Aplicada a la actualización de la media, Ec. (4.8), ésta se convierte en una media temporal de la apariencia del objeto. Una segunda posibilidad es utilizar:

$$\alpha_t = e^{-t} \ . \tag{4.10}$$

La interpretación de esta segunda aproximación es que se le da más importancia a los cambios en las primeras estimaciones. Esto viene condicionado porque en la práctica las mejores estimaciones se realizan al comienzo del seguimiento. Ambas posibilidades se han probado en los experimentos del algoritmo.

[2]Que se puede interpretar como un coeficiente de aprendizaje.

4.2.5 Método computacional.

Antes de descibir el algoritmo que permite calcular la Ec. (4.2), repasaremos las condiciones impuestas por el modelo dinámico y la función de *likelihood*:

1. El proceso dinámico se descompone en tres procesos independientes: uno para la posición, que dependerá de la estimación de la posición y velocidad anterior; otro para la velocidad, que dependerá del estado de velocidad en tiempo $t-1$; y el último, para el tamaño del objeto, que sólo depende de su valor en el instante de tiempo anterior. Además, todos estos procesos son independientes a la evolución del modelo de apariencia:

$$\begin{aligned} p(\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{w}_t, \mathbf{M}_t | \boldsymbol{x}_{t-1}, \boldsymbol{u}_{t-1}, \boldsymbol{w}_{t-1}, \mathbf{M}_{t-1}) = \\ p(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}, \boldsymbol{u}_{t-1}) p(\boldsymbol{u}_t | \boldsymbol{u}_{t-1}) p(\boldsymbol{w}_t | \boldsymbol{w}_{t-1}) p(\mathbf{M}_t | \mathbf{M}_{t-1}) \ . \end{aligned}$$

2. La función de *likelihood* no depende de la estimación de velocidad:

$$p(\mathbf{I}_t | \boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{w}_t, \mathbf{M}_t) = p(\mathbf{I}_t | \boldsymbol{x}_t, \boldsymbol{w}_t, \mathbf{M}_t) \ .$$

La expresión final del posterior aplicando estas restricciones queda:

$$\begin{aligned} p(\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{w}_t, \mathbf{M}_t | \mathcal{I}_t) \propto \\ p(\mathbf{I}_t | \boldsymbol{x}_t, \boldsymbol{w}_t, \mathbf{M}_t) \int & \ p(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}, \boldsymbol{u}_{t-1}) p(\boldsymbol{u}_t | \boldsymbol{u}_{t-1}) p(\boldsymbol{w}_t | \boldsymbol{w}_{t-1}) p(\mathbf{M}_t | \mathbf{M}_{t-1}) \\ & p(\boldsymbol{x}_{t-1}, \boldsymbol{u}_{t-1}, \boldsymbol{w}_{t-1}, \mathbf{M}_{t-1} | \mathcal{I}_{t-1}) d\boldsymbol{x}_{t-1} d\boldsymbol{u}_{t-1} d\boldsymbol{w}_{t-1} d\mathbf{M}_{t-1} \ . \end{aligned} \tag{4.11}$$

Utilizaremos el esquema del filtraje de partículas, tal y como se muestra en la sección 3.5, para mantener una buena representación del posterior. Además este esquema nos permitirá realizar la estimación del estado de múltiples objetos ya que por medio del Filtro de Partículas es posible representar una densidad multimodal.

En nuestro método, el estado del objeto, $\boldsymbol{s}_t$, se representará por un conjunto de muestras $\{\boldsymbol{s}_t^i\}$ donde $i = \{1, \ldots, N\}$. En cada muestra se incluirán las componentes de posición, velocidad y tamaño, es decir, $\boldsymbol{s}_t^i = (\boldsymbol{x}_t^i, \boldsymbol{u}_t^i, \boldsymbol{w}_t^i)$. Este hecho es debido a que no es factible tener un conjunto de muestras de modelos de apariencia y a que éste es posible construirlo y actualizarlo a partir de los valores estimados de posición y tamaño.

Utilizando el modelo dinámico y la función de *likelihood* descritos anteriormente, describiremos el método de cálculo de la representación del posterior en forma de los pasos típicos de los filtros de estimación, predicción y corrección.

Predicción:

Se dispone de la representación muestral de la densidad del estado del objetivo en tiempo $t-1$, que viene dada por:

$$\{\boldsymbol{s}_{t-1}^i\} \quad i = 1, \ldots, N \ ,$$

donde el superíndice indica el número de muestra, y el subíndice el instante de tiempo. A cada muestra, se le aplica el modelo dinámico:

$$\begin{aligned} \boldsymbol{x}_t^{i,-} &= \boldsymbol{x}_{t-1}^i + \boldsymbol{u}_{t-1}^i + \boldsymbol{\xi}_x^i \ , \\ \boldsymbol{u}_t^{i,-} &= \boldsymbol{u}_{t-1}^i + \boldsymbol{\xi}_u^i \ , \\ \boldsymbol{w}_t^{i,-} &= \boldsymbol{w}_{t-1}^i + \boldsymbol{\xi}_w^i \ , \end{aligned}$$

donde $\boldsymbol{\xi}_x^i$ es un vector aleatorio escogido de una distribución Normal de media 0 y matriz de covarianza diagonal Σ^x, lo mismo para ξ_u^i y ξ_w^i, con Σ^v y Σ^w respectivamente. Es posible interpretar el paso de predicción como desplazar la densidad de probabilidad según un modelo de velocidad constante más un componente estocástico que provoca el aumento de la incertidumbre del estado, ver Fig. 4.2.

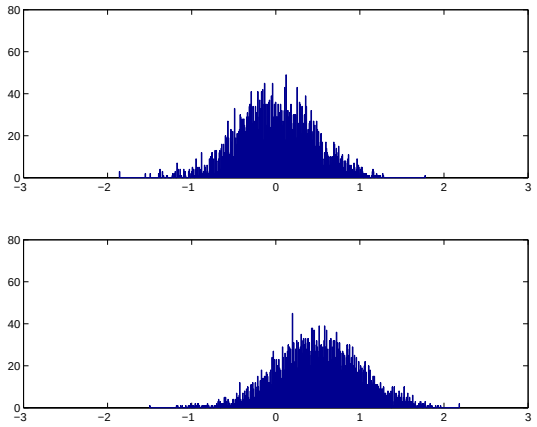

Figura 4.2: Desplazamiento estocástico de una densidad muestral.

Corrección:

La corrección es simple, la predicción actúa como *prior*, y tenemos que convertirlo en posterior. Esto se realiza evaluando la predicción de cada muestra por medio de la función de *likelihood*:

$$\pi_t^i = p(\mathbf{I}_t | \boldsymbol{x}_t = \boldsymbol{x}_t^{i,-}, \boldsymbol{w}_t = \boldsymbol{w}_t^{i,-}, \mathbf{M}_{t-1}) \ ,$$

donde π_t^i se puede interpretar como el peso que cada muestra. Así, la representación del posterior queda:

$$\{(\boldsymbol{s}_t^{i,-}, \pi_t^i)\} \ .$$

El objetivo de la corrección es el de eliminar las muestras menos probables y aumentar las más probables por medio de las observaciones. El efecto es estrechar la densidad muestral cerca de la estimación más probable, ver Fig. 4.3.

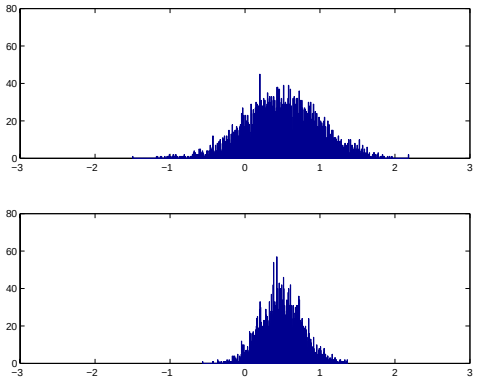

Figura 4.3: Aplicación de la corrección a las muestras del paso de predicción.

Los pesos hallados en este paso se utilizan para construir la representación del posterior. El procedimiento consiste en seleccionar con reemplazamiento N muestras del conjunto $\{\boldsymbol{s}_t^{i,-}\}$, utilizando los pesos, π_t^i, como probabilidad de escoger una muestra. Como vimos en el capítulo anterior, esto se consigue a partir de la densidad acumulada, utilizada como función de distribución para generar las nuevas muestras. De esta forma, obtenemos la representación muestral del posterior en tiempo t: $\{\boldsymbol{s}_t^i\}$.

Estado inicial y reinicialización:

La expresión general de la definición del método, Ec. (4.11), es recursiva debido a que se incluye la estimación del estado anterior condicionado a las medidas realizadas hasta ese instante, $p(\boldsymbol{s}_{t-1}|\mathbf{I}_{t-1})$. Denominaremos esta densidad como **prior temporal**. Sin embargo, también es posible tener en cada instante de tiempo una función de densidad de probabilidad que no tenga en cuenta ninguna medida sobre la imagen. Denotaremos esta función, $p(\boldsymbol{s}_t)$, como densidad **prior**. Con la generación de muestras del *prior* en cada iteración del proceso se consiguen dos cosas: permitimos la inicialización de nuevos objetos a seguir, indispensable para el seguimiento de múltiples objetos; y permitimos la reinicialización del objeto a seguir si hay errores graves en el proceso de estimación como por ejemplo una oclusión completa. El algoritmo completo se muestra en la Fig. 4.4.

iTrack

La representación del posterior en tiempo $t-1$ viene dada por un conjunto de muestras, $\{\boldsymbol{s}_{t-1}^i\}$, donde $i = \{1, \ldots, N\}$. También se conoce la forma de la densidad *prior*, $p(\boldsymbol{s}_t)$.

Generar la muestra i-ésima de las N que representarán el posterior en tiempo t como sigue:

1. **Predicción:** Generar un número aleatorio, α entre 0 y 1 con distribución uniforme,
   (a) Si $\alpha < r$ utilizar el *prior* de inicialización, $p(\boldsymbol{s}_t)$, para generar $\boldsymbol{s}_t^{i,-}$.
   (b) Si $\alpha \geq r$ aplicar el modelo dinámico a la muestra $\boldsymbol{s}_{t-1}^i$:

$$\boldsymbol{s}_t^{i,-} = p(\boldsymbol{s}_t | \boldsymbol{s}_{t-1} = \boldsymbol{s}_{t-1}^i) \ ,$$

   que es lo mismo que calcular:

$$\begin{aligned} \boldsymbol{x}_t^{i,-} &= \boldsymbol{x}_{t-1}^i + \boldsymbol{u}_{t-1}^i + \xi_x^i \ , \\ \boldsymbol{u}_t^{i,-} &= \boldsymbol{u}_{t-1}^i + \xi_u^i \ , \\ \boldsymbol{w}_t^{i,-} &= \boldsymbol{w}_{t-1}^i + \xi_w^i \ , \end{aligned}$$

2. **Corrección:** Calcular el peso de $\boldsymbol{s}_t^{i,-}$ aplicando la función de *likelihood* de la Ec. (4.5):

$$\pi_t^i = p(\mathbf{I}_t | \boldsymbol{x}_t = \boldsymbol{x}_t^{i,-}, \boldsymbol{w}_t = \boldsymbol{w}_t^{i,-}, \mathbf{M}_{t-1}) \ .$$

Una vez generadas todas las muestras, normalizar los pesos de forma que $\sum_i \pi_t^i = 1$, y contruir las probabilidades acumuladas:

$$\begin{aligned} c_t^0 &= 0 \ , \\ c_t^i &= c_t^{i-1} + \pi_t^i \quad \forall i = 1, \ldots, N \ . \end{aligned}$$

Utilizar los valores de las probabilidades acumuladas para generar las muestras que representan el posterior en tiempo t, $\{\boldsymbol{s}_t^i\}$.

Estimar el nuevo estado del objeto calculando la media de la nueva representación muestral:

$$\hat{\boldsymbol{s}}_t = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{s}_t^i \ .$$

y utilizar el nuevo estado para actualizar el modelo de apariencia.

Figura 4.4: Algoritmo *iTrack*.