

Seguimiento visual de personas usando métodos integrativos

por

Elizabeth Guevara Martínez

Tesis

Presentada al Programa de Graduados en Electrónica, Computación, Información y
Comunicaciones

como requisito parcial para obtener el grado académico de

**Maestro en Ciencias
con Especialidad en Sistemas Inteligentes**

Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Monterrey

Monterrey, N.L., Diciembre de 2003

Seguimiento visual de personas usando métodos integrativos

por

Elizabeth Guevara Martínez

Tesis

Presentada al Programa de Graduados en Electrónica, Computación, Información y
Comunicaciones

como requisito parcial para obtener el grado académico de

**Maestro en Ciencias
con Especialidad en Sistemas Inteligentes**

Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Monterrey

Monterrey, N.L., Diciembre de 2003

**Instituto Tecnológico y de Estudios Superiores de
Monterrey
Campus Monterrey**

**División de Electrónica, Computación, Información y Comunicaciones
Programa de Graduados**

Los miembros del comité de tesis recomendamos que la presente tesis de Elizabeth Guevara Martínez sea aceptada como requisito parcial para obtener el grado de **Maestro en Ciencias con Especialidad en Sistemas Inteligentes**.

Comité de Tesis:

Dr. José Luis Gordillo Moscoso
Asesor de la tesis

Dr. Carlos M. Hinojosa Espinosa
Sinodal

M.C. Silvia Patricia Mora Castro
Sinodal

Dr. David Garza Salazar
Director del Programa de Graduados en Electrónica,
Computación, Información y Comunicaciones

Diciembre de 2003

Contenido

1	Introducción	1
2	Seguimiento utilizando la estructura cinemática	7
2.1	Modelo general de estructura	9
2.2	Seguimiento frontal	11
2.2.1	Modelo virtual, modelo cinemático y animación	13
2.3	Seguimiento con profundidad	13
3	Algoritmo de desplazamiento medio	19
3.1	Notación	19
3.2	Adquisición del modelo del objeto de interés	21
3.2.1	Estimación de <i>densidad de núcleo</i>	22
3.2.2	Distribución modelo del objeto de interés	24
3.3	Coficiente de similitud	25
3.4	Vector de desplazamiento medio	27
3.5	Algoritmo de seguimiento	30
3.6	Actualización del modelo del objeto de interés	31
3.7	Adaptación de escala	32
4	Implementación y experimentos	35
4.1	Implementación	35
4.2	Seguimiento con profundidad	36
4.2.1	Medición del error de posición	38
4.2.2	Análisis de resultados	39
4.3	Seguimiento frontal	44
4.3.1	Medición del error de posición	45
4.3.2	Análisis de resultados	45
5	Conclusiones	55
A	Modelo geométrico de perspectiva	59
A.1	Notación matemática	59
A.2	Transformación de coordenadas	60
A.2.1	Proyección sobre el eje Y	61
A.2.2	Proyección sobre el eje X	64

B	Expansión de Taylor del coeficiente de Bhattacharyya	67
C	Interfaz computacional	69
C.1	Archivos	70
C.2	Captura	70
C.3	Seguimiento	71
C.3.1	Seguimiento	71
C.3.2	Seguimiento con perspectiva	72
C.3.3	Seguimiento frontal	76

Lista de Figuras

1.1	Modelo general de estructura jerárquica del cuerpo humano	2
1.2	Método de seguimiento visual de una persona con base en la estructura cinemática del cuerpo humano	3
2.1	Modelo general del cuerpo humano	8
2.2	Modelos de abstracción del cuerpo humano para el seguimiento frontal y con profundidad de la persona	8
2.3	Posiciones de las entidades $\mathbf{p}_i(t)$ y $\mathbf{p}_{i+1}(t)$ en los tiempos t y $t + 1$. . .	10
2.4	Posición de la cámara en el ambiente, para el seguimiento frontal	11
2.5	Modelo jerárquico para el seguimiento de los movimientos de una persona frente a la cámara.	12
2.6	Posiciones del codo y de la mano en los tiempos t y $t + 1$	12
2.7	Elementos y ejes referenciales del modelo virtual	13
2.8	Secuencia de movimientos de una persona frente a la cámara	14
2.9	Posición de la cámara en el ambiente, para el seguimiento con profundidad	15
2.10	Modelo jerárquico para el seguimiento de una persona caminando	15
2.11	Posiciones del tronco y la cabeza en los tiempos t y $t + 1$	16
3.1	Diagrama general del algoritmo de desplazamiento medio	20
3.2	Histograma de una imagen en escala de grises	22
3.3	Núcleo de Epanechnikov	23
3.4	Estimación de densidad de una región de la imagen	25
3.5	Valores del coeficiente de Bhattacharyya correspondientes a la ventana de seguimiento seleccionada en la figura 3.4. El valor inicial del coeficiente se representa con un círculo y el valor máximo se denota mediante un asterisco.	26
3.6	Ejemplo de la aplicación del algoritmo de desplazamiento medio en un conjunto de puntos	29
3.7	Trayectoria definida por los puntos obtenidos mediante el vector de desplazamiento medio aplicado en la imagen 10 y 15 de la secuencia, después de 6 iteraciones.	29
3.8	Diagrama del algoritmo de desplazamiento medio	31
3.9	Relación de escala entre dos posiciones diferentes de la persona en el ambiente	33
3.10	Esquema del ambiente utilizado para la escala	33

4.1	Interfaz computacional para el seguimiento con perspectiva	36
4.2	Interfaz computacional para la representación virtual de los movimientos de una persona	37
4.3	Medidas del ambiente en el que se realizaron los experimentos	38
4.4	Trayectoria establecida para el recorrido de las marcas	38
4.5	Secuencia de imágenes que muestra los resultados del seguimiento de una persona (<i>Prueba 1</i>)	41
4.6	Gráfica de los errores correspondientes al cálculo de la posición utilizando la información del seguidor de la cabeza, del tronco y de los pies.	43
4.7	Medidas del ambiente captado por la cámara en el seguimiento frontal .	44
4.8	Posición de inicio para la selección de las ventanas de seguimiento . . .	45
4.9	Secuencia de imágenes que muestran el seguimiento e imitación de los movimientos realizados por una persona frente a la cámara	47
4.10	Secuencia de imágenes que muestran el seguimiento e imitación de los movimientos realizados por una persona frente a la cámara	48
4.11	Trayectoria generada por el movimiento del torso. La gráfica muestra las posiciones correspondientes a las imágenes analizadas de la secuencia y el resto de los datos se obtienen interpolando cada 5 imágenes	51
4.12	Ángulos definidos por el movimiento del torso	52
4.13	Ángulos definidos durante el movimiento de la cabeza	52
4.14	Ángulos estimados durante el movimiento del codo izquierdo	53
4.15	Ángulos definidos por el movimiento de la mano izquierda	53
A.1	Triángulo $\triangle abc$	59
A.2	Representación de un ángulo	60
A.3	Esquema del montaje de la cámara y ejes de referencia establecidos para el plano del ambiente	60
A.4	Imagen obtenida por la cámara	61
A.5	Proyección del espacio de trabajo en la lente de la cámara	62
A.6	Consideración de la altura de la persona ph en la posición sobre el eje Y	64
A.7	Posición de la persona sobre el eje X en el área de trabajo	65
C.1	Ventana principal de la interfaz computacional del sistema de seguimiento	69
C.2	Opciones del menú <i>Archivo</i>	70
C.3	Opciones del menú <i>Captura</i>	71
C.4	Opciones del menú <i>Seguimiento</i>	72
C.5	Opciones del menú <i>Seguimiento</i> sin consideraciones de estructura . . .	72
C.6	Opciones del menú <i>Seguimiento</i> con perspectiva	73
C.7	Ventana de selección de las partes del cuerpo de interés	73
C.8	Ventana de opciones para el seguimiento con profundidad	74
C.9	Ventanas de configuración de parámetros y de posiciones en el ambiente	75
C.10	Ventana de selección para mostrar las gráficas de trayectorias reales . .	76
C.11	Opciones del menú <i>Seguimiento</i> frontal	77
C.12	Modelo virtual de la persona	77

Lista de Tablas

4.1	Valores de los parámetros del sistema utilizados en los experimentos. . .	37
4.2	Posiciones obtenidas de acuerdo al seguidor de la cabeza	39
4.3	Posiciones obtenidas de acuerdo al seguidor del tronco	39
4.4	Posiciones obtenidas de acuerdo al seguidor de los pies	40
4.5	Errores correspondientes a las posiciones obtenidas mediante el seguidor de la cabeza	40
4.6	Errores correspondientes a las posiciones obtenidas mediante el seguidor del tronco	42
4.7	Errores correspondientes a las posiciones obtenidas mediante el seguidor de los pies	42
4.8	Errores promedio de posiciones	43
4.9	Posiciones en pixeles obtenidas en el seguimiento de la cabeza y el torso	46
4.10	Posiciones en pixeles obtenidas en el seguimiento del codo y la mano izquierda	46
4.11	Posiciones en pixeles obtenidas en el seguimiento del codo y la mano derecha	49
4.12	Errores correspondientes a las posiciones de la cabeza y el torso	49
4.13	Errores correspondientes a las posiciones del codo y la mano derecha .	50
4.14	Errores correspondientes a las posiciones del codo y la mano izquierda .	50
4.15	Error promedio correspondiente a la estimación de la posición de las partes del cuerpo de interés.	51

Capítulo 1

Introducción

El seguimiento de objetos en movimiento se ha estudiado ampliamente en el campo de la Visión Computacional. Sin embargo, la mayoría de las investigaciones se han enfocado en el estudio del movimiento y la estructura de objetos rígidos debido a su simplicidad. Un objeto rígido es aquél en el cual la distancia en tres dimensiones entre cualquier par de puntos del mismo no cambia con el paso del tiempo, lo cual facilita el seguimiento del objeto a través de una secuencia de imágenes. A diferencia del seguimiento de un objeto rígido, cuando se pretende seguir a una persona el problema se vuelve más complejo, ya que el cuerpo humano no es rígido, ni plano, es deformable y se compone de varios grados de libertad [Song 00]. A pesar de su dificultad, el seguimiento de personas es un tema de investigación que ha generado gran interés debido a sus numerosas aplicaciones, las cuales se dividen en tres áreas principales: vigilancia, control y análisis [Moeslund 01].

En el área de vigilancia algunas aplicaciones se basan en la detección del movimiento de uno o más sujetos, y su seguimiento a través de la escena con el fin de evaluar si se ha cometido algún delito. Un ejemplo clásico son los sistemas de vigilancia de estacionamientos y de supermercados [Siebel 01]. En el área de control se utiliza la captura del movimiento para realizar interfaces para juegos, ambientes virtuales, animación, cuartos inteligentes o aplicaciones de robótica móvil [Wren 97]. La tercer área está relacionada con el análisis detallado de los datos de movimiento capturados, para utilizarse en estudios clínicos como el diagnóstico de pacientes de ortopedia, o para ayudar a atletas a entender y mejorar su desempeño [Chaudhari 01, Pers 00]. Con estas aplicaciones se busca mejorar la interacción hombre-máquina con el fin, a largo plazo, de generar máquinas capaces de interactuar de forma inteligente con los seres humanos.

Tradicionalmente, para el seguimiento de objetos móviles se han empleado diversos métodos, los cuales se pueden dividir en dos grupos: los métodos derivativos y los métodos integrativos. Los métodos derivativos detectan el movimiento mediante un análisis en los cambios de intensidades en la imagen (derivada espacial) [Tomasi 91, Lipton 98] y en los cambios de intensidades que se presentan de imagen a imagen (derivada temporal) [Fleet 92]. Por otra parte, los métodos integrativos se basan en

características de área; se enfocan en el conjunto de píxeles que conforman al objeto y utilizan técnicas estadísticas para realizar el seguimiento [Comaniciu 00, Isard 98].

Algunas de estas técnicas se han aplicado al seguimiento específico de una persona. Sin embargo, la mayoría de estos métodos realizan el seguimiento sin considerar la estructura del cuerpo humano y sus restricciones cinemáticas, por lo que ignoran información importante para disminuir la ambigüedad durante el proceso de correspondencia, y evitar la pérdida del objeto de interés debido a cambios de iluminación, ruido, oclusiones, etc. Además, seguir las partes del cuerpo de la persona, en lugar de realizar un seguimiento global, facilita la recuperación de estructura, lo cual es de gran utilidad para aplicaciones de simulación y análisis del movimiento.

El objetivo de esta tesis es desarrollar un método basado en técnicas integrativas que permita modelar, localizar y seguir a una persona utilizando una cámara conectada a la computadora. Se propone el seguimiento de entidades significativas del cuerpo humano considerando sus restricciones cinemáticas, con el fin de recuperar información de estructura para reproducir o trazar los movimientos realizados por la persona. Debido a la necesidad de seguir varias entidades del cuerpo, se aplica el método integrativo de desplazamiento medio [Comaniciu 00], localizando diferentes procesos de seguimiento sobre múltiples objetos.

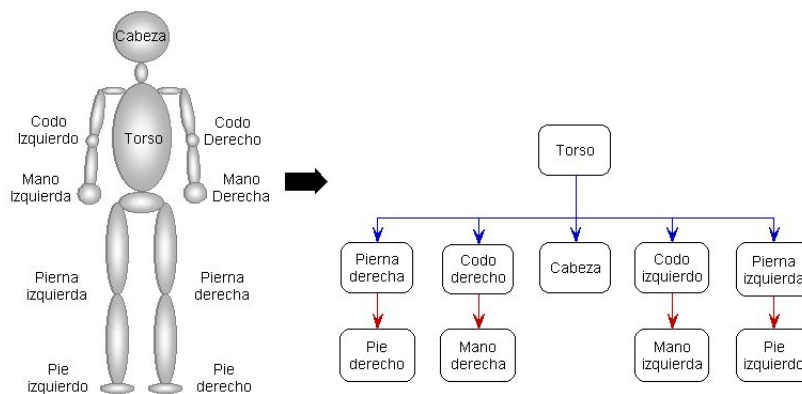


Figura 1.1: Modelo general de estructura jerárquica del cuerpo humano

El modelo utilizado para definir la estructura articulada del cuerpo consiste en un árbol de dependencia cinemática donde cada elemento corresponde a una parte del cuerpo de la persona. A partir del elemento raíz o padre, en este caso ubicado en el torso, se aplican restricciones cinemáticas (distancias, ángulos, etc.) para predecir la posición probable de las entidades dependientes, de acuerdo a un modelo jerárquico de estructura. Gracias al proceso de predicción relativa, se facilita el seguimiento y se asegura que las posiciones obtenidas cumplen con las restricciones cinemáticas del cuerpo humano [Olvera 00].

Para la parte experimental, se definieron dos niveles de abstracción con base en el modelo general de estructura jerárquica que se muestra en la figura 1.1. En el primero se incluyen: la cabeza, el tronco y los pies; el objetivo es el seguimiento de una persona caminando con el propósito de determinar su posición real en el ambiente y trazar su trayectoria a partir de un modelo de proyección de la cámara. El segundo nivel comprende la cabeza, el torso, los codos y las manos; se utiliza para la reproducción de los movimientos efectuados por una persona frente a la cámara, por medio de un modelo virtual en dos dimensiones.

La figura 1.2 muestra el esquema general del método de seguimiento propuesto. Dicho método consiste en seleccionar las partes de interés del cuerpo de la persona, determinar sus características distintivas (distribución estadística y modelo de estructura), obtener la posición relativa de cada entidad con respecto al elemento inmediato superior en el modelo de estructura cinemática y finalmente, a partir de la posición relativa, aplicar el algoritmo de seguimiento de desplazamiento medio.

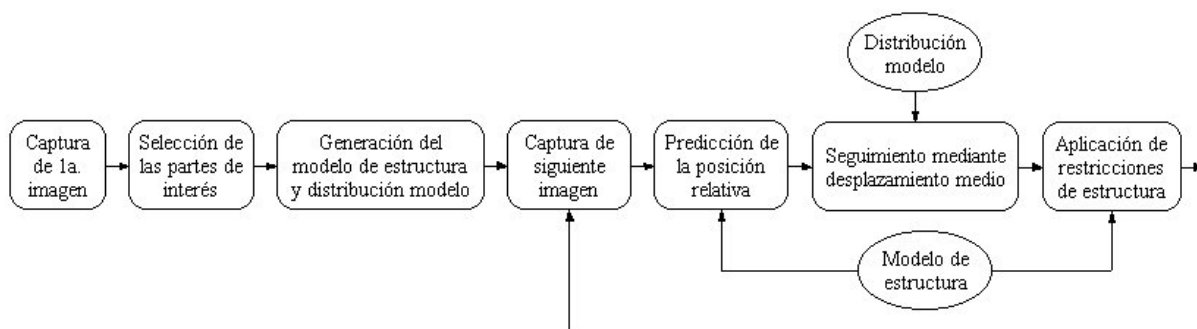


Figura 1.2: Método de seguimiento visual de una persona con base en la estructura cinemática del cuerpo humano

De acuerdo al esquema, el proceso de *Captura de la 1a. imagen* consiste en adquirir la imagen inicial de la secuencia. Las imágenes se obtienen de una cámara de video o de una secuencia predeterminada de imágenes. El propósito es generar una representación computacional de la escena, que sea procesada para obtener la información requerida.

Sobre la primera imagen de la secuencia se realiza la *Selección de las partes del cuerpo de interés* que se van a seguir; y a partir de esta información se lleva a cabo la *Generación del modelo de estructura y el cálculo de la distribución modelo*. El *Modelo de estructura* se utiliza como base para la predicción y reproducción del movimiento de la persona. Por otro lado, para determinar la similitud, al aplicar el algoritmo de desplazamiento medio, se requiere la *Distribución modelo* de cada una de las entidades significativas. Las distribuciones estadísticas se generan a partir de cualquier característica discriminante como pueden ser características de color, de textura o las generadas al obtener el gradiente de la imagen. De manera específica, en este trabajo se utilizan características de color.

Después de obtener los modelos sobre la primera imagen de la secuencia, se *Captura la siguiente imagen* y se inicia el seguimiento.

Para describir la dependencia cinemática entre las entidades del cuerpo que se van a seguir, se utiliza el modelo de estructura jerárquica. Tomando como referencia la posición de la parte del cuerpo inmediata superior, en el árbol de dependencia cinemática, el módulo de *Predicción de la posición relativa* determina la localización del elemento dependiente, considerando que permanece estático con respecto a la articulación superior. Por ende, la posición de la entidad de inicio o base del modelo jerárquico, en este caso el torso, se estima de forma independiente aplicando únicamente el algoritmo de desplazamiento medio.

A partir de la posición determinada mediante la predicción relativa, se aplica el *Seguimiento mediante el algoritmo de desplazamiento medio* para encontrar la nueva localización de cada entidad significativa. Así, la predicción facilita el seguimiento, al iniciar en la posición más probable de acuerdo al modelo de estructura. El algoritmo de seguimiento mediante desplazamiento medio consiste en obtener la distribución estadística del objeto de interés (distribución modelo) y del objeto candidato, para comparar enseguida ambas distribuciones mediante un coeficiente de similitud y estimar la posición del objeto de interés a partir del *vector de desplazamiento medio*. Este vector se calcula iterativamente hasta encontrar la distribución más similar a la del objeto modelo [Comaniciu 00].

Con base en la posición obtenida en la etapa de predicción relativa, se genera un arco cuya circunferencia determina las posiciones válidas de cada entidad significativa. Dicho arco es la base para la *Aplicación de restricciones de estructura* del cuerpo humano. Las principales restricciones son: las distancias relativas (la distancia del codo a la mano por ejemplo) así como los ángulos máximos entre las diferentes partes del cuerpo. Las restricciones de estructura aumentan la robustez del seguidor al evitar posiciones imposibles de realizar por el ser humano. Por tanto, la búsqueda de la posición final de un elemento se restringe a posiciones válidas sobre el arco de circunferencia que describe el movimiento del citado elemento. Posteriormente se continúa con la siguiente imagen de la secuencia.

El algoritmo de desplazamiento medio se basa en un modelo de distribución estático. Sin embargo, la distribución original del objeto de interés cambia a lo largo del tiempo debido a cambios de iluminación, rotaciones, etc. Siendo así, se modificó el algoritmo de desplazamiento medio para ajustar dinámicamente la distribución de probabilidad del objeto de interés.

Por otra parte, durante el seguimiento de una persona al caminar, el individuo se puede alejar o acercar a la cámara, provocando cambios de escala y en consecuencia variaciones en las distribuciones de las ventanas de seguimiento. El algoritmo original de desplazamiento medio supone que el tamaño de las ventanas permanece fijo, de forma que no considera los cambios de escala del objeto de interés, dificultando el seguimiento.

Se propuso, por tanto, explotar la información de perspectiva del ambiente de trabajo para determinar el factor de escala de acuerdo a la posición de la persona en el ambiente.

Para validar el método propuesto, se desarrolló un programa para el seguimiento en tiempo real de una persona. La aplicación recibe como entrada la selección de las partes del cuerpo a seguir y proporciona la posición relativa de la persona, en función de sus elementos en el ambiente. Además, el sistema toma en cuenta la información de estructura para reproducir los movimientos efectuados por el sujeto.

La principal característica del método radica en efectuar el seguimiento de los movimientos ejecutados por una persona, sin la necesidad de utilizar marcas especiales en el cuerpo, dando como resultado una técnica práctica y poco costosa. Otros sistemas utilizan seguidores ópticos [Paradiso 97] o magnéticos [Anisfield 00], que requieren un sistema de cableado o bien sistemas en los que la persona debe usar ropa especial. Asimismo, se generó un procedimiento para adecuar el seguidor a cambios de escala, para el caso específico de contar con la información de perspectiva del ambiente captado por la cámara.

Para describir el método propuesto en detalle, el Capítulo 2 presenta la estructura cinemática que se emplea tanto para el seguimiento y trazo de la trayectoria seguida por una persona caminando (cabeza, tronco y pies), como para el seguimiento y reproducción de los movimientos realizados por la persona frente a la cámara (cabeza, torso y brazos). El Capítulo 3 describe el método integrativo de desplazamiento medio utilizado en la etapa de seguimiento. El Capítulo 4 muestra los detalles de la implementación del programa, así como las pruebas experimentales que permitieron validar el método de seguimiento. Por último, el Capítulo 5 presenta las conclusiones obtenidas durante el desarrollo de la tesis. Se incluyen además varios anexos: el apéndice A describe el modelo geométrico de perspectiva, el apéndice B presenta la expansión de Taylor del coeficiente de Bhattacharyya, mientras que el apéndice C contiene el manual de usuario de la interfaz computacional desarrollada.

Capítulo 2

Seguimiento utilizando la estructura cinemática

Este capítulo describe los detalles del método de seguimiento visual basado en la estructura cinemática del cuerpo humano. El método propuesto considera la estructura del cuerpo y sus restricciones cinemáticas para disminuir la ambigüedad durante el proceso de correspondencia y recuperar información necesaria para reproducir los movimientos efectuados por una persona. La información de estructura es de gran utilidad para aplicaciones de simulación y análisis del movimiento en el campo de la Visión Computacional.

El problema de capturar información necesaria para reproducir una secuencia de movimientos mediante un modelo virtual, se ha resuelto en otros trabajos, colocando sensores físicos [Paradiso 97, Anisfield 00] o marcas [Herda 00] sobre la persona que se desea imitar. Sin embargo, debido al acondicionamiento del ambiente y a la ubicación de los elementos utilizados en el seguimiento, dichas técnicas resultan complicadas y costosas. La ventaja principal del método propuesto es que no requiere sensores, ni marcas especiales para aplicar un conjunto de reglas de estructura cinemática y recuperar información para simular los movimientos de la persona.

El modelo de la estructura articulada del cuerpo humano consiste en un árbol de dependencia cinemática, donde cada nodo corresponde a un elemento del cuerpo de la persona [Olvera 00]. A partir del nodo raíz o padre, en este caso ubicado en el torso, se aplican restricciones cinemáticas (distancias, ángulos, etc.) para predecir la posición probable de las entidades dependientes, de acuerdo a un modelo jerárquico como el que aparece en la figura 2.1. Gracias al proceso de predicción relativa, se valida el espacio de búsqueda del seguidor y se asegura que las posiciones obtenidas cumplen con las restricciones cinemáticas.

El modelo puede utilizarse a diferentes niveles de resolución y abstracción de acuerdo a cada aplicación. Así, en este capítulo, se definen dos niveles de abstracción del cuerpo humano, con base en el modelo general, para tratar la estructura de la persona desde dos puntos de vista diferentes: vista frontal y vista en perspectiva.

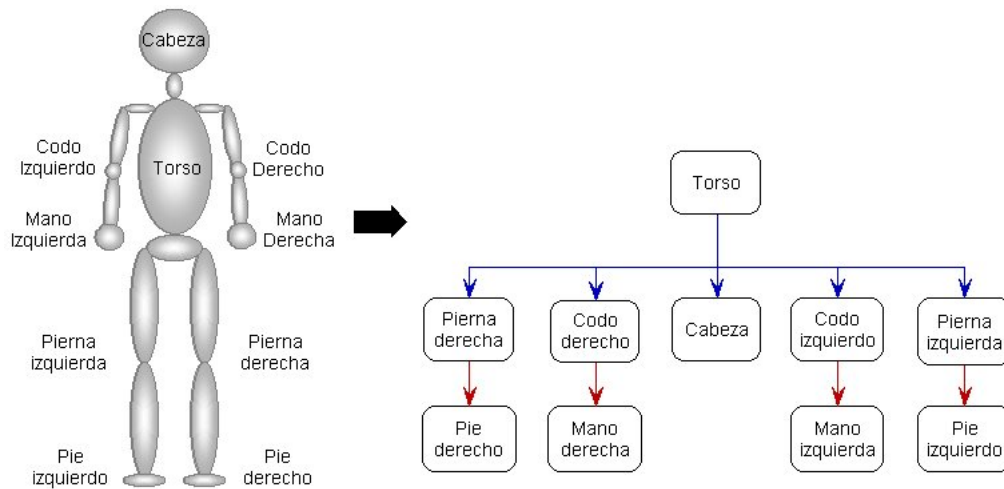


Figura 2.1: Modelo general del cuerpo humano

Para el seguimiento frontal se considera la parte superior del cuerpo: cabeza, torso, codos y manos (figura 2.2a). El objetivo es utilizar el modelo de estructura cinemática para el seguimiento visual de las diferentes entidades del cuerpo y la reproducción de los movimientos realizados por la persona mediante un modelo virtual en dos dimensiones.

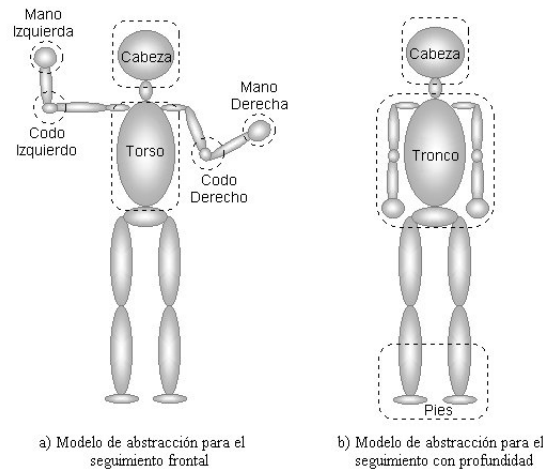


Figura 2.2: Modelos de abstracción del cuerpo humano para el seguimiento frontal y con profundidad de la persona

Para el seguimiento con profundidad, la cámara se coloca en perspectiva con respecto al espacio de trabajo; el propósito es captar el área en donde se desplaza la persona y determinar su posición real en el ambiente mediante un modelo matemático de proyección de la cámara. Para esta aplicación se considera un nivel de abstracción que comprende únicamente: la cabeza, el tronco y los pies; utilizándose el término

tronco para referirse al conjunto de torso y brazos y el término pies para denotar al grupo de piernas y pies, como se muestra en la figura 2.2b. Se utiliza este modelo debido a que al caminar, los brazos y las piernas presentan oclusiones que dificultan su seguimiento, por lo que el nivel de abstracción propuesto ayuda a obtener la posición de la persona en el ambiente de una manera más confiable.

Este capítulo describe los detalles del modelo cinemático y las restricciones de estructura para cada nivel de abstracción.

2.1 Modelo general de estructura

El objetivo principal del modelo de estructura es definir el estado de las articulaciones en un momento dado y facilitar la predicción de su posición en el siguiente instante de tiempo. Para simplificar la localización de la partes del cuerpo de la persona durante el seguimiento, se utiliza un árbol de dependencia cinemática. De manera que, una vez encontrado el nodo base, los elementos de los siguientes niveles del árbol quedan delimitados, en cada caso, por las características geométricas y cinemáticas definidas en el modelo de estructura.

Para determinar la posición de los diferentes elementos del cuerpo, la etapa de predicción relativa busca la entidad a seguir en la última posición en que fue encontrada relativa a la parte del cuerpo inmediata superior, considerando las restricciones cinemáticas y de estructura del cuerpo humano [Olvera 00]. Por ende, la predicción relativa disminuye el espacio de búsqueda al iniciar el seguimiento en la posición más probable de cada entidad. El algoritmo de seguimiento con predicción relativa realiza la exploración del árbol de estructura mediante el procedimiento *depth-first*, como se muestra a continuación.

Pseudocódigo 2.1:

Procedimiento: *jerárquica(m)*

Estimar la posición del elemento base mediante el seguidor

1. **inicio**
2. visitar el elemento y marcar (m)
3. obtener la posición relativa con base en la posición de la entidad inmediata superior en el árbol de estructura
4. aplicar el seguidor en la posición relativa y determinar la posición final del elemento
5. **mientras** exista un elemento sin marcar (sm) adyacente a (m) **hacer**
6. *jerárquica(sm)*
7. **fin** {mientras}
8. **fin**

A continuación se describe la predicción relativa para el caso de dos entidades del árbol de dependencia cinemática. Sean $\mathbf{p}_i(t)$ la posición de la entidad padre, y

$\mathbf{p}_{i+1}(t)$ la posición de la entidad inmediata inferior; donde $\mathbf{p}_i(t)$ es un vector en dos dimensiones (x_i, y_i) que denota la posición de la entidad i en el tiempo t . Supongamos que en el tiempo $t+1$, la posición del nodo padre $\mathbf{p}_i(t+1)$ se ha encontrado, a partir del algoritmo de seguimiento, y se requiere estimar la posición de la entidad $\mathbf{p}_{i+1}(t+1)$. La predicción relativa sugiere que $\mathbf{p}_{i+1}(t+1)$ se encontrará en la misma posición relativa a $\mathbf{p}_i(t)$, donde se encontraba en el tiempo t , tal como se muestra en la figura 2.3a; siendo $\mathbf{r}_i(t)$ el vector relativo de $\mathbf{p}_{i+1}(t)$ con respecto a $\mathbf{p}_i(t)$,

$$\mathbf{r}_i(t) = \mathbf{p}_{i+1}(t) - \mathbf{p}_i(t) \quad (2.1)$$

Dado que se conoce la nueva posición de $\mathbf{p}_i(t+1)$ en el tiempo $t+1$, se calcula el ángulo de rotación relativo a su última posición,

$$\mathbf{r}_i(t+1) = \text{rot}\theta_i \cdot \mathbf{r}_i(t) \quad (2.2)$$

donde θ_i es el ángulo generado por el movimiento del nodo $\mathbf{p}_i(t+1)$ con respecto a la posición anterior, por tanto,

$$\text{rot}\theta_i = \begin{bmatrix} \cos\theta_i & -\text{sen}\theta_i \\ \text{sen}\theta_i & \cos\theta_i \end{bmatrix} \quad (2.3)$$

De esta manera, la nueva posición de la entidad $\mathbf{p}_{i+1}(t+1)$ se obtiene incluyendo el desplazamiento rotacional,

$$\mathbf{p}_{i+1}(t+1) = \mathbf{p}_i(t+1) + \mathbf{r}_i(t+1) \quad (2.4)$$

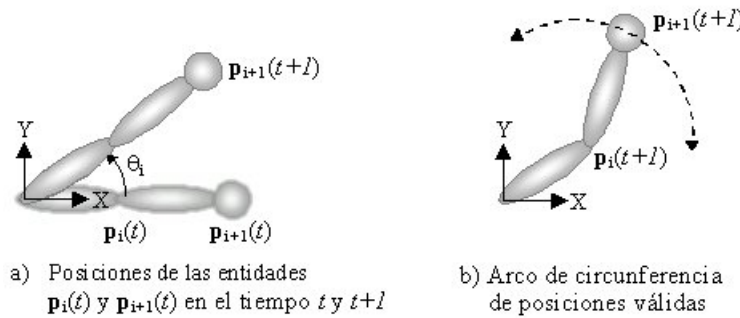


Figura 2.3: Posiciones de las entidades $\mathbf{p}_i(t)$ y $\mathbf{p}_{i+1}(t)$ en los tiempos t y $t+1$.

La predicción relativa estima la posición de la entidad inferior $\mathbf{p}_{i+1}(t+1)$, suponiendo que se encuentra estática con respecto al padre $\mathbf{p}_i(t+1)$, lo cual puede ser inexacto si $\mathbf{p}_{i+1}(t+1)$ también se está moviendo. Por lo tanto, se procede a generar un arco circular de posiciones válidas a partir de la posición estimada de $\mathbf{p}_{i+1}(t+1)$ [Olvera 00], como se muestra en la figura 2.3b. El radio válido se obtiene considerando la distancia entre ambas partes del cuerpo así como el ángulo máximo entre ellas.

Finalmente, se aplica el seguidor visual (descrito en el Capítulo 3), en la posición encontrada en la etapa de predicción relativa, para obtener la posición de $\mathbf{p}_{i+1}(t+1)$, asegurando que dicha posición se encuentra en el arco de circunferencia de posiciones válidas.

2.2 Seguimiento frontal

El seguimiento de una persona frente a la cámara tiene como propósito la recuperación del movimiento articulado de la parte superior del cuerpo humano, a través de una secuencia de imágenes y su reproducción mediante un modelo virtual de la persona. Una suposición importante en la implementación del algoritmo de seguimiento frontal es que los movimientos de la persona siempre deben ser paralelos al plano de visión de la cámara, lo cual limita los posibles movimientos a sólo dos dimensiones.

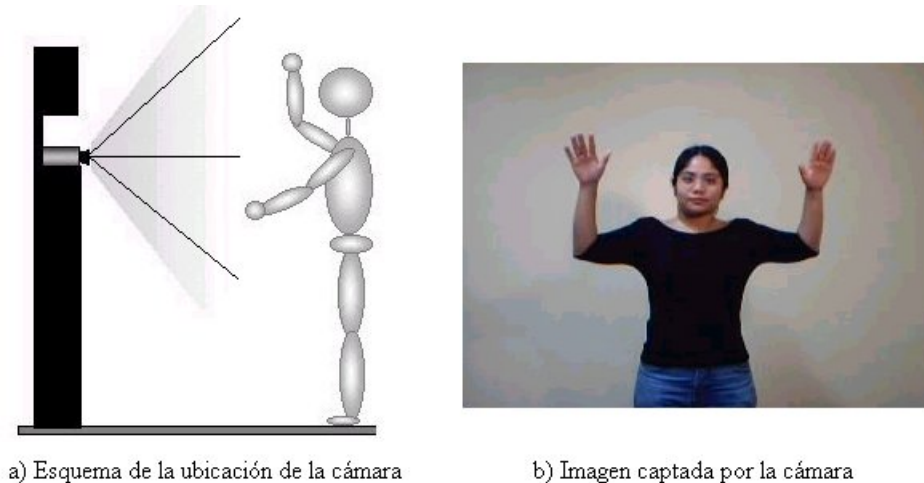


Figura 2.4: Posición de la cámara en el ambiente, para el seguimiento frontal

El escenario utilizado para seguir y reproducir los movimientos de una persona, realizados enfrente de una cámara, se presenta en la figura 2.4. La descripción del modelo del cuerpo humano para el seguimiento frontal consta de: cabeza, torso, codos y manos; como se observa en la figura 2.5. Donde $\mathbf{p}_i(t)$ representa las posiciones (x_i, y_i) de los elementos correspondientes al nivel i del árbol de estructura.

El seguimiento frontal se inicia con la localización del nodo base del modelo de estructura mediante el algoritmo de seguimiento. En este caso, se seleccionó el torso como la parte del cuerpo base, por lo que sus movimientos afectan las posiciones de las demás entidades de acuerdo al modelo establecido. Para calcular el ángulo de rotación del torso, a partir del cual se va a realizar la etapa de predicción, se consideró un punto de referencia estático en la cintura de la persona.

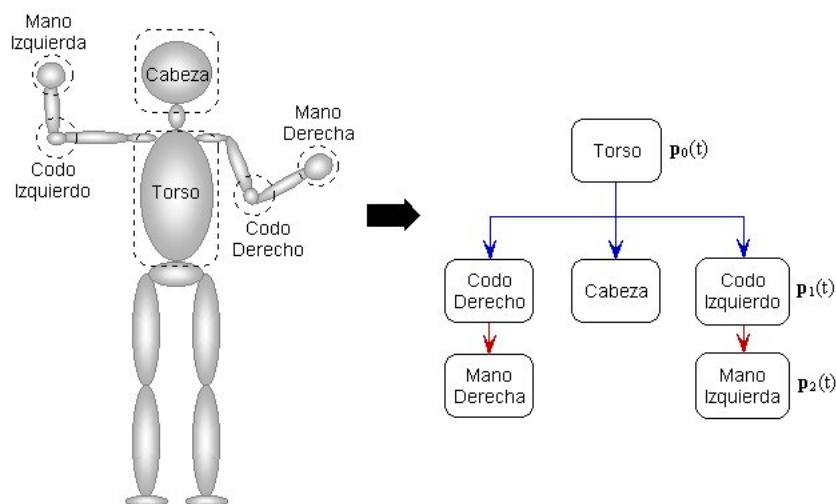


Figura 2.5: Modelo jerárquico para el seguimiento de los movimientos de una persona frente a la cámara.

Para fines de presentación, la figura 2.6 ilustra la etapa de predicción relativa para el caso del codo y la mano. Si la posición del codo $\mathbf{p}_1(t+1)$ ya ha sido encontrada a partir de la posición del torso $\mathbf{p}_0(t+1)$, se requiere determinar la posición del siguiente elemento inferior correspondiente a la mano $\mathbf{p}_2(t+1)$. De esta manera, se aplica la ecuación (2.4) para encontrar la posición probable del codo, a partir de la posición del torso dada por el seguidor. Una vez encontrada la posición del codo se utiliza esta información para predecir la posición de la mano.

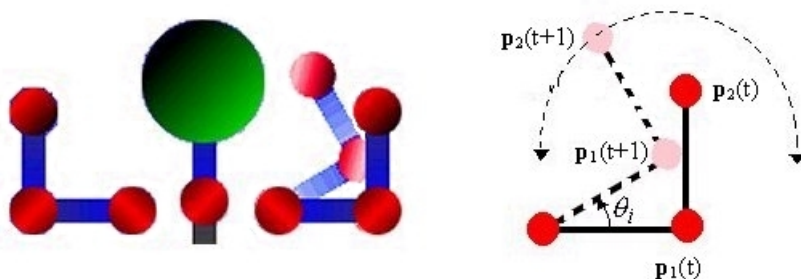


Figura 2.6: Posiciones del codo y de la mano en los tiempos t y $t+1$.

Para considerar el movimiento de la mano, después de aplicar el procedimiento de predicción relativa, se genera un arco circular de posiciones válidas a partir de la posición estimada, como se muestra en la figura 2.6. El radio válido se obtiene considerando la distancia entre ambas partes del cuerpo así como el ángulo máximo que puede tener la mano con respecto al codo. Finalmente, se aplica el algoritmo de seguimiento para obtener la posición de la mano, asegurando que dicha posición se encuentra en el radio válido obtenido mediante las restricciones cinemáticas.

2.2.1 Modelo virtual, modelo cinemático y animación

Los movimientos realizados por la persona se codifican, con el fin de reproducir la secuencia de animación en el modelo virtual de la persona. Así, mediante el uso de la cinemática del cuerpo humano, se establece la secuencia de instrucciones (ángulos) del mecanismo articulado que llevan a la posición realizada por la persona.

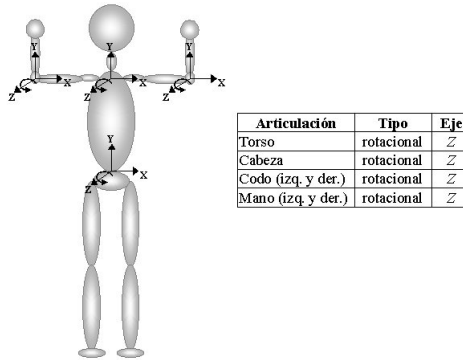


Figura 2.7: Elementos y ejes referenciales del modelo virtual

El modelo virtual es una representación computacional que refleja el estado de los movimientos efectuados por el individuo. El modelo cuenta con una descripción geométrica y cinemática equivalente a la persona. En el presente trabajo, se modela la geometría de las partes del cuerpo mediante líneas y articulaciones que representan sus diferentes componentes. Mientras que el modelo cinemático se define mediante la estructura jerárquica que representa la composición de las articulaciones y las relaciones entre ellas. Debido a que sólo se consideran movimientos en dos dimensiones, la rotación de cada una de las articulaciones se realiza con respecto al eje Z , como se describe en la figura 2.7. La figura 2.8 presenta un ejemplo de la reproducción del movimiento realizado por una persona. Así, la animación consiste en la aplicación de la transformación de coordenadas directa, con base en los ángulos que deben adoptar las articulaciones correspondientes al modelo virtual.

2.3 Seguimiento con profundidad

El seguimiento de una persona cuando camina se efectúa en el escenario que describe la figura 2.9, donde la cámara se encuentra colocada en perspectiva con respecto al ambiente. El seguidor capta el cuerpo completo de la persona al evolucionar dentro de una región suficientemente grande para que se desplace sin salir del área de visión de la cámara.

Al caminar, considerando la vista a distancia, los brazos y las piernas presentan oclusiones que dificultan su seguimiento. Por tanto, el modelo de abstracción utilizado,

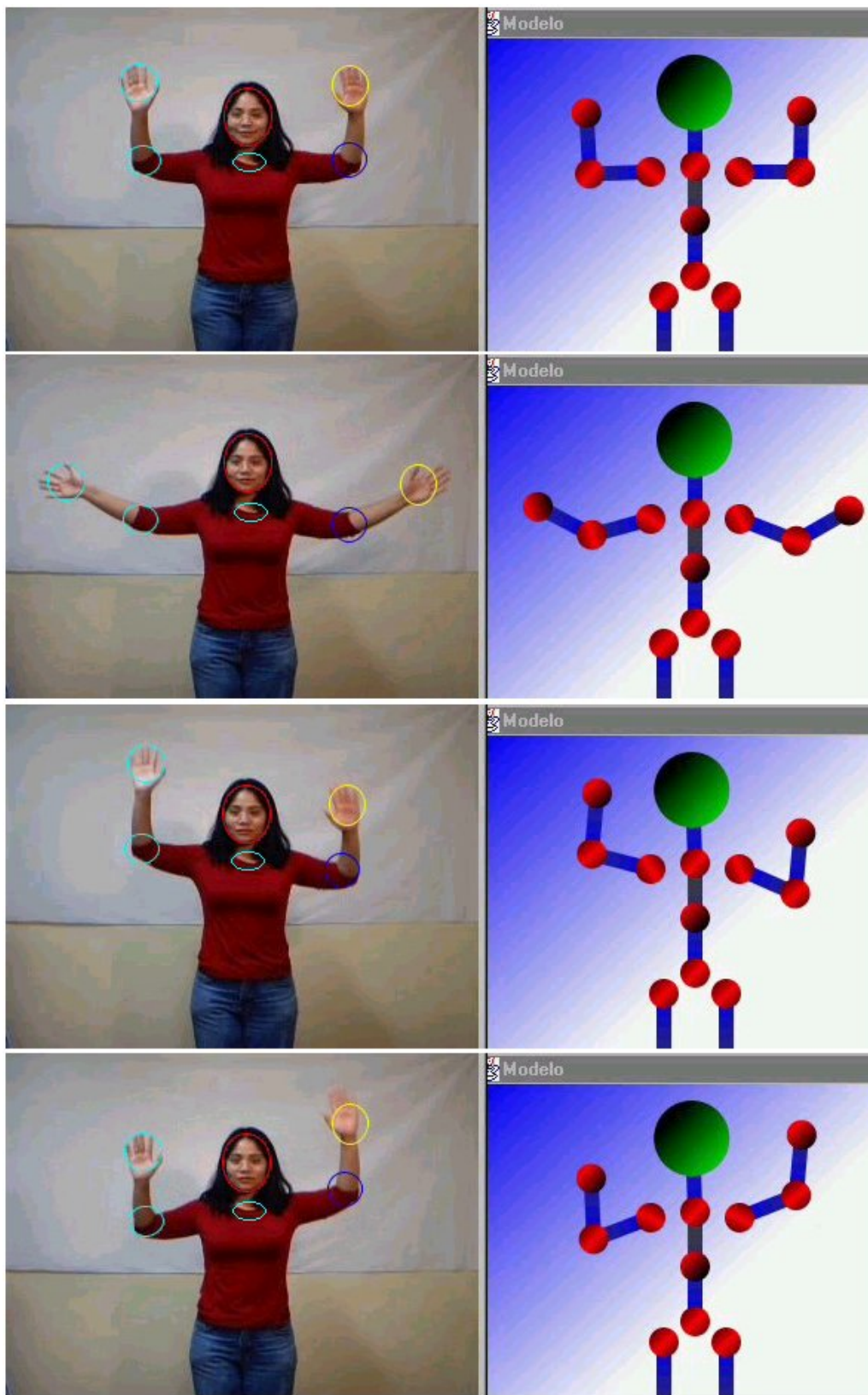


Figura 2.8: Secuencia de movimientos de una persona frente a la cámara

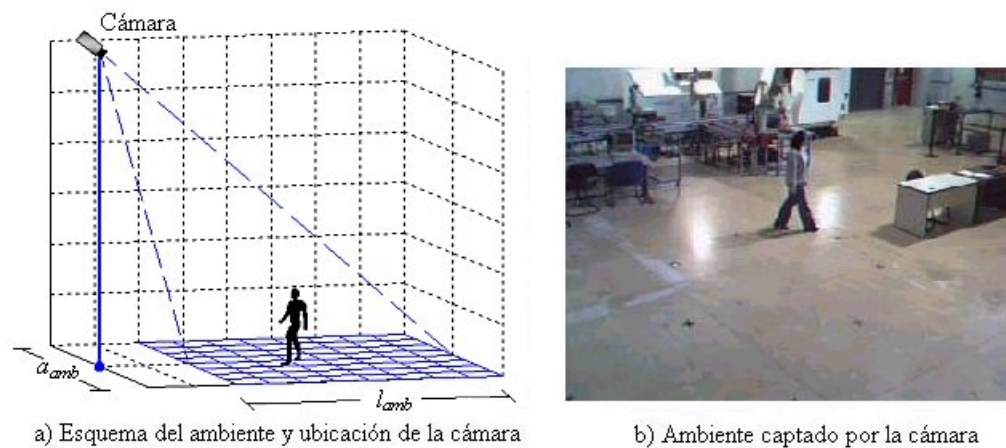


Figura 2.9: Posición de la cámara en el ambiente, para el seguimiento con profundidad

en este caso, comprende únicamente la cabeza, el tronco y los pies, usando el término tronco para el conjunto de torso y brazos, y el término pies para denotar al grupo de piernas y pies, como se muestra en la figura 2.10.

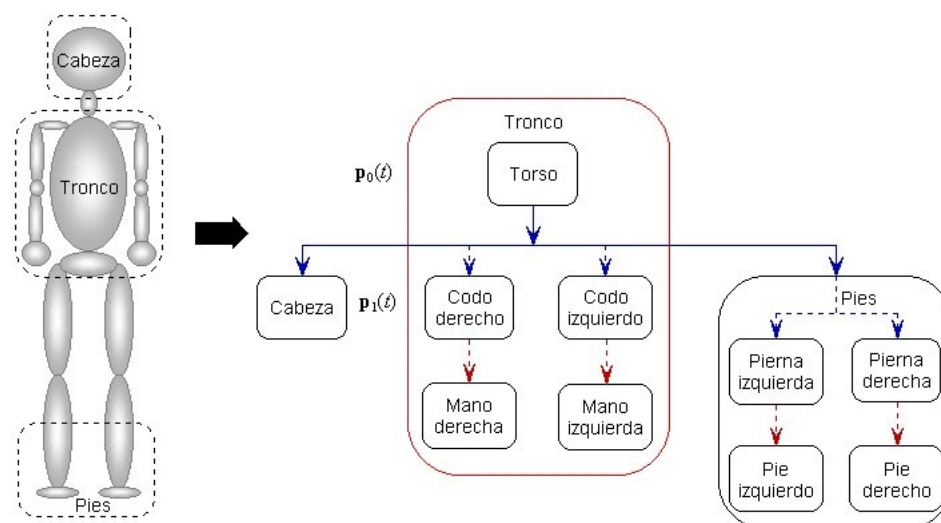


Figura 2.10: Modelo jerárquico para el seguimiento de una persona caminando

Para el seguimiento con profundidad se aplica el modelo de predicción relativa, considerando al tronco como el punto de inicio o base a partir del cual se predice la posición de la cabeza y de los pies. La posición del tronco se estima de manera independiente aplicando el algoritmo de seguimiento, de acuerdo a criterios de desplazamiento sobre un plano.

La predicción relativa de las entidades cabeza y pies, utiliza el mismo principio de seguimiento sobre la estructura del árbol cinemático. En este caso, ambos elementos (cabeza y pies) dependen del tronco. Por ejemplo, sean $\mathbf{p}_0(t)$ y $\mathbf{p}_1(t)$ la posición del tronco y la posición de la cabeza en el tiempo t , respectivamente. Si en el tiempo $t + 1$, se ha encontrado la posición del tronco $\mathbf{p}_0(t + 1)$ mediante el algoritmo de seguimiento, la posición de la cabeza $\mathbf{p}_1(t + 1)$ se estima utilizando la etapa de predicción relativa. La predicción relativa sugiere que $\mathbf{p}_1(t + 1)$ se encontrará en la misma posición relativa a $\mathbf{p}_0(t)$ en que se encontraba en el tiempo t , a partir de la nueva posición del tronco $\mathbf{p}_0(t + 1)$ en el tiempo $t + 1$, de manera general

$$\mathbf{p}_{i+1}(t + 1) = \mathbf{p}_i(t + 1) + \mathbf{p}_{i+1}(t) \quad (2.5)$$

El proceso de predicción relativa estima la posición de la cabeza $\mathbf{p}_1(t + 1)$, a partir del desplazamiento del tronco, sin considerar el movimiento propio de la cabeza. Como el movimiento es rotacional, la siguiente etapa genera un arco circular de posiciones válidas para la cabeza a partir de la posición del tronco. El radio válido se obtiene considerando la distancia y el ángulo máximo entre ambas partes del cuerpo. Finalmente, se aplica el algoritmo de seguimiento para obtener la posición más probable de la cabeza, asegurando que dicha posición se encuentra en el arco de circunferencia válido. En la figura 2.11 se muestra un ejemplo de la predicción relativa y el arco de circunferencia de posiciones válidas para la cabeza.

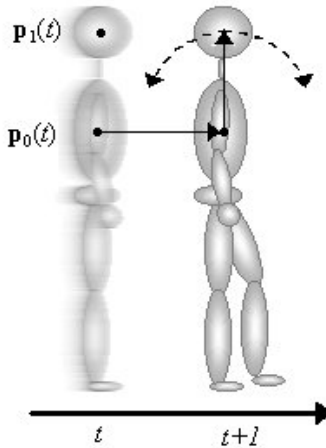


Figura 2.11: Posiciones del tronco y la cabeza en los tiempos t y $t + 1$

El objetivo del seguimiento con profundidad es obtener la posición de la persona en el ambiente a partir de las imágenes capturadas por una cámara que se encuentra ubicada en perspectiva con relación al ambiente. Para determinar la posición real de la persona, se aplica el modelo matemático de perspectiva propuesto por [Vázquez 02, Aguilar 03]. Dicho modelo se basa en medidas del ambiente, como la altura a la que se encuentra la cámara y su ángulo de inclinación, para obtener la proyección de un punto

en la imagen en su posición real sobre el ambiente. Así, la transformación tiene como entrada las coordenadas (x_{img}, y_{img}) que corresponden a la ubicación de la persona en la imagen y como salida las coordenadas de la posición de la persona en el ambiente real (x_{amb}, y_{amb}) . De manera que, el modelo de perspectiva determina la posición de la persona en el espacio de trabajo a partir de su posición correspondiente en la imagen y tomando en cuenta la altura de la persona. Los detalles del modelo matemático de perspectiva se presentan en el apéndice A.

Capítulo 3

Algoritmo de desplazamiento medio

El seguimiento de las partes del cuerpo de la persona se realizó mediante el método integrativo basado en el algoritmo de desplazamiento medio [Comaniciu 00]. Dicho algoritmo es una técnica de estimación no paramétrica que maximiza la correlación que existe entre dos distribuciones estadísticas o funciones de densidad: la correspondiente al objeto de interés y la del objeto candidato. El término *objeto de interés* se refiere a la parte del cuerpo que se va a seguir y el término *objeto candidato* denota a la región de la imagen que tiene la distribución más similar a la del objeto de interés.

Las distribuciones estadísticas se obtienen usando alguna característica que permita delimitar al objeto que se desea seguir. Entre las características que pueden emplearse se encuentran: el color, la textura o el gradiente de la imagen; en este caso se utilizaron características de color.

En general, el algoritmo de seguimiento mediante desplazamiento medio consiste en obtener la distribución del objeto de interés (distribución modelo) y del objeto candidato, comparar enseguida ambas distribuciones mediante un coeficiente de similitud y estimar la posición probable del objeto de interés a partir del *vector de desplazamiento medio*. Este vector se calcula iterativamente hasta encontrar la distribución más similar a la del objeto modelo. El diagrama general del algoritmo se muestra en la figura 3.1.

En este capítulo se describe detalladamente el procedimiento para obtener las distribuciones estadísticas de las partes del cuerpo a seguir, el algoritmo de desplazamiento medio, la actualización del modelo o distribución del objeto de interés, y el método de adaptación de escala propuesto.

3.1 Notación

A lo largo del documento se utilizará la siguiente notación:

Vector: Un vector en dos dimensiones es un par ordenado de números reales,

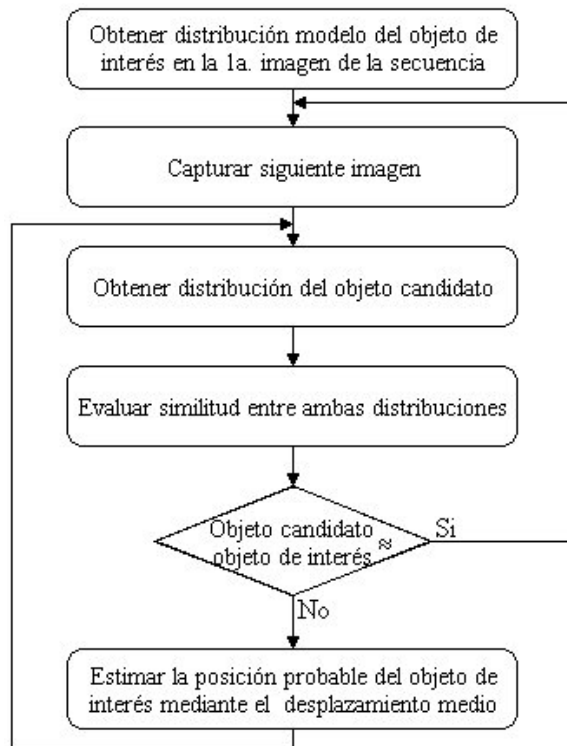


Figura 3.1: Diagrama general del algoritmo de desplazamiento medio

$$\mathbf{x} = (x_1, x_2)$$

donde x_1, x_2 son los componentes del vector \mathbf{x} . Por tanto, los vectores se denotarán con letras minúsculas en negritas (\mathbf{x}, \mathbf{y} , etc.), y los escalares se representarán con letras minúsculas en cursivas (x, y , etc.).

Norma: La norma o magnitud de un vector se representa por

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2}$$

Funciones de estimación de densidad: El símbolo “ $\hat{\cdot}$ ” denotará a las funciones de estimación de densidad (por ejemplo $\hat{f}(\mathbf{y})$ es la estimación de densidad de la función $f(\mathbf{y})$).

Para simplificar la descripción del algoritmo de seguimiento se utilizará la notación de vector para representar a un punto en la imagen, de manera que $f(\mathbf{y}) = f(x, y)$. Por otra parte, se usará \mathbf{y} para identificar al pixel correspondiente al centro de la ventana de seguimiento y \mathbf{x}_i para denotar a los pixeles que conforman dicha ventana.

3.2 Adquisición del modelo del objeto de interés

La primera etapa del algoritmo de seguimiento es obtener el modelo del objeto de interés a partir de las características discriminantes que van a identificarlo a lo largo de la secuencia de imágenes. En este caso, se van a utilizar características de color definidas mediante una distribución estadística.

La estimación de densidad genera una distribución estadística a partir de un conjunto de datos pertenecientes a una función de probabilidad desconocida. Para definir una función de estimación existen dos aproximaciones: estimación paramétrica y estimación no paramétrica [Silverman 86]. La estimación de densidad paramétrica supone que los datos pertenecen a una familia conocida de distribuciones (distribución Gaussiana, exponencial, etc.) cuyas características dependen del valor de un conjunto de parámetros.

Por otra parte, la aproximación no paramétrica obtiene la función de densidad de los datos sin determinar ningún modelo a priori para la distribución de probabilidad. Los estimadores no paramétricos son populares en el campo de la Visión Computacional, ya que la mayoría de los datos analizados son complejos y no se ajustan a una distribución conocida. Por esta razón, se propuso utilizar una técnica de estimación no paramétrica para obtener el modelo del objeto de interés.

Un estimador de densidad no paramétrico simple y utilizado comúnmente es el histograma. Para construir un histograma se divide el espacio de los datos en pequeños intervalos iguales y se determina el número de muestras que caen en un intervalo en particular. En imágenes, cada uno de estos intervalos corresponde a un nivel de intensidad diferente. Recordando, una imagen se define como $I(x, y)$, donde x, y son las coordenadas espaciales, e I es una función bidimensional que asigna un nivel de gris a cada par de coordenadas (x, y) . El histograma estadístico de una imagen en escala de grises se representa como:

$$H(k) = \frac{a_k}{N} \quad (3.1)$$

donde k es el k -ésimo nivel de gris, $k = 0, 1, \dots, L - 1$, (L corresponde al número de intensidades de gris en la imagen), a_k es el número de píxeles con el nivel de gris k y N es el número total de píxeles en la imagen. En la figura 3.2 se presenta un ejemplo del histograma de una imagen en escala de grises.

Para el caso de una imagen a color, la función $I(x, y)$ asigna un color a cada par de coordenadas (x, y) . En el modelo *RGB* cada color consta de tres elementos que corresponden a los niveles de intensidad rojo (*Red*), verde (*Green*) y azul (*Blue*), respectivamente. Así, el histograma de color es:

$$H(u) = \frac{a_u}{N} \quad (3.2)$$

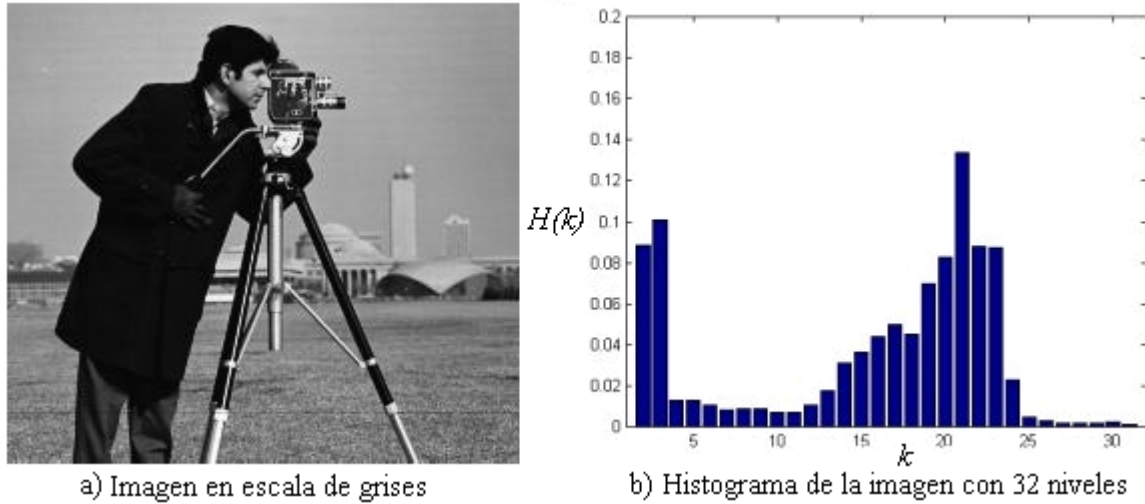


Figura 3.2: Histograma de una imagen en escala de grises

donde u es el color correspondiente a los niveles de intensidad r -ésimo para el rojo, g -ésimo para el verde y b -ésimo para el azul, el rango de cada uno de los niveles de intensidad r, g y b es $0, 1, \dots, L - 1$ para $L \times L \times L$ niveles de color, y a_u es el número de pixeles en la imagen con el color u .

3.2.1 Estimación de *densidad de núcleo*

La estimación de *densidad de núcleo* es muy similar a los histogramas excepto que remueve las discontinuidades de la función de densidad estimada, interpolando los valores del histograma en intervalos infinitesimales. Los métodos de estimación de densidad de núcleo se basan en la premisa de que funciones continuas y diferenciables se pueden modelar exactamente por la suma infinita de otra función llamada núcleo, elegida apropiadamente. Usar la estimación de densidad para modelar el color tiene muchas ventajas. A diferencia de los histogramas, la estimación de densidad de núcleo permite obtener la distribución de una manera suave, continua y diferenciable, con un número pequeño de muestras.

La estimación de densidad de núcleo $\hat{f}(\mathbf{y})$ define una distribución de probabilidad arbitraria $f(\mathbf{y})$, en el espacio bidimensional,

$$\hat{f}(\mathbf{y}) = \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{\mathbf{y} - \mathbf{x}_i}{h}\right) \quad (3.3)$$

donde $\mathbf{x}_i = (x_i, y_i)$ son las posiciones de los n pixeles que conforman a la ventana de interés cuyo centro $\mathbf{y} = (x, y)$, $h = \left\| \left(\frac{h_x}{2}, \frac{h_y}{2} \right) \right\|$ es el radio de la región determinada por una ventana de dimensiones (h_x, h_y) ; y $K(\mathbf{y})$ es el núcleo definido por una función definida positiva, simétrica e integrable a uno.

Entre las funciones núcleo comúnmente utilizadas en la estimación de densidad se encuentran: el núcleo Gaussiano, el núcleo de Epanechnikov, el núcleo rectangular y el núcleo uniforme, entre otros [Cheng 95]. Sin embargo, se ha obtenido que la función óptima que minimiza el error global entre el estimado y la densidad real es el núcleo multivariable de Epanechnikov [Scott 92], definido para dos dimensiones como

$$K_E(\mathbf{y}) = \begin{cases} \frac{2}{\pi} (1 - \|\mathbf{y}\|^2) & \text{si } \|\mathbf{y}\| < 1 \\ 0 & \text{en otro caso} \end{cases} \quad (3.4)$$

donde $\mathbf{y} = (x, y)$ es un pixel en la imagen. La figura 3.3 muestra el núcleo de Epanechnikov para el caso bidimensional. Como se observa en la gráfica, este núcleo es una función esférica, decreciente, monótonica y convexa.

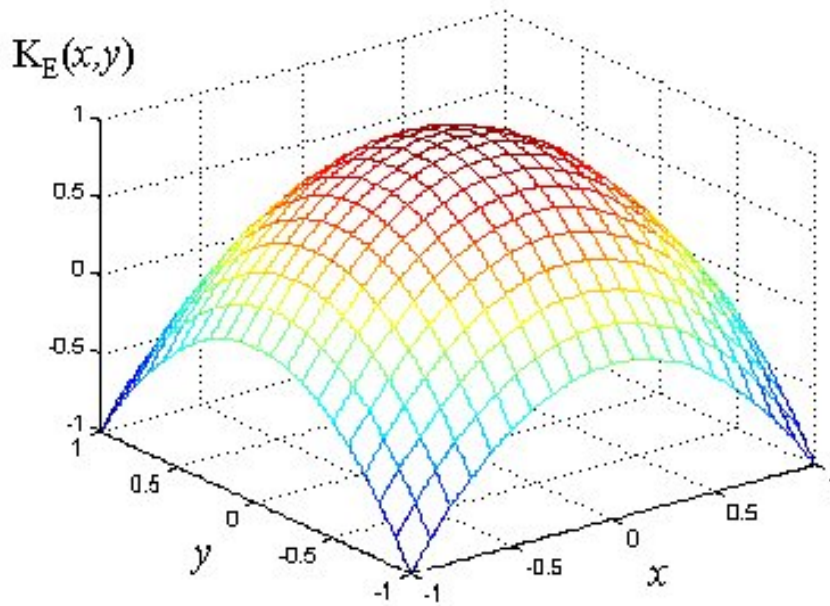


Figura 3.3: Núcleo de Epanechnikov

El núcleo de Epanechnikov incrementa la robustez de la estimación de densidad en imágenes, debido a que asigna un mayor peso a los píxeles que se encuentran en el centro de la ventana de interés y un peso mínimo a los píxeles más alejados. Tal asignación ayuda a disminuir los errores al ser los píxeles de la periferia los menos confiables, pues a menudo son afectados por oclusiones.

3.2.2 Distribución modelo del objeto de interés

Para obtener la distribución modelo del objeto de interés $\hat{\mathbf{q}}_u$ se aplica la estimación de densidad de núcleo con la función de Epanechnikov [Comaniciu 00],

$$\hat{\mathbf{q}}_u = C \sum_{i=1}^n K_E \left(\frac{\mathbf{x}_i}{h} \right) \delta [b(\mathbf{x}_i) - u], \quad (3.5)$$

donde el subíndice u indica que la distribución $\hat{\mathbf{q}}_u$ se obtiene con base en características de color $u = 1 \dots m$, m es el número total de colores en la imagen, n es el número total de píxeles de la ventana de interés, $\mathbf{x}_i = (x_i, y_i)$ son las posiciones de los píxeles que conforman a la ventana tomando como origen el centro de la región de interés, h es el radio de la ventana de seguimiento con dimensiones (h_x, h_y) y $b(\mathbf{x}_i)$ es una función que asigna a cada píxel \mathbf{x}_i el índice de la distribución correspondiente al color en dicho píxel. Finalmente, δ es la función delta de Kronecker definida por

$$\delta = \begin{cases} 1 & \text{si } b(\mathbf{x}_i) = u \\ 0 & \text{si } b(\mathbf{x}_i) \neq u \end{cases} \quad (3.6)$$

La constante de normalización C se deriva imponiendo la condición $\sum_{u=1}^m \hat{\mathbf{q}}_u = 1$,

$$C = \frac{1}{\sum_{i=1}^n K_E \left(\frac{\mathbf{x}_i}{h} \right)}, \quad (3.7)$$

puesto que la suma de las funciones delta para $u = 1 \dots m$ es igual a uno. La distribución $\hat{\mathbf{q}}_u$ sólo se obtiene en la primera imagen de la secuencia para determinar el modelo con el cual se comparan las distribuciones candidatas $\hat{\mathbf{p}}_u(\mathbf{y})$.

Para obtener la distribución del objeto candidato $\hat{\mathbf{p}}_u(\mathbf{y})$ en la imagen actual se utiliza la ecuación,

$$\hat{\mathbf{p}}_u(\mathbf{y}) = C_h \sum_{i=1}^n K_E \left(\frac{\mathbf{y} - \mathbf{x}_i}{h} \right) \delta [b(\mathbf{x}_i) - u] \quad (3.8)$$

donde $\mathbf{y} = (x, y)$ es una variable que corresponde al centro de la ventana candidata y \mathbf{x}_i son las posiciones de los píxeles del objeto candidato. Aplicando la condición $\sum_{u=1}^m \hat{\mathbf{p}}_u(\mathbf{y}) = 1$, la constante de normalización C_h se define como:

$$C_h = \frac{1}{\sum_{i=1}^n K_E \left(\frac{\mathbf{y} - \mathbf{x}_i}{h} \right)} \quad (3.9)$$

La figura 3.4 muestra un ejemplo de la distribución del objeto modelo obtenida al aplicar la ecuación (3.5) en una imagen en escala de grises. La distribución de una imagen en escala de grises facilita la visualización, pues es bidimensional; una dimensión proporciona el nivel de gris y la otra el número de píxeles correspondientes a cada nivel. Mientras que en una imagen a color, la distribución se define en cuatro

dimensiones lo cual es difícil de visualizar. La figura 3.4-a presenta el objeto de interés (cabeza de la persona) seleccionado en la imagen mediante una ventana de dimensiones $h_x = 27$ y $h_y = 25$, mientras la figura 3.4-b muestra la función de densidad estimada considerando 16 niveles de intensidad.

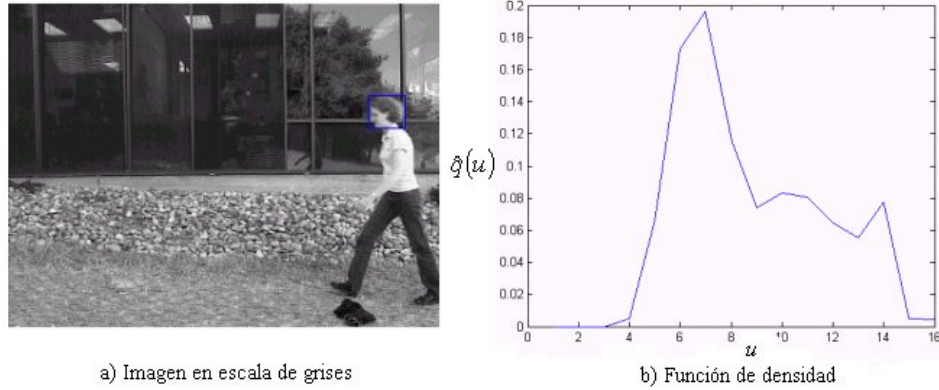


Figura 3.4: Estimación de densidad de una región de la imagen

3.3 Coeficiente de similitud

Existen varias técnicas para obtener la similitud entre dos distribuciones estadísticas, entre las más sencillas se encuentran la intersección y la norma L_1 que consiste en ir sumando las diferencias entre los correspondientes niveles de cada distribución. También se aplican la distancia de Jeffrey, la distancia de Minkowski, el coeficiente de Bhattacharyya, y la estadística- χ^2 , entre otras. Sin embargo, se han realizado análisis comparativos [Roberts 02] que determinan que el coeficiente que obtiene mejores resultados en aplicaciones de seguimiento es el coeficiente de Bhattacharyya [Bhattacharyya 43],

$$\rho [\hat{\mathbf{p}}_u(\mathbf{y}), \hat{\mathbf{q}}_u] = \sum_{u=1}^m \sqrt{\hat{p}_u(\mathbf{y}) \hat{q}_u} \quad (3.10)$$

donde $\hat{\mathbf{p}}_u(\mathbf{y})$ es la distribución estadística del objeto candidato y $\hat{\mathbf{q}}_u$ es la distribución estadística del objeto de interés. Mientras mayor sea el valor del coeficiente de Bhattacharyya ρ más similares son las distribuciones. Para dos distribuciones idénticas se obtiene $\rho = 1$ lo que indica una correspondencia biunívoca, entre cada elemento de ambas distribuciones (objeto candidato y de interés). En este caso, el coeficiente de Bhattacharyya tiene el significado de una función de correlación cuyo máximo local en la imagen indica la presencia de un objeto con una distribución similar a la distribución original del objeto de interés.

En la figura 3.5 se muestra una gráfica de los valores del coeficiente de Bhattacharyya obtenidos entre la distribución de un objeto de interés (cabeza de la persona

en la figura 3.4) y la distribución del objeto candidato en la siguiente imagen de la secuencia. El valor inicial del coeficiente se representa con un círculo y el valor máximo, que corresponde a la posición de la distribución más similar a la del objeto de interés, se denota mediante un asterisco.

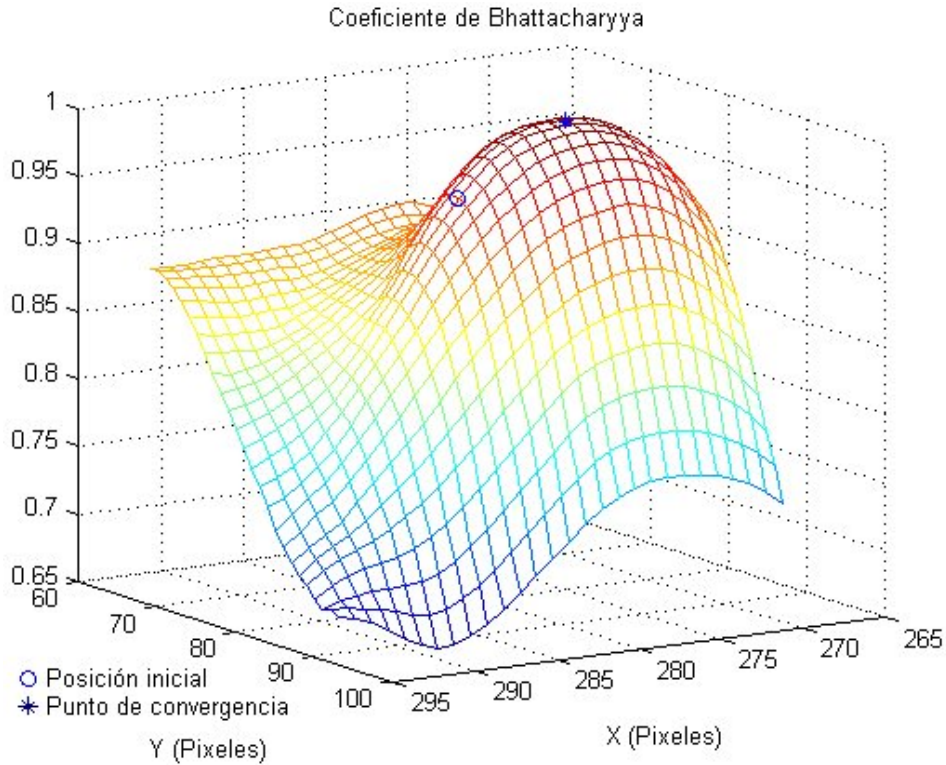


Figura 3.5: Valores del coeficiente de Bhattacharyya correspondientes a la ventana de seguimiento seleccionada en la figura 3.4. El valor inicial del coeficiente se representa con un círculo y el valor máximo se denota mediante un asterisco.

Para obtener la localización más probable del objeto de interés en la secuencia de imágenes se busca minimizar la distancia que existe entre las dos distribuciones en cada par consecutivo de imágenes

$$d(\mathbf{y}) = \sqrt{1 - \rho[\hat{\mathbf{p}}_u(\mathbf{y}), \hat{\mathbf{q}}_u]} \quad (3.11)$$

lo cual es equivalente a maximizar el coeficiente de Bhattacharyya. El coeficiente de Bhattacharyya hereda las propiedades de la función núcleo cuando las distribuciones del objeto modelo y del objeto candidato se calculan mediante un estimador de densidad de núcleo. Una función núcleo diferenciable hace que el coeficiente de similitud sea diferenciable y se puedan utilizar procedimientos de optimización basados en gradiente para encontrar el máximo.

La búsqueda de la nueva localización del objeto en la imagen actual se inicia en $\mathbf{y}_0 = (x_0, y_0)$ que corresponde a la posición del objeto de interés en la imagen anterior

de la secuencia. Por tanto, la búsqueda se realiza en la vecindad de \mathbf{y}_0 , por lo que es conveniente encontrar una aproximación de la función de similitud, en las cercanías de este punto. Tal aproximación se realiza mediante la expansión de Taylor alrededor de los valores de la función de densidad del objeto candidato $\hat{\mathbf{p}}_u(\mathbf{y}_0)$, obteniéndose la siguiente expresión para el coeficiente después de algunas manipulaciones (ver Apéndice B)

$$\rho[\hat{\mathbf{p}}_u(\mathbf{y}), \hat{\mathbf{q}}_u] \approx \frac{1}{2} \sum_{u=1}^m \sqrt{\hat{p}_u(\mathbf{y}_0) \hat{q}_u} + \frac{1}{2} \sum_{u=1}^m \hat{p}_u(\mathbf{y}) \sqrt{\frac{\hat{q}_u}{\hat{p}_u(\mathbf{y}_0)}} \quad (3.12)$$

La expresión (3.12) es válida si se supone que la distribución del objeto candidato $\hat{\mathbf{p}}_u(\mathbf{y})$ no cambia drásticamente con respecto a la distribución anterior $\hat{\mathbf{p}}_u(\mathbf{y}_0)$ y que $\hat{p}_u(\mathbf{y}_0) > 0$ para toda $u = 1, \dots, m$. Sustituyendo la expresión para la distribución $\hat{\mathbf{p}}_u(\mathbf{y})$ en la ecuación (3.12),

$$\rho[\hat{\mathbf{p}}_u(\mathbf{y}), \hat{\mathbf{q}}_u] \approx \frac{1}{2} \sum_{u=1}^m \sqrt{\hat{p}_u(\mathbf{y}_0) \hat{q}_u} + \frac{C_h}{2} \sum_{i=1}^{n_h} \omega_i K_E \left(\frac{\mathbf{y} - \mathbf{x}_i}{h} \right), \quad (3.13)$$

donde,

$$\omega_i = \sum_{u=1}^m \delta[b(\mathbf{x}_i) - u] \sqrt{\frac{\hat{q}_u}{\hat{p}_u(\mathbf{y}_0)}} \quad (3.14)$$

La aproximación realizada es de gran utilidad ya que se puede observar que para minimizar la distancia entre las dos distribuciones, se debe maximizar el segundo término en la ecuación (3.13), puesto que el primer término es independiente de \mathbf{y} . Por otra parte, el segundo término representa la estimación de densidad calculada con el núcleo $K_E(\mathbf{y})$ en la imagen actual de la secuencia y con los datos ponderados mediante ω_i . La maximización del coeficiente de Bhattacharyya se puede obtener eficientemente mediante las iteraciones del vector de desplazamiento medio, como se verá en la siguiente sección.

3.4 Vector de desplazamiento medio

El vector de desplazamiento medio fue propuesto en 1975 por Fukunaga y Hostetler [Fukunaga 75]; y se ha utilizado como una herramienta para el análisis de características espaciales en muchas áreas de Visión Computacional [Comaniciu 99].

El uso de un núcleo diferenciable como el núcleo de Epanechnikov permite definir la estimación del gradiente de densidad como el gradiente de la estimación de densidad de núcleo [Fukunaga 75];

$$\hat{\nabla} f(\mathbf{y}) \equiv \nabla \hat{f}(\mathbf{y}) = \frac{1}{nh^2} \sum_{i=1}^n \nabla K_E \left(\frac{\mathbf{y} - \mathbf{x}_i}{h} \right). \quad (3.15)$$

Considerando que la derivada del núcleo de Epanechnikov es el núcleo uniforme $g(\mathbf{y})$ definido como

$$g(\mathbf{y}) = \begin{cases} 1 & \text{si } \|\mathbf{y}\| < 1 \\ 0 & \text{en otro caso} \end{cases} \quad (3.16)$$

y después de ciertas manipulaciones algebraicas ([Comaniciu 99]) se obtiene

$$\frac{\sum_{i=1}^n \mathbf{x}_i g\left(\frac{\mathbf{y}-\mathbf{x}_i}{h}\right)}{\sum_{i=1}^n g\left(\frac{\mathbf{y}-\mathbf{x}_i}{h}\right)} - \mathbf{y} = \frac{h^2}{4} \frac{\widehat{\nabla} f(\mathbf{y})}{\widehat{f}(\mathbf{y})} \quad (3.17)$$

donde el término de la izquierda representa al vector de desplazamiento medio. La expresión (3.17) muestra que el vector de desplazamiento medio obtenido con el núcleo g es una estimación del gradiente de densidad normalizado obtenido con el núcleo K_E . Aprovechando, en este caso, que el coeficiente de Bhattacharyya es una función suavizada, Comaniciu propuso maximizar la similitud aplicando el vector de desplazamiento medio para encontrar la estimación del gradiente. De esta manera, se puede determinar la zona con mayor densidad de puntos pertenecientes a los pixeles que conforman al objeto de interés.

Por tanto, la posición que maximiza el coeficiente de similitud se obtiene mediante

$$\mathbf{y}_1 = \frac{\sum_{i=1}^n \mathbf{x}_i \omega_i g\left(\frac{\mathbf{y}_0-\mathbf{x}_i}{h}\right)}{\sum_{i=1}^n \omega_i g\left(\frac{\mathbf{y}_0-\mathbf{x}_i}{h}\right)} \quad (3.18)$$

donde ω_i son los pesos provenientes de la ecuación (3.14), $g(\mathbf{y})$ es un núcleo uniforme y \mathbf{y}_0 es el centro de la ventana en la imagen anterior de la secuencia.

La idea básica del algoritmo de desplazamiento es definir una región de búsqueda, la cual tiene como origen el centro de la ventana de seguimiento. A partir de la posición inicial \mathbf{y}_0 , el algoritmo calcula la media espacial de los datos en la región de búsqueda (en este caso, los datos corresponden al valor de similitud de cada uno de los pixeles de la región de interés), y desplaza la ventana para centrarse en la media estimada hasta que converge. En la figura 3.6 se presenta un ejemplo del desplazamiento medio aplicado a un conjunto de puntos. Se puede observar que en cada iteración la ventana inicial se desplaza hacia la región con mayor densidad de puntos hasta que converge.

El cálculo sucesivo del vector de desplazamiento medio garantiza que siempre se converge a un punto cercano donde la estimación tiene gradiente cero. La garantía de convergencia se debe a la magnitud adaptable del vector de desplazamiento medio, la cual elimina la necesidad de utilizar procedimientos adicionales para elegir el tamaño adecuado de los pasos. Esta capacidad de adaptación es la principal ventaja sobre los métodos tradicionales basados en gradiente; en los cuales si el tamaño del paso es muy largo, el algoritmo tiende a divergir, mientras que si el tamaño del paso es muy pequeño, el tiempo de convergencia tiende a ser muy grande. Por lo tanto, el método de desplazamiento medio es un método adaptable de ascenso de gradiente que

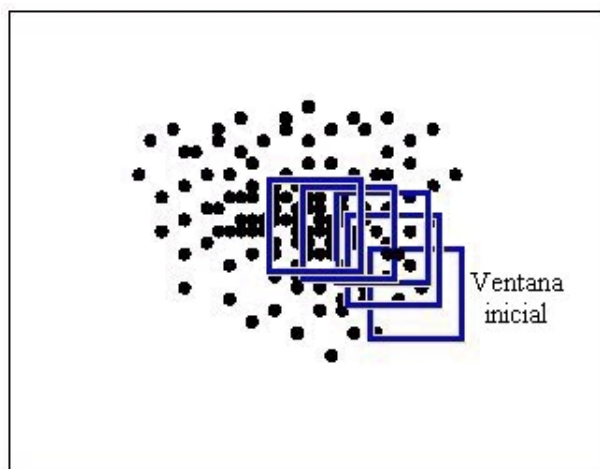


Figura 3.6: Ejemplo de la aplicación del algoritmo de desplazamiento medio en un conjunto de puntos

avanza con pasos largos en las regiones con coeficientes de similitud pequeños, ya que tales regiones no corresponden al modelo del objeto de interés. Análogamente, cerca del máximo local del coeficiente de similitud los pasos son pequeños, de forma que el análisis de la posición probable sea más refinado.

La figura 3.7 muestra el resultado de los puntos encontrados mediante el vector de desplazamiento medio aplicado entre dos imágenes de una secuencia. En la figura se observa que la distancia entre los puntos, obtenidos mediante el vector de desplazamiento medio, disminuye a medida que se acercan a la posición del objeto de interés (región con mayor similitud al objeto modelo).

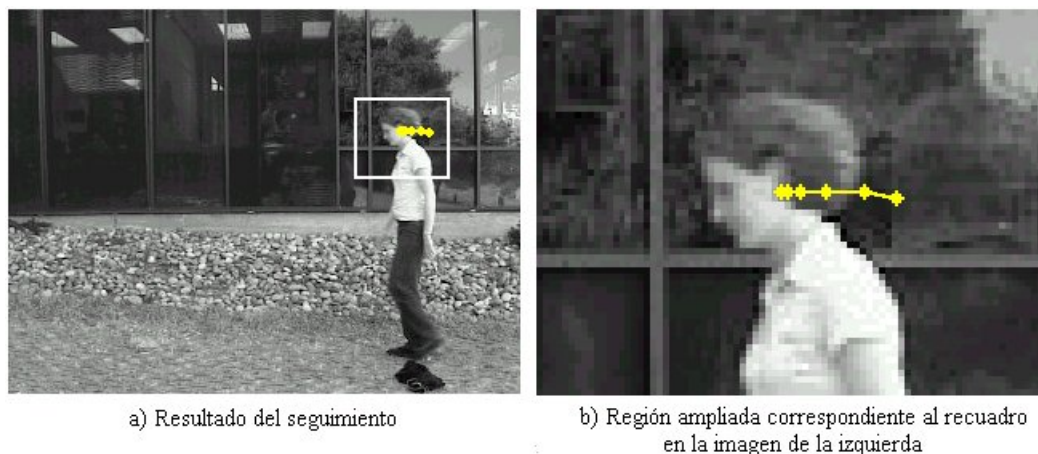


Figura 3.7: Trayectoria definida por los puntos obtenidos mediante el vector de desplazamiento medio aplicado en la imagen 10 y 15 de la secuencia, después de 6 iteraciones.

3.5 Algoritmo de seguimiento

El algoritmo de seguimiento implica la predicción de la posición del objeto candidato más similar al objeto modelo en la imagen actual de la secuencia. Por tanto, el problema consiste en encontrar la función de densidad $\hat{\mathbf{p}}_u(\mathbf{y})$ centrada en la posición \mathbf{y} cuya distribución sea la más cercana a la distribución modelo $\hat{\mathbf{q}}_u$. La similitud entre ambas distribuciones se obtiene mediante el coeficiente de Bhattacharyya, y la posición más probable del objeto de interés se determina a partir del vector de desplazamiento medio. De manera general, el algoritmo de seguimiento consta de las etapas que aparecen en el diagrama 3.8:

1. Seleccionar el objeto de interés en la imagen inicial de la secuencia, mediante una ventana de dimensiones h_x y h_y .
2. Determinar la distribución modelo del objeto de interés $\hat{\mathbf{q}}_u$ definido por la ecuación (3.5).
3. Obtener la siguiente imagen de la secuencia.
4. Determinar la función de densidad $\hat{\mathbf{p}}_u(\mathbf{y}_0)$ correspondiente al objeto candidato, a partir de la región definida por el rectángulo centrado en la posición \mathbf{y}_0 (posición del objeto de interés en la imagen anterior de la secuencia). La función de densidad $\hat{\mathbf{p}}_u(\mathbf{y}_0)$ se obtiene mediante la ecuación (3.8).
5. Evaluar la similitud entre la distribución del objeto de interés $\hat{\mathbf{q}}_u$ y la distribución del objeto candidato $\hat{\mathbf{p}}_u(\mathbf{y}_0)$ mediante el coeficiente de Bhattacharyya definido por la ecuación (3.10)
6. Verificar si $\hat{\mathbf{q}}_u \approx \hat{\mathbf{p}}_u(\mathbf{y}_0)$, i.e. si $|1 - \rho[\hat{\mathbf{p}}_u(\mathbf{y}_0), \hat{\mathbf{q}}_u]| < \epsilon_0$ continuar con la siguiente imagen de la secuencia de lo contrario,
7. Calcular los pesos ω_i definidos mediante la ecuación (3.14) y
8. Aplicar el vector de desplazamiento medio (3.18) para encontrar la dirección hacia la posición \mathbf{y}_1 del objeto candidato más similar al objeto de interés.
9. Si la posición \mathbf{y}_1 regresada por el desplazamiento medio no varía con respecto a la posición anterior \mathbf{y}_0 (i.e. $\|\mathbf{y}_1 - \mathbf{y}_0\| < \epsilon_1$) detener las iteraciones y continuar con la siguiente imagen de la secuencia; de lo contrario reemplazar $\mathbf{y}_0 \leftarrow \mathbf{y}_1$ para seguir iterando hasta que se alcance el número máximo de iteraciones (i_{max}) sobre la imagen actual.

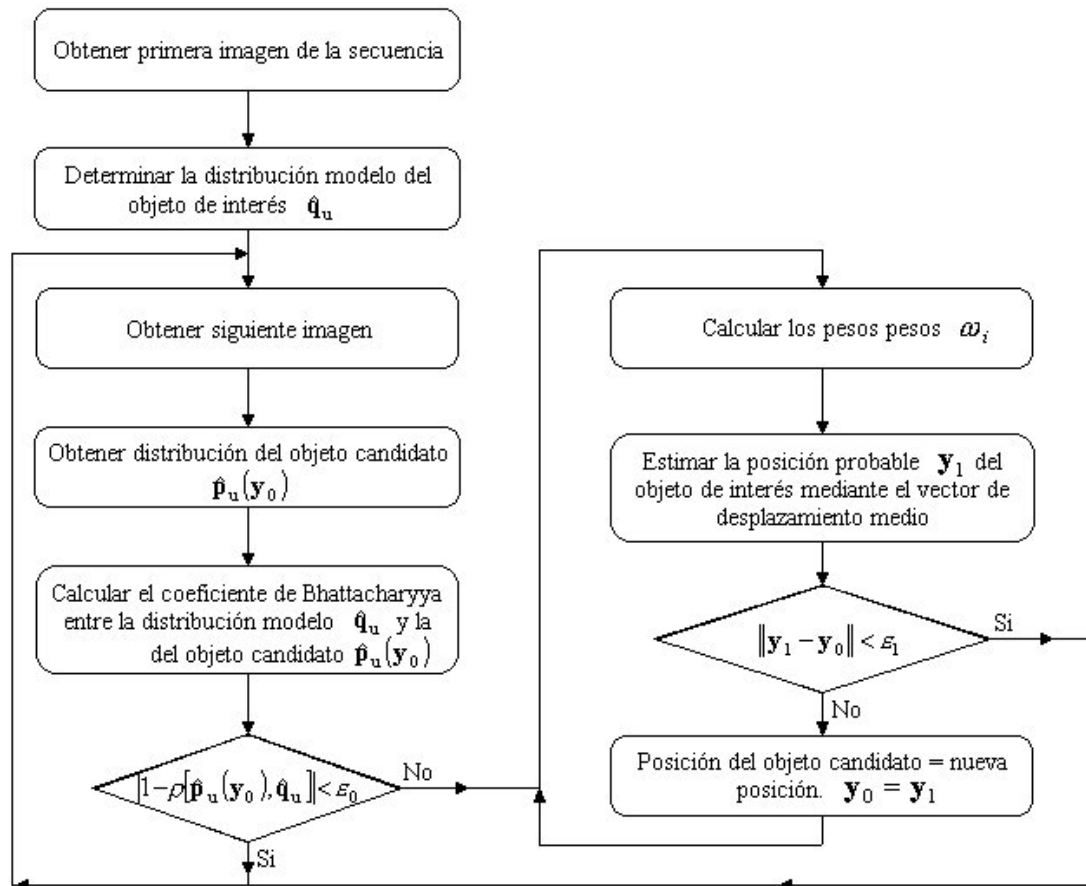


Figura 3.8: Diagrama del algoritmo de desplazamiento medio

3.6 Actualización del modelo del objeto de interés

La distribución original del objeto de interés varía a lo largo del tiempo, debido entre otras causas a movimientos de rotación del objeto y cambios en las condiciones de iluminación. Para actualizar la distribución modelo, y mejorar el seguimiento a lo largo de todo el proceso, es necesario modificar el algoritmo de desplazamiento medio con el fin de ajustar la distribución de probabilidad del objeto.

Para determinar cuando se debe ajustar el modelo de distribución se utiliza como criterio el coeficiente de Bhattacharyya. Como se describió en la Sección 2.2.1, el coeficiente de Bhattacharyya determina la similitud que existe entre la distribución original y la distribución del objeto candidato. Un valor del coeficiente igual a uno indica que existe una correspondencia biunívoca entre las dos distribuciones; por lo tanto, cuando el coeficiente entre ambas distribuciones sea menor que un umbral establecido debe modificarse el modelo original, de manera que, la distribución se adapte a los cambios del objeto de interés.

El nuevo modelo se obtiene mediante un promedio ponderado entre la distribución original y la distribución en la imagen actual utilizando la siguiente ecuación:

$$\hat{\mathbf{q}}_u = \beta \hat{\mathbf{q}}_u + (1 - \beta) \hat{\mathbf{p}}_u(\mathbf{y}) \quad (3.19)$$

donde $\mathbf{y} = (x, y)$ es el centro de la ventana de seguimiento actual y β es una constante que ajusta el porcentaje de contribución de las dos distribuciones y se obtiene empíricamente. Para evitar cambios drásticos en el modelo original del objeto de interés se le otorga menor peso a la distribución actual, típicamente $\beta = 0.9$.

3.7 Adaptación de escala

El algoritmo original de desplazamiento medio supone que el tamaño de las ventanas permanece fijo, de forma que no considera cambios de escala del objeto de interés, dificultando así el seguimiento cuando la cámara se coloca en perspectiva con respecto al ambiente de trabajo (seguimiento con profundidad). La ventaja principal de la ubicación en perspectiva del sistema de visión es captar un espacio de trabajo mayor al de una ubicación perpendicular; sin embargo, esta ubicación de la cámara implica cambios de escala del objeto a seguir.

Para resolver el problema de adaptación de escala que presenta el algoritmo de desplazamiento medio, se explotó la propiedad de invarianza en escala de las distribuciones y el conocimiento que a priori se tiene del ambiente. Debido a la colocación en perspectiva de la cámara, los objetos se ven grandes en la parte inferior de la imagen, mientras que en la parte superior su tamaño disminuye. Por tanto, se propone un método que utiliza el modelo matemático de perspectiva, según se describe en el apéndice A, para determinar el factor de escala de acuerdo a la posición de la persona en el ambiente.

El factor de escala propuesto consiste en la relación de la proyección del tamaño de la persona en el punto actual del ambiente, y la altura de la persona en el primer punto visto por la cámara. Se considera que en el primer punto visto por la cámara, el factor de escala tiene un valor unitario puesto que en este punto el tamaño de la persona es máximo. El método de adaptación de escala propuesto supone que la altura de la persona no cambia a lo largo del tiempo, considerando actividades como caminar y correr.

Así, el factor de escala se calcula mediante la relación entre las proyecciones, sobre el plano de la imagen, de la altura de la persona en la posición actual en el ambiente (a_c) y de la altura en el primer punto visto por la cámara (a_m),

$$\Delta e = \frac{a_c}{a_m} \quad (3.20)$$

Esta proporción equivale a la relación entre el tamaño de la persona en cualquier punto del ambiente y el tamaño máximo que se tiene en la imagen. De esta manera,

se determina la razón en la que varía la escala de las ventanas de seguimiento en el algoritmo de desplazamiento medio, aprovechando el conocimiento que se tiene del ambiente y del modelo de perspectiva. En la figura 3.9 se observa un ejemplo de la relación de escala que existe en dos posiciones diferentes del espacio de trabajo.

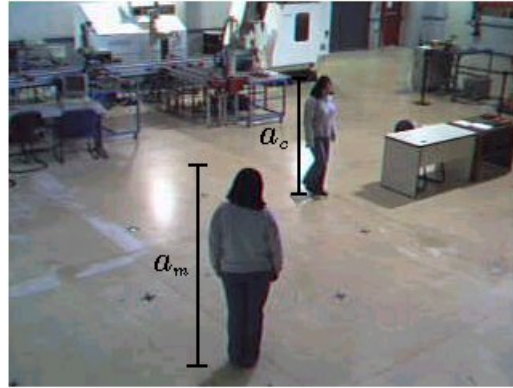


Figura 3.9: Relación de escala entre dos posiciones diferentes de la persona en el ambiente

A continuación se obtienen las proyecciones a partir del modelo de perspectiva cuyo esquema se muestra en la figura 3.10. La notación que se utilizará en esta sección fue definida para el modelo de perspectiva en el apéndice A. En esta notación, un segmento de línea del punto a al punto b se representa como \overline{ab} y su magnitud se define como ab , un triángulo cuyos vértices son los puntos a , b y c , se denota como $\triangle abc$ y para denotar un ángulo se utilizan letras del alfabeto griego (α , β , etc.).

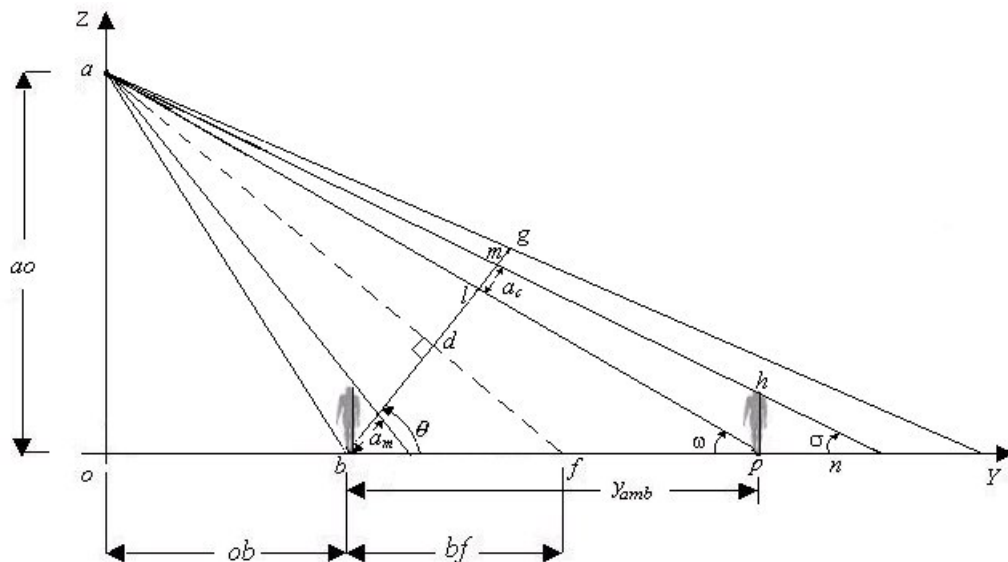


Figura 3.10: Esquema del ambiente utilizado para la escala

Para generar la proporción de escala se utiliza la proyección del tamaño de la persona sobre el segmento \overline{bg} que corresponde a la resolución total de la cámara en el eje Y . Por tanto, a partir de la figura 3.10 se observa la proyección de la persona a_c en el segmento \overline{bg} ,

$$a_c = lm = bm - bl \quad (3.21)$$

La magnitud del segmento \overline{bl} se obtiene utilizando ley de senos en el triángulo Δblp ,

$$bl = y_{amb} \frac{\text{sen}(\omega)}{\text{sen}(\pi - \theta - \omega)} \quad (3.22)$$

donde θ se define del modelo de perspectiva,

$$\theta = \tan^{-1} \left(\frac{ob + bf}{ao} \right) \quad (3.23)$$

mientras que ω se obtiene del triángulo Δoap ,

$$\omega = \tan^{-1} \left(\frac{ao}{ob + y_{amb}} \right) \quad (3.24)$$

Por su parte, la magnitud bm se calcula utilizando ley de senos en el triángulo Δbmn ,

$$bm = (y_{amb} + pn) \frac{\text{sen}(\sigma)}{\text{sen}(\pi - \theta - \sigma)} \quad (3.25)$$

donde pn corresponde a la proyección de la altura de la persona en el plano del ambiente. Debido a que los triángulos Δoan y Δphn son semejantes por tener dos ángulos iguales (el ángulo recto y σ),

$$\frac{pn}{hp} = \frac{ob + y_{amb} + pn}{ao} \quad (3.26)$$

y despejando

$$pn = \frac{hp(y_{amb} + ob)}{ao - hp} \quad (3.27)$$

Finalmente, el ángulo σ se obtiene del triángulo Δoan ,

$$\sigma = \tan^{-1} \left(\frac{ao}{ob + y_{amb} + pn} \right) \quad (3.28)$$

Por otro lado, la proyección de la persona en el primer punto de vista a_m se obtiene de las ecuaciones 3.21 y 3.25, considerando que la posición de la persona coincide con el primer punto visto por la cámara ($y_{amb} = 0$).

Capítulo 4

Implementación y experimentos

Este capítulo describe la implementación del método de seguimiento visual propuesto, así como los resultados de los experimentos realizados para validar la propuesta de tesis. Como el sistema está implementado mediante un programa computacional, los recursos utilizados se limitan a una cámara y una computadora con una tarjeta de captura y procesamiento de video.

A continuación se detalla la implementación del método de seguimiento, se describen los experimentos realizados para evaluar el seguimiento frontal y con profundidad de una persona, y finalmente se analizan los resultados obtenidos en cada caso.

4.1 Implementación

Para validar el método de seguimiento propuesto y considerar los dos niveles de abstracción del cuerpo humano, se realizaron montajes diferentes. Para el seguimiento con profundidad, la cámara se colocó en perspectiva con respecto al espacio de trabajo (figura 2.9), con el fin de captar un área donde la persona pueda desplazarse, y calcular su posición real en el ambiente a partir de un modelo matemático de proyección de la cámara.

Por otra parte, para el seguimiento frontal, la cámara se colocó enfrente de la persona, como se muestra en la figura 2.4, con el propósito de seguir las partes del cuerpo de interés y reproducir los movimientos realizados mediante un modelo virtual en dos dimensiones.

El sistema de seguimiento se programó en JAVA a excepción de los métodos nativos, utilizados durante el proceso de adquisición de imágenes, que se programaron en C++. La computadora que alberga el sistema es una PC Hewlett-Packard con un procesador Pentium 4 a 1.7GHz, 512 Mb de memoria RAM y sistema operativo Windows 98. Para capturar las imágenes se utilizó una cámara CCD de Sun Microsystems y una tarjeta de procesamiento de video Meteor II de Matrox.

El sistema computacional implementado opera sobre imágenes capturadas en tiempo real y en secuencias de imágenes previamente grabadas en archivos con formato JPEG. En promedio se procesan 15 imágenes/segundo para el seguimiento de una persona caminando (seguimiento con profundidad) y 10 imágenes/segundo para el seguimiento y reproducción de los movimientos realizados por una persona frente a la cámara (seguimiento frontal).

La interfaz computacional desarrollada se muestra en las figuras 4.1 y 4.2. En tal interfaz se implementó el algoritmo de seguimiento visual basado en la estructura del cuerpo humano.



a) Seguimiento de una persona caminando

b) Ventanas de configuración de parámetros del modelo de perspectiva

Figura 4.1: Interfaz computacional para el seguimiento con perspectiva

La figura 4.1 muestra un ejemplo del seguimiento con profundidad, así como las ventanas de configuración de los parámetros correspondientes al modelo de perspectiva. Por otra parte, la figura 4.2 presenta el modelo virtual generado para la reproducción de los movimientos de la persona en el caso del seguimiento frontal. El apéndice C describe con detalle el funcionamiento de la interfaz gráfica del sistema.

4.2 Seguimiento con profundidad

Las pruebas realizadas para el seguimiento de una persona caminando se llevaron a cabo en el Laboratorio de Manufactura del ITESM, Campus Monterrey. Se colocó una cámara en perspectiva con respecto al espacio de trabajo del laboratorio y se tomaron las medidas requeridas por el modelo de proyección. La figura 4.3 muestra las dimensiones del ambiente donde se realizaron las pruebas.

Los valores establecidos para los parámetros del sistema se muestran en la tabla 4.1. Dichos valores se utilizaron en los experimentos para validar tanto el seguimiento con profundidad como el seguimiento frontal.

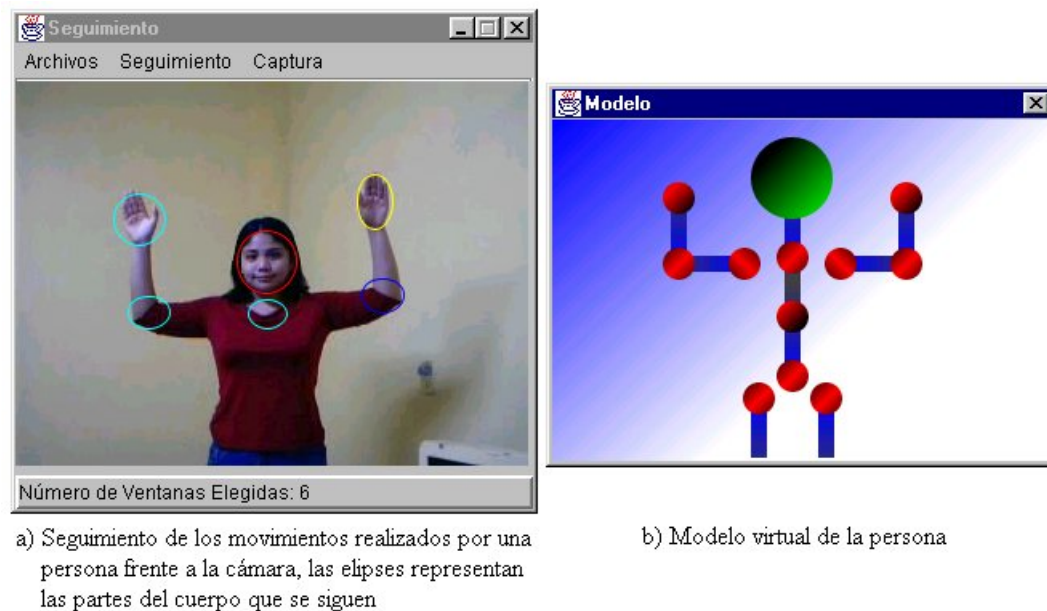


Figura 4.2: Interfaz computacional para la representación virtual de los movimientos de una persona

Tabla 4.1: Valores de los parámetros del sistema utilizados en los experimentos.

Niveles de Color	$\mu = 16 \times 16 \times 16$
Máximo número de iteraciones del vector de desplazamiento medio	$i_{max} = 15$
Umbral de similitud	$\epsilon_0 = 0.05$
Umbral de posición	$\epsilon_1 = 1$
Umbral de ajuste del modelo	$\beta = 0.9$

Para calcular el error de la posición de la persona en el plano XY , se obtuvo la diferencia entre la posición real de la persona y la posición calculada mediante el algoritmo de seguimiento al aplicar el modelo de proyección de la cámara. Debido a que no se cuenta con un dispositivo para obtener la posición real de la persona, para el cálculo del error se colocaron marcas sobre el piso y se midió su posición en el ambiente, con el fin de utilizar tales posiciones como puntos de referencia.

Basándose en el experimento diseñado por [Aguilar 03], la trayectoria definida para el seguimiento se muestra en la figura 4.4, donde aparecen los puntos de control como marcas colocadas en el piso. La ruta consiste en que la persona recorra las marcas en el sentido de las manecillas del reloj. Se seleccionaron este tipo de trayectorias para tener un recorrido sencillo y fácil de realizar por la persona. Además, la ruta es suficientemente completa en desplazamientos para que las validaciones aporten datos relevantes.

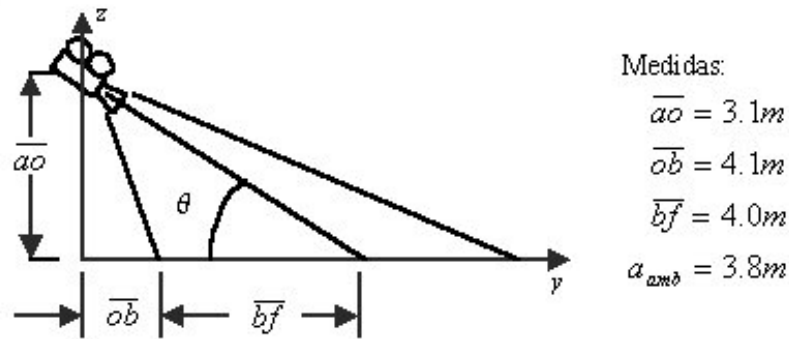


Figura 4.3: Medidas del ambiente en el que se realizaron los experimentos

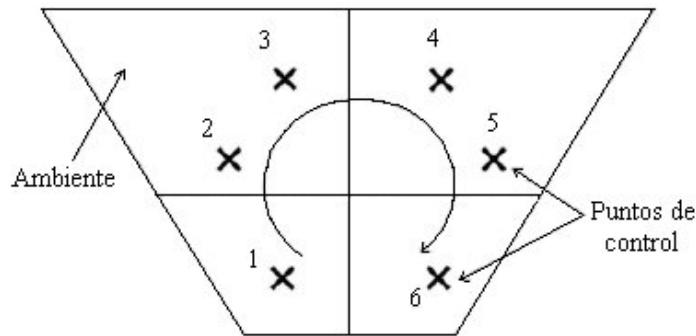


Figura 4.4: Trayectoria establecida para el recorrido de las marcas

4.2.1 Medición del error de posición

El error de posición se define como la distancia que existe entre la posición real de la persona y la posición calculada por el modelo de perspectiva. La posición real corresponde al punto sobre el cual se proyecta perpendicularmente el centro de masa de la persona. Para calcular el error, se utilizó la fórmula de la distancia euclidiana entre dos puntos

$$d_e = \sqrt{(x_{amb} - x_{real})^2 + (y_{amb} - y_{real})^2} \quad (4.1)$$

donde (x_{real}, y_{real}) representa la posición en el ambiente y (x_{amb}, y_{amb}) denota la posición calculada mediante el modelo de perspectiva.

El error obtenido es aproximado, debido a que no se cuenta con una manera de calcular la posición exacta de la persona en todo momento.

4.2.2 Análisis de resultados

A lo largo del seguimiento se estima la posición en la imagen de la cabeza, el tronco y los pies de la persona. Posteriormente, aplicando el modelo de perspectiva, se proyectan dichas posiciones sobre el plano del ambiente y se obtiene la localización real de la persona.

En las tablas 4.2, 4.3 y 4.4 se presentan los resultados que se obtuvieron de acuerdo al plano de la cabeza, el tronco y los pies respectivamente. Se analizaron tres recorridos de la ruta establecida, cada recorrido lo realizó una persona distinta, cuyas alturas son $1.55m$, $1.65m$ y $1.70m$ respectivamente. Cada persona corresponde a una columna de prueba de las tablas. La columna *Puntos de Control*, especifica el número de la marca colocada en el suelo. La columna *Posición Original* indica la posición real de la marca con respecto al referencial establecido para el ambiente. Finalmente, las columnas *Prueba1*, *Prueba2* y *Prueba3* presentan la posición calculada de cada persona mediante el modelo de proyección.

Tabla 4.2: Posiciones obtenidas de acuerdo al seguidor de la cabeza

<i>Puntos de Control</i>	<i>Posición Original (m)</i>	<i>Prueba 1 (1.55m)</i>	<i>Prueba 2 (1.65 m)</i>	<i>Prueba 3 (1.70 m)</i>
1	(-1.19,1.34)	(-1.22,1.25)	(-1.37,1.48)	(-1.26,1.36)
2	(-1.7,2.95)	(-1.78,3.13)	(-1.75,2.87)	(-1.83,3.35)
3	(-0.80,4.05)	(-0.84,3.78)	(-0.59,3.78)	(-0.59,4.12)
4	(0.70,3.95)	(0.79,3.88)	(0.72,3.67)	(0.90,4.20)
5	(1.25,2.79)	(1.27,2.88)	(1.22,2.61)	(1.20,2.61)
6	(0.75,1.40)	(0.84,1.35)	(0.64,1.10)	(0.74,1.51)

Tabla 4.3: Posiciones obtenidas de acuerdo al seguidor del tronco

<i>Puntos de Control</i>	<i>Posición Original (m)</i>	<i>Prueba 1 (1.55m)</i>	<i>Prueba 2 (1.65m)</i>	<i>Prueba 3 (1.70 m)</i>
1	(-1.19,1.34)	(-1.24,1.14)	(-1.45,1.5)	(-1.24,1.15)
2	(-1.70,2.95)	(-1.74,2.89)	(-1.92,3.10)	(-1.84,3.06)
3	(-0.80,4.05)	(-0.84,3.80)	(-0.46,3.95)	(-0.59,3.78)
4	(0.70,3.95)	(0.68,3.74)	(0.81,3.67)	(0.79,3.75)
5	(1.25,2.79)	(1.21,2.89)	(1.17,2.56)	(1.31,2.65)
6	(0.75,1.40)	(0.75,1.31)	(0.68,1.24)	(0.60,1.15)

Tabla 4.4: Posiciones obtenidas de acuerdo al seguidor de los pies

<i>Puntos de Control</i>	<i>Posición Original (m)</i>	<i>Prueba 1 (1.55m)</i>	<i>Prueba 2 (1.65m)</i>	<i>Prueba 3 (1.70m)</i>
1	(-1.19,1.34)	(-1.29,1.27)	(-1.33,1.34)	(-1.20,1.10)
2	(-1.7,2.95)	(-1.79,3.01)	(-1.71,2.69)	(-1.78,2.75)
3	(-0.80,4.05)	(-0.96,3.95)	(-0.54,3.83)	(-0.65,3.67)
4	(0.70,3.95)	(0.70,4.01)	(0.90,4.37)	(0.75,3.51)
5	(1.25,2.79)	(1.27,3.15)	(1.10,2.95)	(1.18,2.48)
6	(0.75,1.40)	(0.76,1.31)	(0.59,1.45)	(0.61,1.04)

Tabla 4.5: Errores correspondientes a las posiciones obtenidas mediante el seguidor de la cabeza

<i>Puntos de Control</i>	<i>Error Prueba 1</i>	<i>Error Prueba 2</i>	<i>Error Prueba 3</i>	<i>Error Promedio</i>
1	0.09m	0.23m	0.07m	0.13m
2	0.20m	0.09m	0.42m	0.24m
3	0.27m	0.34m	0.22m	0.28m
4	0.11m	0.28m	0.32m	0.24m
5	0.09m	0.18m	0.19m	0.15m
6	0.10m	0.32m	0.11m	0.18m

Para representar de forma gráfica la precisión del modelo de proyección a lo largo de la ruta definida por las marcas, se trazaron sobre la imagen las posiciones obtenidas por el sistema. En la figura 4.5 se observa un ejemplo (*Prueba1*) del seguimiento de una persona, donde se muestra la ruta definida por las marcas colocadas en el suelo y las trayectorias obtenidas con base en los resultados de las ventanas de la cabeza, del tronco y de los pies.

Los errores de medición se calcularon usando la distancia euclidiana, definida en la ecuación (4.1), entre las posiciones de los puntos de control y las posiciones registradas en las tablas de resultados. Las tablas 4.5, 4.6 y 4.7 presentan dichos errores, además de un promedio del error correspondiente a cada marca. El error promedio para cada una de las partes del cuerpo se muestra en la tabla 4.8.

Con base en los resultados obtenidos, se observa que el cálculo de la posición es más preciso cuando la persona se encuentra cerca de la cámara. Esta característica se debe a la resolución de la imagen utilizada, ya que al observarse un área de tamaño considerable, los píxeles que corresponden a posiciones alejadas de la cámara cubren un área mayor en el ambiente que los que están más cerca, lo cual incrementa el error de dichos píxeles. Por otra parte, la cabeza genera la posición más confiable, ya que

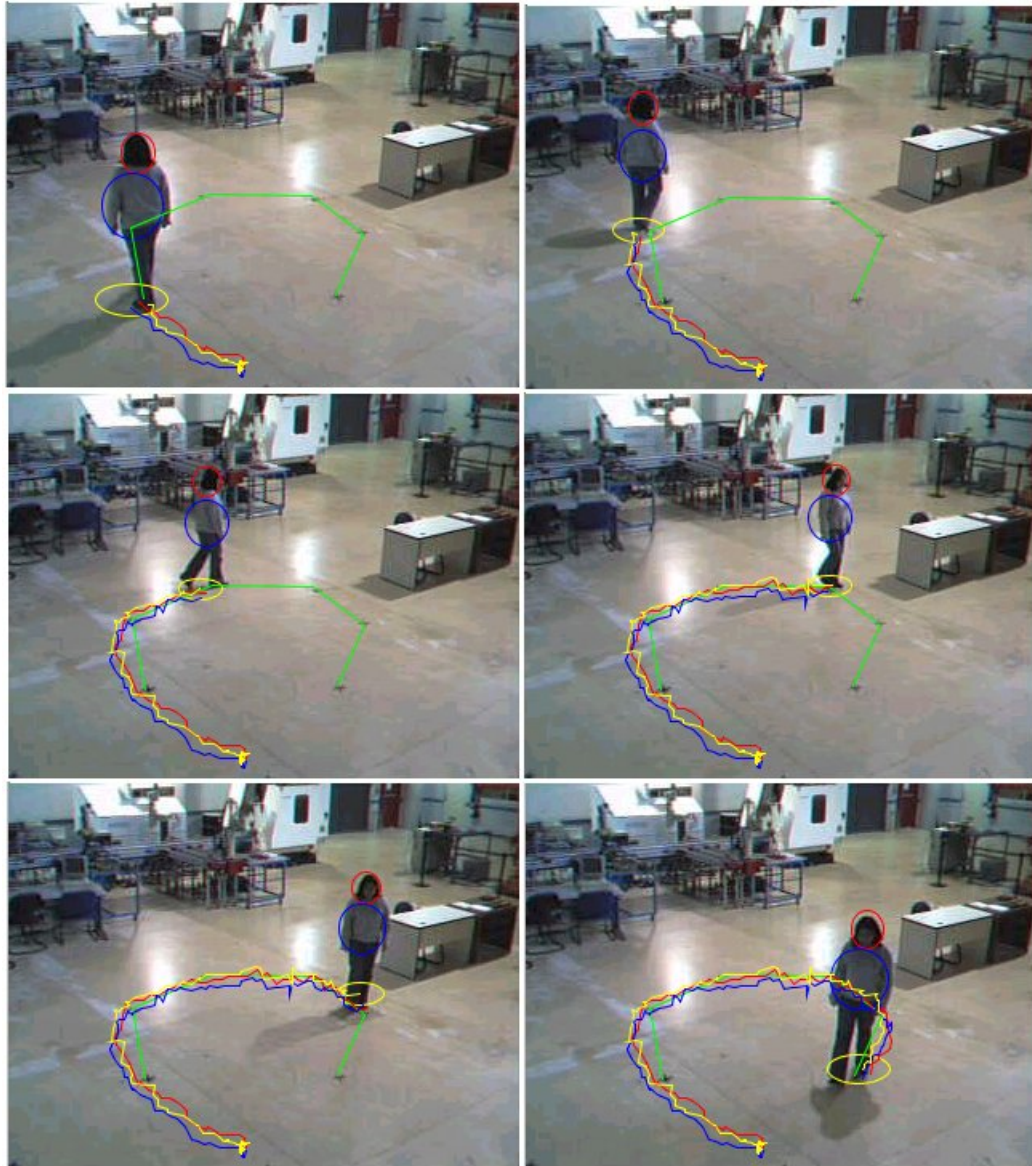


Figura 4.5: Secuencia de imágenes que muestra los resultados del seguimiento de una persona (*Prueba 1*)

Tabla 4.6: Errores correspondientes a las posiciones obtenidas mediante el seguidor del tronco

<i>Puntos de Control</i>	<i>Error Prueba 1</i>	<i>Error Prueba 2</i>	<i>Error Prueba 3</i>	<i>Error Promedio</i>
1	0.21m	0.31m	0.20m	0.24m
2	0.07m	0.27m	0.18m	0.17m
3	0.25m	0.35m	0.34m	0.32m
4	0.21m	0.30m	0.22m	0.24m
5	0.11m	0.24m	0.15m	0.17m
6	0.09m	0.17m	0.29m	0.19m

Tabla 4.7: Errores correspondientes a las posiciones obtenidas mediante el seguidor de los pies

<i>Puntos de Control</i>	<i>Error Prueba 1</i>	<i>Error Prueba 2</i>	<i>Error Prueba 3</i>	<i>Error Promedio</i>
1	0.12m	0.14m	0.24m	0.17m
2	0.11m	0.26m	0.22m	0.19m
3	0.19m	0.34m	0.41m	0.31m
4	0.06m	0.47m	0.44m	0.32m
5	0.36m	0.22m	0.32m	0.30m
6	0.09m	0.17m	0.39m	0.21m

es la parte del cuerpo que proporciona el menor error promedio. La proyección de la cabeza es más exacta debido a que el seguidor inicia en una posición probable, gracias al proceso de predicción relativa, lo cual facilita el seguimiento al igual que el hecho de que su distribución estadística es distintiva y prácticamente constante a lo largo del recorrido. Este situación no ocurre para el caso de los pies, cuya distribución cambia drásticamente al caminar. En el caso del tronco, puesto que se trata de la parte del cuerpo base, no existe punto de referencia durante el seguimiento. Además la altura utilizada por el modelo de perspectiva no siempre corresponde al centro de la ventana de seguimiento del tronco, lo que genera errores en la proyección.

La figura 4.6 muestra las posiciones obtenidas en las pruebas realizadas y la posición real de los puntos de control. Las posiciones de los seguidores de la cabeza, el tronco y los pies se identifican en la gráfica por signos diferentes. En la figura se observa que las posiciones obtenidas no tienen una distribución radial con respecto a la posición de cada marca. Por ejemplo, en el caso de las marcas 1 y 2 se observa una tendencia de acumulación hacia la izquierda. Esta tendencia se explica puesto que las personas no pasan exactamente por el centro de la marca, sino a un lado. Por otra parte, los errores

Tabla 4.8: Errores promedio de posiciones

	<i>Error Promedio</i>
Cabeza	0.20m
Tronco	0.22m
Pies	0.25m

obtenidos se atribuyen principalmente al método de seguimiento, ya que las ventanas no se ajustan exactamente a la forma del cuerpo de la persona. Esta variación provoca que el centro de la ventana de seguimiento no coincida con el centroide de la parte del cuerpo de interés, generando errores en el cálculo de las posiciones.

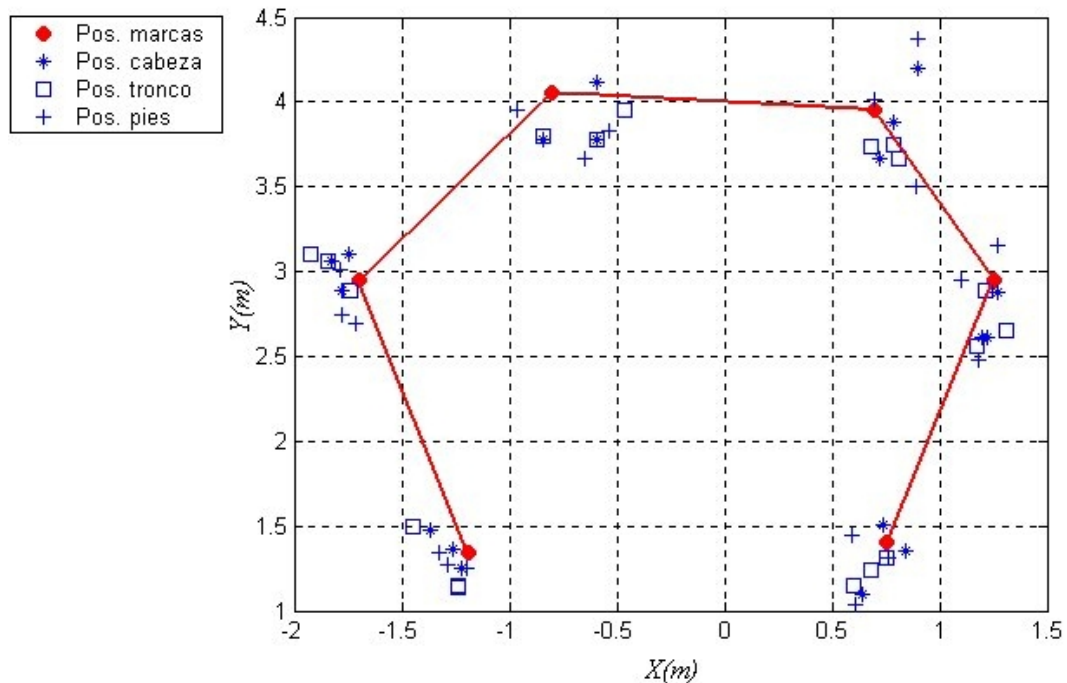


Figura 4.6: Gráfica de los errores correspondientes al cálculo de la posición utilizando la información del seguidor de la cabeza, del tronco y de los pies.

Asimismo, el error también tiene orígenes humanos: la medición de la posición de las marcas, la forma de caminar de la persona (debido a la longitud del paso) y el propio error de la persona al pasar por las marcas. Debe considerarse que no se cuenta con un mecanismo preciso para la medición del desplazamiento de la persona, motivo por el cual se utilizan las marcas en el piso, con la consabida imprecisión que acarrear los factores enunciados.

4.3 Seguimiento frontal

Los experimentos para validar el seguimiento de los movimientos de una persona frente a la cámara se realizaron en un cuarto con una pared lisa con el fin de facilitar el seguimiento de cada una de las partes del cuerpo. La figura 4.7 muestra las dimensiones del ambiente donde se realizaron las pruebas.

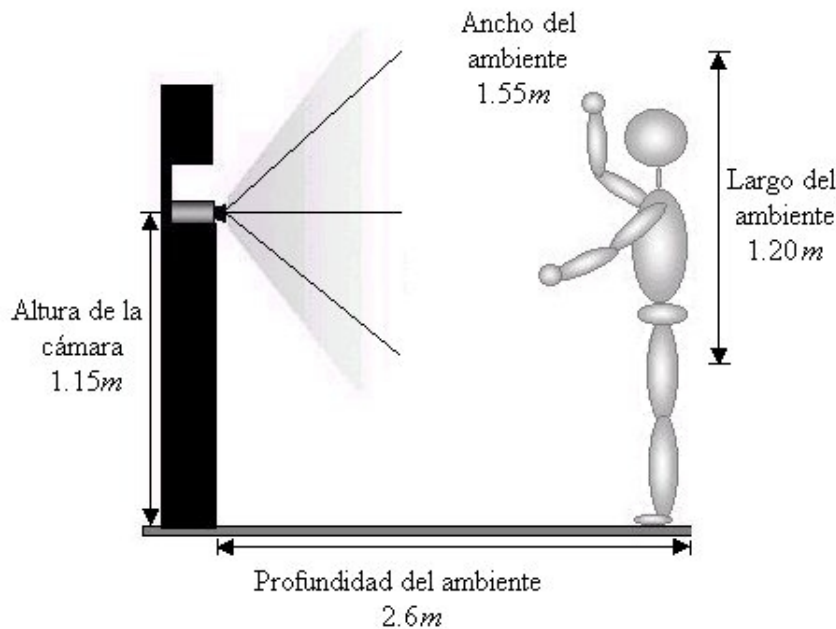


Figura 4.7: Medidas del ambiente captado por la cámara en el seguimiento frontal

Para un buen funcionamiento del sistema se recomienda respetar las siguientes condiciones:

1. El fondo del escenario debe contrastar con las partes del cuerpo de interés, para evitar que el seguidor pierda el objetivo.
2. Debido a que el seguidor compara la similitud entre dos distribuciones estadísticas se requiere que las partes del cuerpo a seguir sean distintivas. Por tanto, se recomienda que la persona utilice camisas a tres cuartos y que no se usen cuellos altos, con el propósito de utilizar el cambio entre la piel y la ropa para facilitar el seguimiento.

Por otra parte, para seleccionar las partes del cuerpo a seguir, es necesario que la persona adopte la posición de inicio que se muestra en la figura 4.8 y permanezca estática en esa posición hasta terminar el proceso de selección.

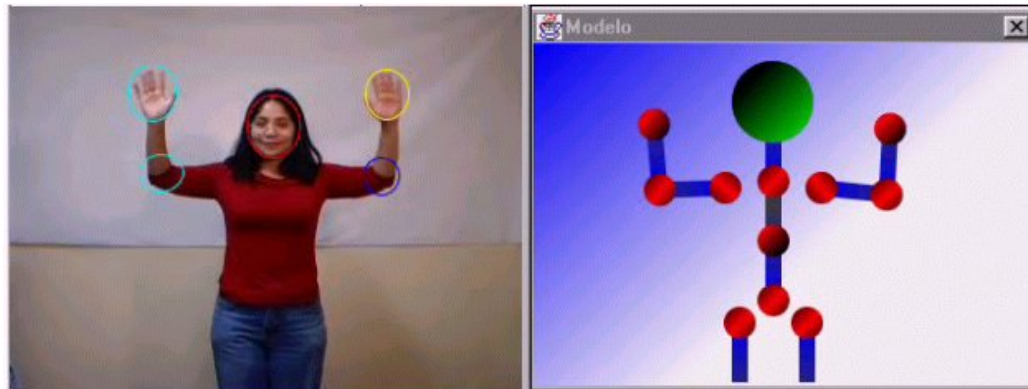


Figura 4.8: Posición de inicio para la selección de las ventanas de seguimiento

4.3.1 Medición del error de posición

Para medir el error de las posiciones obtenidas por el seguidor se requiere la posición real de las partes del cuerpo consideradas en el seguimiento frontal. Debido a que no se contó con algún sensor o parámetro real que permitiera obtener la posición exacta de las entidades durante el seguimiento, se realizó una aproximación para determinar la posición real. Tal aproximación consiste en obtener manualmente, en cada imagen de la secuencia, el punto en píxeles correspondiente al centro de la parte del cuerpo de interés. De manera que, para medir el error entre el movimiento realizado por la persona y la posición de la ventana dada por el seguidor, se obtuvo en cada imagen la posición real de la parte del cuerpo analizada y se comparó con el punto devuelto por el seguidor mediante el cálculo de la distancia euclidiana entre dos puntos, según la ecuación (4.1). Por tanto, el error obtenido es aproximado puesto que se midió de manera relativa y manual.

La posición del centro de cada una de las partes del cuerpo se determina en píxeles. Para obtener una analogía en centímetros se utilizó la relación que existe entre el número de píxeles que hay del inicio de la muñeca al codo y su medida en centímetros. Así, se obtuvo un valor aproximado de la equivalencia en centímetros de un píxel para una secuencia en particular.

4.3.2 Análisis de resultados

Como no se cuenta con un método automático para determinar la posición real de cada parte del cuerpo, no fue posible medir los errores en tiempo real. Por tanto, fue necesario utilizar una secuencia grabada de imágenes para determinar la posición real en píxeles de cada una de las partes del cuerpo de interés. La secuencia utilizada para analizar los errores consta de 100 imágenes y se obtuvo mediante una cámara de videoconferencia *PcCam300* de *CreativeLabs*.

En las tablas 4.9, 4.10 y 4.11 se presentan los resultados obtenidos durante el seguimiento frontal en la secuencia de prueba. La primera columna indica el número de imagen correspondiente a cada una de las mediciones. La columna *Cabeza Real* corresponde a las posiciones obtenidas directamente de las imágenes, mientras que la columna *Cabeza Virtual* presenta las posiciones calculadas por el seguidor, lo mismo ocurre para las demás partes del cuerpo. Las figuras 4.9 y 4.10 muestran las imágenes de la secuencia que se consideran en el análisis de resultados.

Tabla 4.9: Posiciones en pixeles obtenidas en el seguimiento de la cabeza y el torso

<i>No. de Cuadro</i>	<i>Cabeza Real</i>	<i>Cabeza Virtual</i>	<i>Torso Real</i>	<i>Torso Virtual</i>
1	(156,112)	(156,112)	(157,145)	(157,145)
15	(156,113)	(155,111)	(156,145)	(156,145)
30	(170,118)	(169,114)	(166,150)	(168,148)
45	(173,123)	(174,120)	(169,155)	(169,156)
60	(141,116)	(142,114)	(150,146)	(150,145)
80	(139,121)	(137,119)	(146,149)	(142,149)

Tabla 4.10: Posiciones en pixeles obtenidas en el seguimiento del codo y la mano izquierda

<i>No. de Cuadro</i>	<i>Codo Izq. Real</i>	<i>Codo Izq. Virtual</i>	<i>Mano Izq. Real</i>	<i>Mano Izq. Virtual</i>
1	(83,144)	(83,144)	(77,86)	(77,86)
15	(81,143)	(82,145)	(78,88)	(76,85)
30	(95,126)	(96,129)	(110,71)	(104,73)
45	(98,124)	(102,125)	(122,73)	(125,71)
60	(74,154)	(77,154)	(64,97)	(69,95)
80	(74,168)	(75,173)	(56,116)	(58,119)

Las tablas 4.12, 4.13 y 4.14 presentan el error promedio en pixeles y en centímetros, correspondientes a cada una de las partes del cuerpo.

Finalmente, en la tabla 4.15 se presenta el error promedio correspondiente a cada una de las partes del cuerpo de interés.

Como se observa, los errores más considerables se obtienen cuando existe una rotación, ya que las ventanas de seguimiento no se adaptan a dichos cambios, puesto que su orientación es fija. Por esta razón, las partes del cuerpo que presentaron mayor

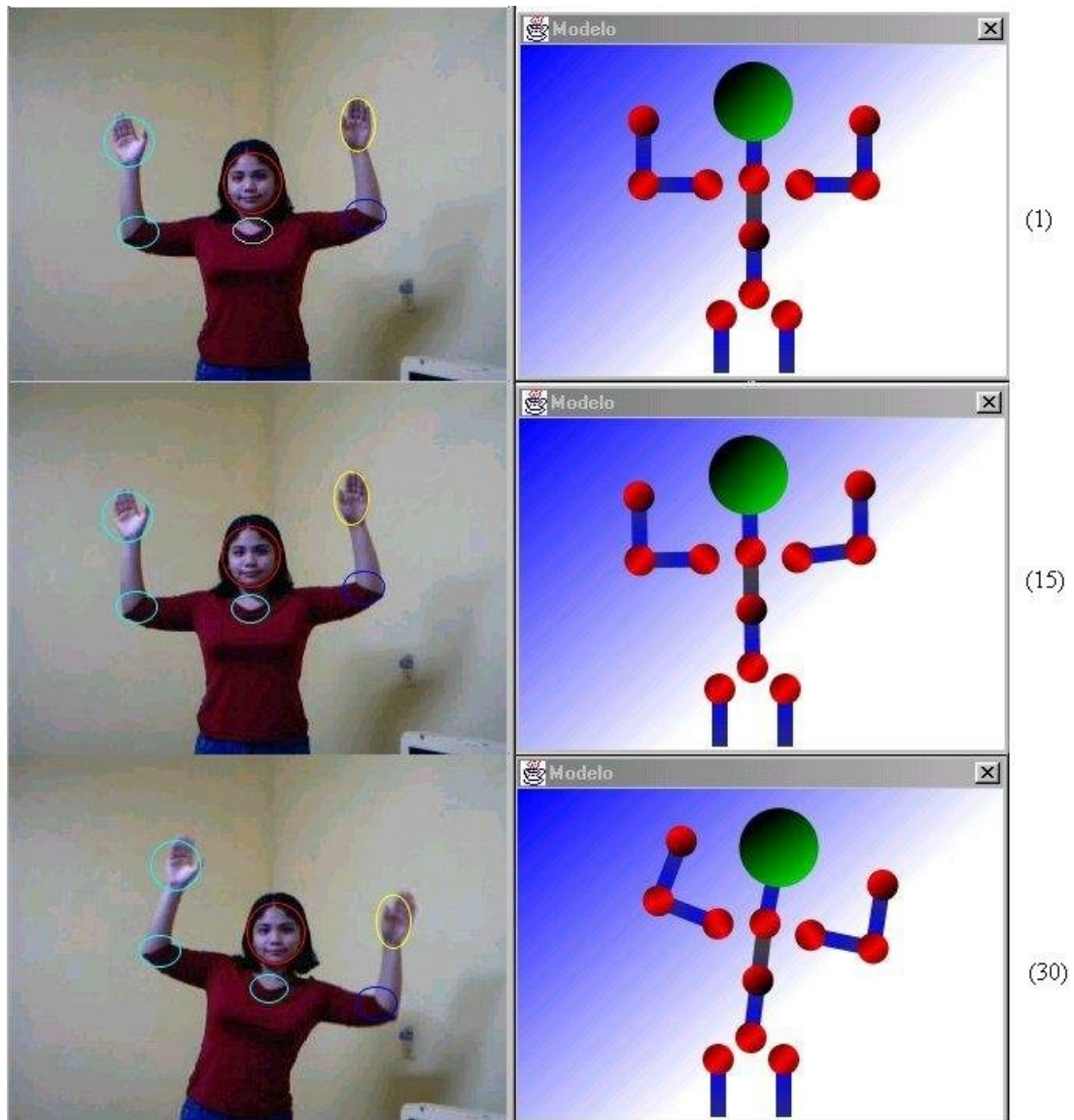


Figura 4.9: Secuencia de imágenes que muestran el seguimiento e imitación de los movimientos realizados por una persona frente a la cámara

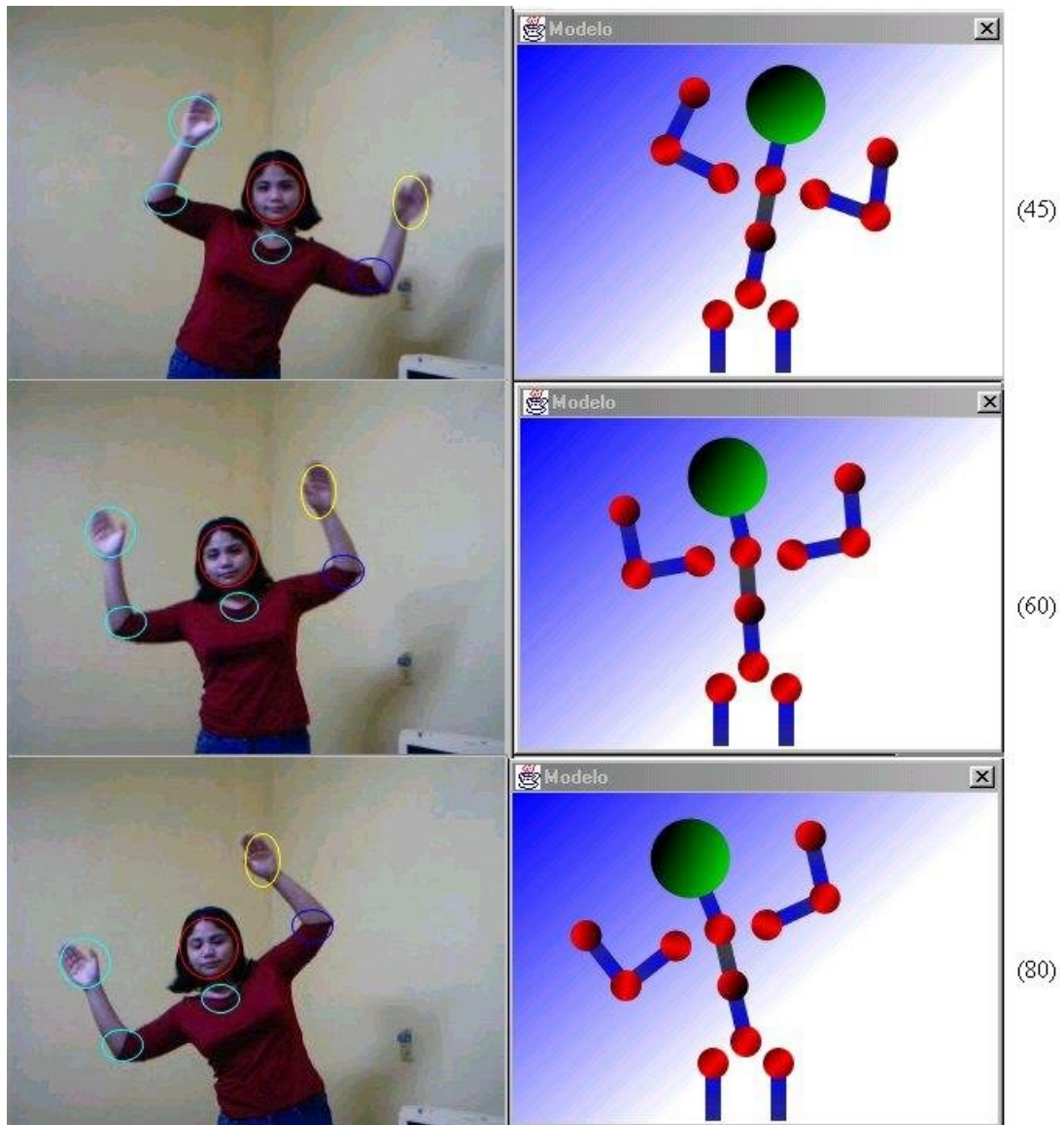


Figura 4.10: Secuencia de imágenes que muestran el seguimiento e imitación de los movimientos realizados por una persona frente a la cámara

Tabla 4.11: Posiciones en pixeles obtenidas en el seguimiento del codo y la mano derecha

No. de Cuadro	Codo Der. Real	Codo Der. Virtual	Mano Der. Real	Mano Der. Virtual
1	(229,133)	(229,133)	(224,75)	(224,75)
15	(231,132)	(227,132)	(221,77)	(221,80)
30	(241,162)	(237,156)	(249,107)	(245,102)
45	(238,175)	(233,176)	(258,124)	(260,131)
60	(218,121)	(218,125)	(201,70)	(205,70)
80	(210,111)	(206,112)	(179,65)	(176,66)

Tabla 4.12: Errores correspondientes a las posiciones de la cabeza y el torso

No. de Cuadro	Cabeza (pixeles)	Cabeza (cm)	Torso (pixeles)	Torso (cm)
1	0	0	0	0
15	2.23	1.34	0	0
30	4.12	2.47	2.82	1.69
45	3.16	1.89	1	0.60
60	1.41	0.84	1	0.60
80	2.82	1.69	4	2.40

error fueron la mano y el codo derecho ya que tienen movimientos más drásticos en cuanto a cambios de orientación.

La reproducción de los movimientos realizados por la persona requiere conocer los ángulos a los que se deben mover cada una de las articulaciones del modelo virtual. Por tanto, es conveniente analizar no sólo los errores en posición, sino también los errores de los ángulos estimados.

La figura 4.11, muestra las gráficas de la trayectoria real y la trayectoria estimada por el seguidor para el caso del torso. La medición de la posición real de cada una de las partes del cuerpo se realizó manualmente en la secuencia, considerando intervalos de 5 imágenes. El resto de las posiciones se obtuvieron mediante interpolación, utilizando el programa MATLAB. En la gráfica 4.11 se indica el número de cuadro correspondiente a las posiciones consideradas en las tablas de posiciones. La descripción de *Modelo virtual* corresponde a los datos obtenidos con base en el algoritmo de seguimiento y la etiqueta *Medición Real* corresponde a los datos que se determinaron manualmente en la secuencia. Como se puede observar, el torso presenta un movimiento de derecha a izquierda a partir de su posición inicial. Los errores más considerables se presentan en las imágenes 30 y 80. En el caso de la imagen 30, el error se debe a que se inicia un

Tabla 4.13: Errores correspondientes a las posiciones del codo y la mano derecha

<i>No. de Cuadro</i>	<i>Codo Der. (pixeles)</i>	<i>Codo Der. (cm)</i>	<i>Mano Der. (pixeles)</i>	<i>Mano Der. (cm)</i>
1	0	0	0	0
15	4	2.40	3.0	1.8
30	7.21	4.32	6.40	3.84
45	5.09	3.05	7.28	4.36
60	4	2.40	4	2.40
80	4.12	2.47	3.16	1.89

Tabla 4.14: Errores correspondientes a las posiciones del codo y la mano izquierda

<i>No. de Cuadro</i>	<i>Codo Izq. (pixeles)</i>	<i>Codo Izq. (cm)</i>	<i>Mano Izq. (pixeles)</i>	<i>Mano Izq. (cm)</i>
1	0	0	0	0
15	2.23	1.34	3.60	2.16
30	3.16	1.89	6.32	3.79
45	4.12	2.47	3.60	2.16
60	3	1.80	5.38	4.24
80	5.09	3.05	3.60	2.16

movimiento rápido en la secuencia, y en la imagen 80 existe un cambio de dirección del movimiento, lo que provoca cambios drásticos en la distribución modelo.

La figura 4.12 muestra los ángulos reales y los estimados a partir de la posición de la ventana de seguimiento del torso. Como se observa en la gráfica, el ángulo inicial del torso es de 90° con respecto a la cintura. A partir del cuadro 15 se presenta un movimiento hacia la derecha de aproximadamente 70° , para después invertir el sentido del movimiento hacia la izquierda. Los errores más notables de la gráfica se encuentran en los cuadros donde el torso presenta un cambio en la dirección del movimiento, como sucede alrededor de los cuadros 40 (de derecha a izquierda) y 70 (de izquierda a derecha) respectivamente.

Las figuras 4.13, 4.14 y 4.15 muestran las gráficas de los ángulos de la cabeza, el codo y la mano izquierda respectivamente. Los ángulos se obtienen con respecto a la parte del cuerpo inmediata superior del modelo de estructura de la figura 2.5. Al igual que en el caso del torso, se cumple que los errores angulares más considerables se presentan cuando existe un cambio drástico en la dirección del movimiento.

Tabla 4.15: Error promedio correspondiente a la estimación de la posición de las partes del cuerpo de interés.

<i>Región de interés</i>	<i>Error Promedio (Píxeles)</i>	<i>Error Promedio (cm)</i>
Cabeza	4.04	2.42cm
Torso	2.29	1.37cm
Codo Derecho	5.12	3.07cm
Mano Derecha	5.25	3.15cm
Codo Izquierdo	3.97	2.38cm
Mano Izquierda	3.97	2.38cm

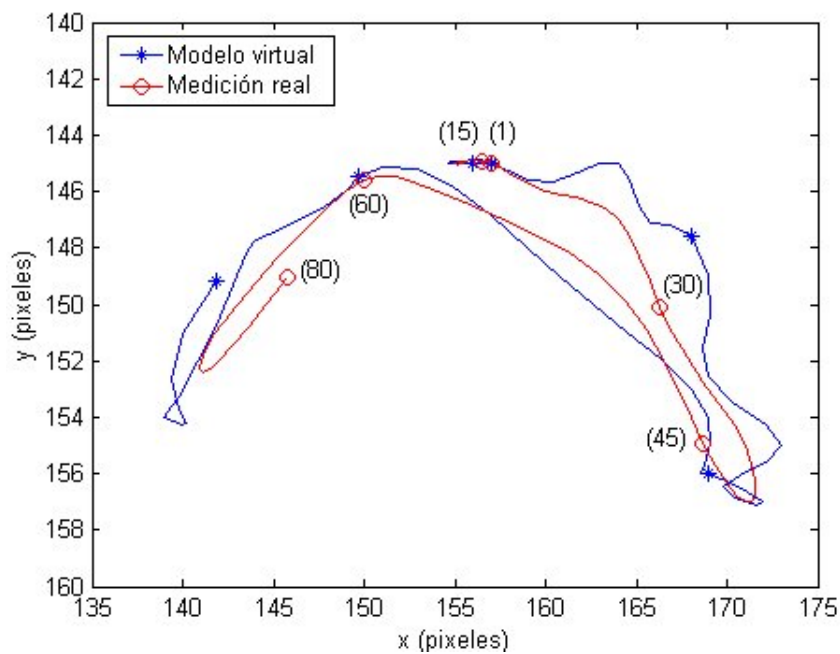


Figura 4.11: Trayectoria generada por el movimiento del torso. La gráfica muestra las posiciones correspondientes a las imágenes analizadas de la secuencia y el resto de los datos se obtienen interpolando cada 5 imágenes

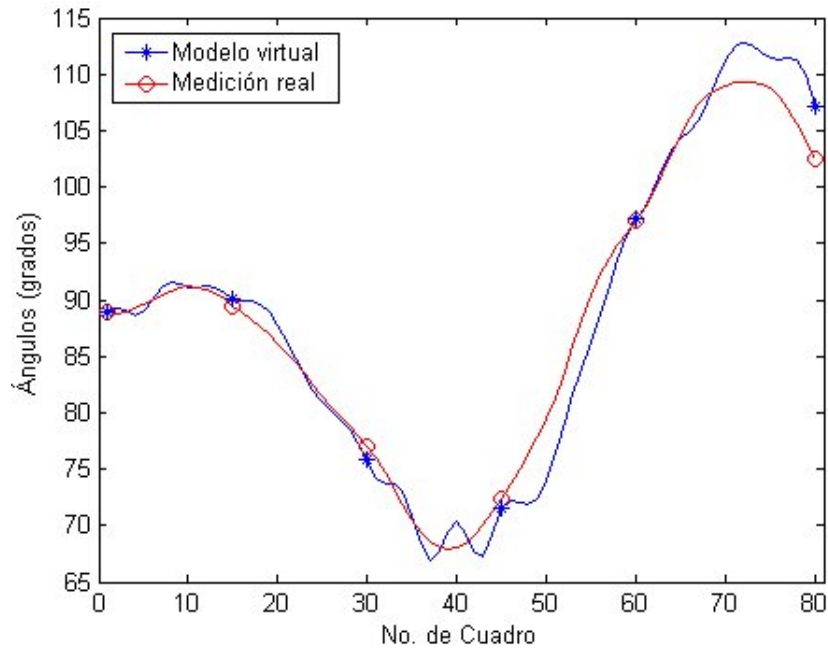


Figura 4.12: Ángulos definidos por el movimiento del torso

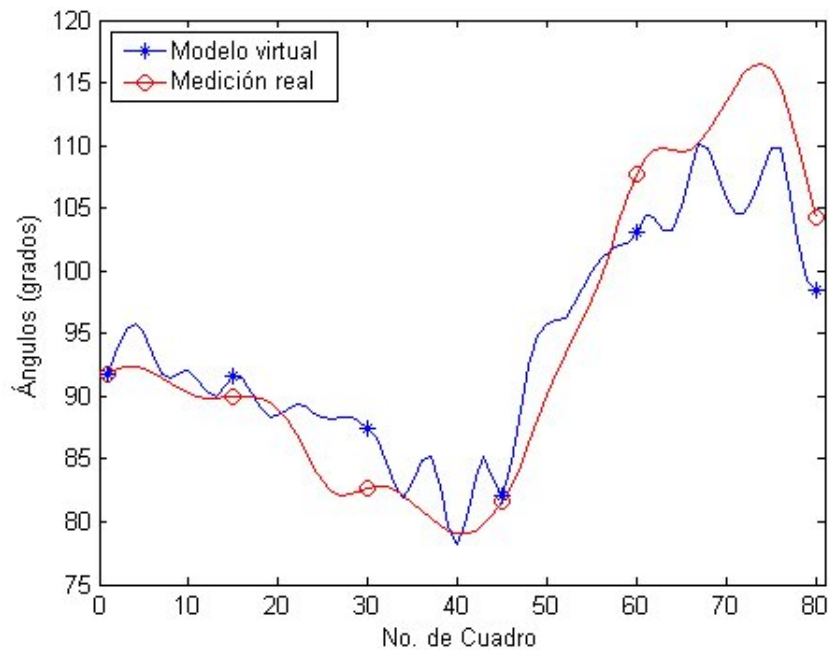


Figura 4.13: Ángulos definidos durante el movimiento de la cabeza

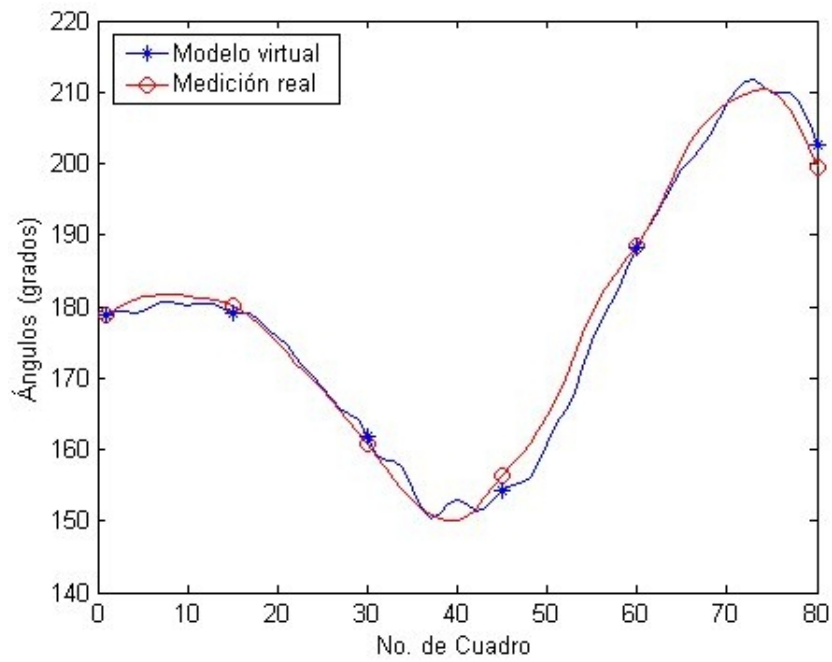


Figura 4.14: Ángulos estimados durante el movimiento del codo izquierdo

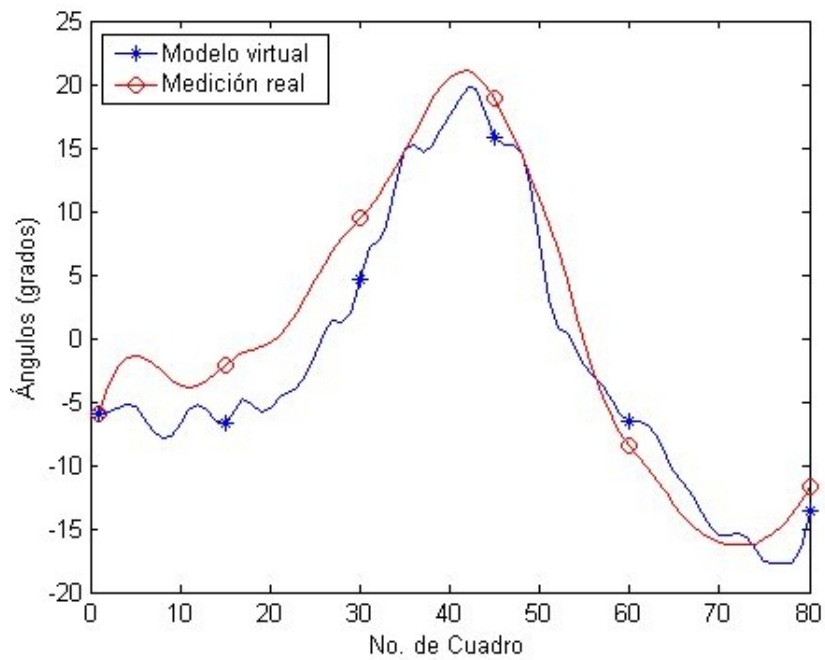


Figura 4.15: Ángulos definidos por el movimiento de la mano izquierda

Capítulo 5

Conclusiones

Esta tesis describe un método de seguimiento de personas basado en la estructura cinemática del cuerpo humano. El método utiliza la técnica integrativa conocida como algoritmo de desplazamiento medio para realizar el seguimiento, y emplea restricciones cinemáticas de la estructura humana para reducir el espacio de búsqueda y validar las posiciones devueltas por el seguidor.

El método propuesto considera al cuerpo humano como una estructura articulada en la que, a partir de una entidad de inicio (torso), se aplican restricciones de estructura (distancias relativas, ángulos, etc.) para predecir la posición de las demás partes del cuerpo, con base a un modelo jerárquico de dependencias cinemáticas. De tal manera, las restricciones cinemáticas del cuerpo humano guían la aplicación del seguidor de desplazamiento medio, para mejorar el seguimiento de la persona y obtener información de estructura necesaria para la reproducción de sus movimientos.

El algoritmo de seguimiento mediante desplazamiento medio consiste en obtener la distribución estadística del objeto de interés (distribución modelo) y del objeto candidato, comparar ambas distribuciones mediante un coeficiente de similitud y estimar la posición del objeto de interés a partir del *vector de desplazamiento medio*, el cual maximiza la correlación que existe entre las dos distribuciones estadísticas. Este vector se calcula iterativamente hasta encontrar la distribución más similar a la del objeto modelo.

Para fines experimentales se utilizaron dos niveles diferentes de abstracción del cuerpo humano. Una abstracción se emplea para seguir a una persona caminando (seguimiento con profundidad). Mientras la otra se usa para seguir los movimientos de una persona frente a la cámara, con el propósito de reproducirlos mediante un modelo virtual de la persona (seguimiento frontal).

El seguimiento de una persona frente a la cámara considera un modelo de abstracción que comprende la cabeza, el torso y los brazos (codos y manos); el objetivo es capturar la información de movimiento y reproducirla a través de un modelo virtual de la persona. El algoritmo propuesto realiza una exploración jerárquica sobre las

articulaciones, tomando en cuenta las restricciones cinemáticas inherentes al cuerpo humano. La principal ventaja del sistema implementado es que no requiere de marcas, ni sensores en el cuerpo, para reproducir los movimientos efectuados por la persona, a diferencia de otros algoritmos. Una suposición importante en la implementación, es que los movimientos del individuo siempre deben ser paralelos al plano de visión de la cámara, lo cual limita los posibles movimientos a dos dimensiones.

Por otra parte, la abstracción utilizada para seguir a una persona caminando consta de la cabeza, el tronco y los pies. La cámara con la cual se capturan las imágenes se encuentra ubicada en perspectiva con relación al ambiente; por tanto, el área de trabajo está delimitada por el área de visión de la cámara. Así, para obtener la posición real de la persona se utiliza un modelo matemático de perspectiva que se basa en medidas del ambiente y en la altura de la persona.

Gracias a la ubicación en perspectiva del sistema de visión se dispone de un espacio de trabajo mayor al de una ubicación perpendicular. En contraparte, esta ubicación implica cambios de escala del objeto a seguir. El algoritmo original de desplazamiento medio no se adapta a cambios de escala, motivo por el cual se propuso utilizar información del ambiente y el modelo geométrico de perspectiva para obtener un factor de escala. En función de este factor se aumenta o disminuye la ventana de seguimiento, dependiendo de la posición de la persona en el ambiente. Además, se modificó el algoritmo de desplazamiento medio para que la distribución modelo no sea estática, sino que se modifique durante el seguimiento. Así, entre las principales contribuciones de este trabajo se encuentran la propuesta de un método de adaptación a los cambios de escala y la incorporación de una distribución estadística dinámica de los objetos a seguir.

En general, los resultados de los experimentos permitieron verificar el método de seguimiento propuesto. En el caso del seguimiento frontal, los resultados mostraron una correspondencia entre el movimiento de la persona y el movimiento simulado por el modelo virtual. Los errores más considerables se presentaron en las partes del cuerpo con mayor rotación en la secuencia. Ahí, el seguimiento se dificulta puesto que la orientación de las ventanas de seguimiento es fija.

Por otra parte, el método para el seguimiento de una persona caminando tuvo un desempeño favorable, siendo la cabeza la parte del cuerpo más confiable para el cálculo de la posición de la persona. El error obtenido se atribuye principalmente a que el método no se ajusta exactamente a la forma del cuerpo de la persona, dando como resultado que el centro de la ventana de seguimiento no coincida exactamente con el centroide de la parte del cuerpo de interés. Además, se generan errores debido a la perspectiva.

Las pruebas realizadas para verificar el desempeño mostraron que el seguimiento se realizó en tiempo real con el equipo disponible, según la tasa de captura de 15 imágenes por segundo. El tiempo de procesado del sistema depende principalmente del tamaño

de las ventanas de seguimiento, del número de niveles utilizados en las distribuciones de color y del número de iteraciones del vector de desplazamiento medio.

El método de seguimiento aprovecha las propiedades de la estructura cinemática del cuerpo humano para disminuir y validar el espacio de búsqueda del algoritmo original de desplazamiento medio, haciéndolo más robusto. En el seguimiento frontal, se comprobó que el sistema recupera las posiciones y los ángulos de las diferentes partes del cuerpo humano, para la simulación de los movimientos realizados por la persona, sin requerir de sensores físicos ni marcas visuales. Por otra parte, en el seguimiento con profundidad se obtuvo la posición real de la persona en el ambiente, a partir del modelo de perspectiva, sin la necesidad de utilizar más de una cámara.

Para efectuar el seguimiento, es deseable que el objeto de interés tenga cierto contraste con el fondo; de preferencia se busca que el objeto tenga textura mientras que el fondo sea liso, de manera que su distribución sea distintiva. Además, la velocidad de captura de las imágenes y la velocidad de procesamiento deben ser lo suficientemente rápidas para evitar cambios drásticos en la posición y apariencia de los objetos a seguir.

Una desventaja del método es que los cambios de iluminación modifican la distribución estadística del objeto de interés, lo que dificulta el seguimiento, puesto que disminuye la similitud con respecto a la distribución modelo.

Para incrementar la robustez del sistema se propone, como trabajo futuro, utilizar una representación de color menos sensible a los cambios de iluminación, como pueden ser el modelo HSV (Hue, Saturation, Value) o un modelo de color basado en lógica difusa. Además, para el caso de seguimiento con cámara fija, se propone realizar una substracción del fondo, considerando un modelo estadístico como el introducido por [Stauffer 99], para mejorar el seguimiento.

Por otro lado, se propone el desarrollo de un algoritmo que permita resolver el problema de escala para cualquier secuencia de imágenes, sin requerir del modelo de perspectiva del ambiente. Una idea es considerar diferentes tamaños de la ventana de seguimiento y seleccionar el que proporcione un mayor coeficiente de similitud.

Por último, se propone la inclusión de movimientos más complejos para su reproducción mediante un modelo virtual de la persona, considerando las restricciones cinemáticas del cuerpo humano. Así, al contar con un modelo más completo de la estructura humana y realizar ciertas consideraciones en el ambiente, como la estimación de profundidad, se podrían seguir movimientos tridimensionales.

Apéndice A

Modelo geométrico de perspectiva

Para obtener la posición de la persona en el ambiente de trabajo, a partir de una imagen capturada por la cámara, se utiliza un modelo matemático de perspectiva basado en [Vázquez 02] y [Aguilar 03]. Gracias a este modelo de perspectiva se determina la proyección de un punto en la imagen en su posición real sobre el ambiente, con base en ciertas medidas del espacio de trabajo y la altura de la persona. En este apéndice se describen los detalles del modelo.

A.1 Notación matemática

La notación que se utilizará a lo largo del apéndice es la siguiente:

Punto: Los puntos se representan con letras minúsculas y en cursiva (a, b, c , etc.)

Segmento de línea: El segmento de línea del punto a al punto b , se representa con el símbolo \overline{ab} , mientras que la magnitud de \overline{ab} se denota como $|\overline{ab}|$ o ab

Triángulo: La notación que se utiliza para representar a un triángulo (figura A.1) es $\triangle abc$, cuyos vértices son los puntos a, b y c , y sus lados son los segmentos de línea \overline{ab} , \overline{bc} , y \overline{ac} .

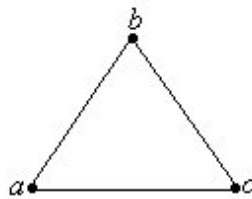


Figura A.1: Triángulo $\triangle abc$

Ángulo: Para denotar un ángulo se utilizan letras del alfabeto griego (α, β , etc.). La amplitud del ángulo sigue el principio de la “mano derecha” como se muestra en la figura A.2.

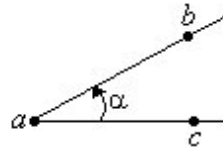


Figura A.2: Representación de un ángulo

A.2 Transformación de coordenadas

El ambiente donde se desplaza la persona se define por el plano XY , mientras que el sistema visual se encuentra ubicado en perspectiva respecto al espacio de trabajo, como se muestra en la figura A.3. Por ende, se realiza una transformación de coordenadas para determinar la correspondencia de las coordenadas en el sistema visual, respecto a las coordenadas en el ambiente real.

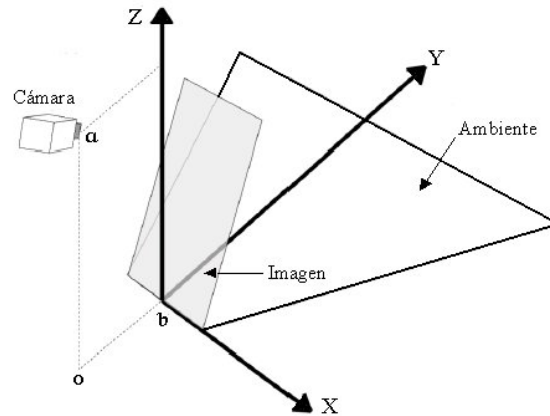


Figura A.3: Esquema del montaje de la cámara y ejes de referencia establecidos para el plano del ambiente

La transformación tiene como entrada las coordenadas (x_{img}, y_{img}) que corresponden a la ubicación de la persona en la imagen y como salida las coordenadas de la posición de la persona en el ambiente real (x_{amb}, y_{amb}) . El tamaño de la imagen obtenida por la cámara es: $a_{img} \times l_{img}$ donde a_{img} denota el ancho y l_{img} el largo de la imagen en píxeles, como se muestra en la figura A.4. Por otro lado, las dimensiones del ambiente de trabajo son: $a_{amb} \times l_{amb}$ donde a_{amb} y l_{amb} denotan el ancho y el largo del ambiente, respectivamente.

En la figura A.5 se observa el esquema del montaje realizado: la cámara está colocada en a , y la posición de la persona en el ambiente (x_{img}, y_{img}) corresponde al punto e , ubicado en el eje Y sobre el espacio de trabajo. Además, en la figura se observan los siguientes elementos obtenidos por medición sobre el ambiente:

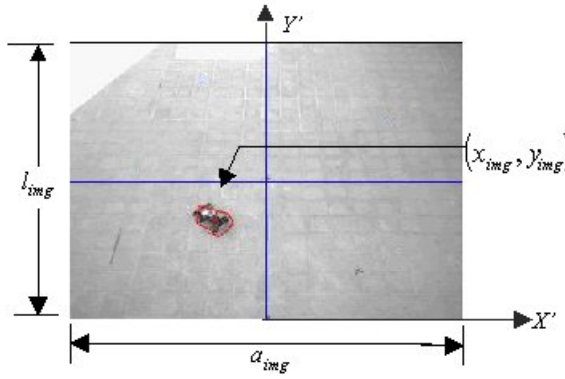


Figura A.4: Imagen obtenida por la cámara

ao: altura a la que se encuentra colocada la cámara.

ob: distancia al primer punto visible del ambiente. La distancia se mide a partir del punto donde se proyecta perpendicularmente la posición de la cámara, sobre el plano del ambiente de trabajo.

bf: distancia del primer punto observable, al punto en el ambiente correspondiente al centro de la imagen captada por la cámara, también sobre el plano del espacio de trabajo.

A.2.1 Proyección sobre el eje *Y*

Basándose en el esquema de la vista transversal del montaje mostrado en la figura A.5 se obtendrá $y_{proy} = be$, que representa la posición de la persona en el eje *Y* del espacio de trabajo, sin considerar la altura de la persona.

El desarrollo comienza con la determinación de la magnitud del segmento \overline{bc} que representa la proyección perpendicular de la posición y_{img} de la persona en el segmento \overline{bg} . La magnitud de \overline{bg} es la distancia que se capta en el sistema de visión y corresponde a la resolución total de la cámara en el eje *Y* (l_{img}).

De acuerdo con la ubicación, en y_{img} , de la persona y el largo de la imagen l_{img} , la longitud del segmento de recta \overline{bc} se determina por la relación,

$$bc = bg \frac{y_{img}}{l_{img}} \quad (A.1)$$

donde y_{img} y l_{img} son medidas en pixeles, mientras que bg y bc se definen en metros

Para obtener la magnitud del segmento \overline{bg} se aplica ley de senos al triángulo $\triangle abg$,

$$bg = \frac{ab \operatorname{sen} \beta}{\operatorname{sen} \gamma} \quad (A.2)$$

Por tanto, sustituyendo la ecuación (A.2) en (A.1),

$$bc = \frac{ab \operatorname{sen}\beta}{\operatorname{sen}\gamma} \frac{y_{img}}{l_{img}} \quad (\text{A.8})$$

donde y_{img} y l_{img} son medidas en pixeles, mientras que bc y ab se definen en metros. La ecuación anterior (A.8) representa la proyección del espacio percibido por la cámara en el ambiente donde se desplaza la persona. Una vez obtenida la magnitud del segmento \overline{bc} el siguiente paso es determinar la expresión para $be = y_{proy}$ correspondiente a la posición de la persona en el ambiente.

En la figura se observa que

$$be = oe - ob \quad (\text{A.9})$$

y del triángulo $\triangle oae$,

$$oe = ao \tan \alpha \quad (\text{A.10})$$

Por suma de ángulos en el mismo triángulo se deduce,

$$\alpha = \delta + \tan^{-1} \left(\frac{ob}{ao} \right) \quad (\text{A.11})$$

donde al utilizar la ley de senos en el triángulo $\triangle abc$ y despejando,

$$\delta = \operatorname{sen}^{-1} \left(\frac{bc \operatorname{sen}\gamma}{ac} \right) \quad (\text{A.12})$$

y al emplear la ley de cosenos

$$ac^2 = ab^2 + bc^2 - 2(ab)(bc) \cos \gamma \quad (\text{A.13})$$

Por tanto, la ubicación de la persona sobre el eje Y a nivel del piso y_{proy} , considerando la posición de los pies de la persona en la imagen, se obtiene sustituyendo (A.10) en (A.9),

$$y_{proy} = be = ao \tan \alpha - ob \quad (\text{A.14})$$

Para el caso del tronco y la cabeza se debe incluir una compensación en la transformación anterior para considerar su altura, tomando en cuenta, que no se encuentran ubicados al nivel del piso. Utilizando la altura de la persona se realiza un ajuste para evitar que, debido a la perspectiva, se estime una posición de la persona sobre el eje Y mayor a la que en realidad se encuentra. En la figura A.6 se observa que la persona se encuentra en la posición p y su altura es la magnitud del segmento \overline{hp} , por lo que la posición y_{amb} que se debe obtener es la proyección al nivel del piso considerando la altura.

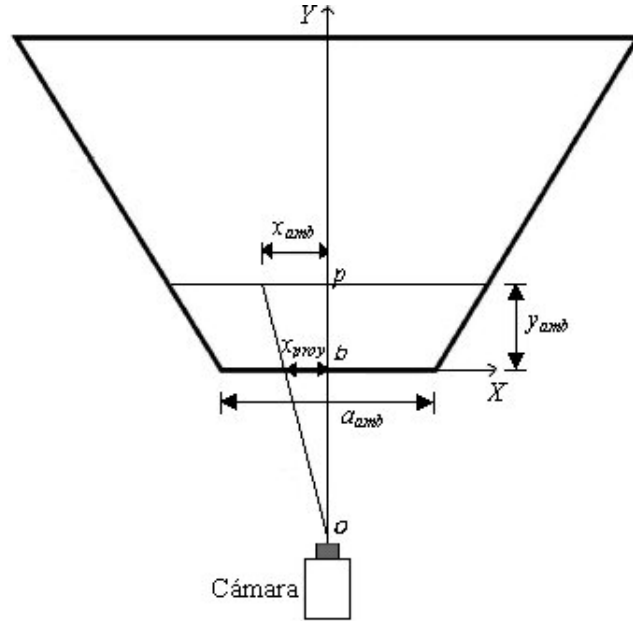


Figura A.7: Posición de la persona sobre el eje X en el área de trabajo

donde x_{proy} corresponde a la proyección de x_{img} en el plano del ambiente,

$$x_{proy} = \frac{a_{amb}}{a_{img}} \left(x_{img} - \frac{a_{img}}{2} \right) \quad (\text{A.19})$$

con x_{img} y a_{img} en píxeles y x_{proy} y a_{amb} en metros. El ancho del ambiente a la altura del primer punto captado por la cámara (a_{amb}), se mide directamente en el espacio de trabajo.

Por tanto, despejando de la ecuación (A.18),

$$x_{amb} = \frac{x_{proy} (ob + y_{amb})}{ob} \quad (\text{A.20})$$

De acuerdo con lo anterior, la posición de la persona en el ambiente real se determina mediante las ecuaciones correspondientes a las coordenadas (x_{amb}, y_{amb}) .

Apéndice B

Expansión de Taylor del coeficiente de Bhattacharyya

El algoritmo de seguimiento mediante desplazamiento medio, inicia la búsqueda de la localización del objeto de interés a lo largo de la secuencia de imágenes, en el punto $\mathbf{y}_0 = (x_0, y_0)$ que corresponde a la posición del objeto de interés en la imagen anterior de la secuencia. Por tanto, la búsqueda se realiza en la vecindad de \mathbf{y}_0 , por lo que es conveniente realizar una aproximación de la función de similitud en las cercanías de este punto. Para este fin se aplica la expansión de Taylor alrededor de los valores de la función de densidad del objeto candidato $\hat{\mathbf{p}}_u(\mathbf{y}_0)$.

El objetivo de la expansión de Taylor es aproximar la evaluación de una función en las cercanías de un punto. La expansión de Taylor de primer orden, en el punto a de la función $f(x)$, se define por:

$$f(x) \approx f(a) + f'(a)(x - a) \quad (\text{B.1})$$

En este caso, la función considerada es el coeficiente de Bhattacharyya,

$$\rho[\hat{\mathbf{p}}_u(\mathbf{y}), \hat{\mathbf{q}}_u] = \sum_{u=1}^m \sqrt{\hat{p}_u(\mathbf{y}) \hat{q}_u} \quad (\text{B.2})$$

donde $\hat{\mathbf{q}}_u$ es la distribución estadística del objeto de interés, $\hat{\mathbf{p}}_u(\mathbf{y})$ es la distribución estadística del objeto candidato, el subíndice u indica que las distribuciones se obtienen con base en características de color $u = 1 \dots m$, donde m es el número total de colores en la imagen. Por su parte, los valores alrededor de los cuales se expande el coeficiente son aquellos de la distribución $\hat{\mathbf{p}}_u(\mathbf{y}_0)$ del objeto candidato en la posición \mathbf{y}_0 (posición del objeto de interés en la imagen anterior de la secuencia). Aplicando la ecuación (B.1) al coeficiente de Bhattacharyya,

$$\rho[\hat{\mathbf{p}}_u(\mathbf{y}), \hat{\mathbf{q}}_u] \approx \sum_{u=1}^m \sqrt{\hat{p}_u(\mathbf{y}_0) \hat{q}_u} + \sum_{u=1}^m \frac{1}{2\sqrt{\hat{p}_u(\mathbf{y}_0) \hat{q}_u}} (\hat{p}_u(\mathbf{y}) \hat{q}_u - \hat{p}_u(\mathbf{y}_0) \hat{q}_u) \quad (\text{B.3})$$

multiplicando términos,

$$\rho[\hat{\mathbf{p}}_u(\mathbf{y}), \hat{\mathbf{q}}_u] \approx \sum_{u=1}^m \sqrt{\hat{p}_u(\mathbf{y}_0) \hat{q}_u} + \sum_{u=1}^m \frac{\hat{p}_u(\mathbf{y}) \hat{q}_u}{2\sqrt{\hat{p}_u(\mathbf{y}_0) \hat{q}_u}} - \sum_{u=1}^m \frac{\hat{p}_u(\mathbf{y}_0) \hat{q}_u}{2\sqrt{\hat{p}_u(\mathbf{y}_0) \hat{q}_u}} \quad (\text{B.4})$$

por propiedad de linealidad de la sumatoria,

$$\rho[\hat{\mathbf{p}}_u(\mathbf{y}), \hat{\mathbf{q}}_u] \approx \sum_{u=1}^m \sqrt{\hat{p}_u(\mathbf{y}_0) \hat{q}_u} + \frac{1}{2} \sum_{u=1}^m \frac{\hat{p}_u(\mathbf{y}) \hat{q}_u}{\sqrt{\hat{p}_u(\mathbf{y}_0) \hat{q}_u}} - \frac{1}{2} \sum_{u=1}^m \frac{\hat{p}_u(\mathbf{y}_0) \hat{q}_u}{\sqrt{\hat{p}_u(\mathbf{y}_0) \hat{q}_u}} \quad (\text{B.5})$$

reordenando,

$$\rho[\hat{\mathbf{p}}_u(\mathbf{y}), \hat{\mathbf{q}}_u] \approx \sum_{u=1}^m \sqrt{\hat{p}_u(\mathbf{y}_0) \hat{q}_u} + \frac{1}{2} \sum_{u=1}^m \hat{p}_u(\mathbf{y}) \sqrt{\frac{\hat{q}_u^2}{\hat{p}_u(\mathbf{y}_0) \hat{q}_u}} - \frac{1}{2} \sum_{u=1}^m \sqrt{\frac{(\hat{p}_u(\mathbf{y}_0) \hat{q}_u)^2}{\hat{p}_u(\mathbf{y}_0) \hat{q}_u}} \quad (\text{B.6})$$

simplificando enseguida,

$$\rho[\hat{\mathbf{p}}_u(\mathbf{y}), \hat{\mathbf{q}}_u] \approx \sum_{u=1}^m \sqrt{\hat{p}_u(\mathbf{y}_0) \hat{q}_u} + \frac{1}{2} \sum_{u=1}^m \hat{p}_u(\mathbf{y}) \sqrt{\frac{\hat{q}_u}{\hat{p}_u(\mathbf{y}_0)}} - \frac{1}{2} \sum_{u=1}^m \sqrt{\hat{p}_u(\mathbf{y}_0) \hat{q}_u} \quad (\text{B.7})$$

por tanto, la expansión de Taylor del coeficiente de Bhattacharyya resulta:

$$\rho[\hat{\mathbf{p}}_u(\mathbf{y}), \hat{\mathbf{q}}_u] \approx \frac{1}{2} \sum_{u=1}^m \sqrt{\hat{p}_u(\mathbf{y}_0) \hat{q}_u} + \frac{1}{2} \sum_{u=1}^m \hat{p}_u(\mathbf{y}) \sqrt{\frac{\hat{q}_u}{\hat{p}_u(\mathbf{y}_0)}} \quad (\text{B.8})$$

La expresión anterior es válida si la distribución del objeto candidato $\hat{\mathbf{p}}_u(\mathbf{y})$ no cambia drásticamente con respecto a la distribución anterior $\hat{\mathbf{p}}_u(\mathbf{y}_0)$ y $\hat{\mathbf{p}}_u(\mathbf{y}_0) > 0$ para toda $u = 1 \dots m$. Sustituyendo la expresión para la distribución $\hat{\mathbf{p}}_u(\mathbf{y})$ en la ecuación (B.8),

$$\rho[\hat{\mathbf{p}}_u(\mathbf{y}), \hat{\mathbf{q}}_u] \approx \frac{1}{2} \sum_{u=1}^m \sqrt{\hat{p}_u(\mathbf{y}_0) \hat{q}_u} + \frac{C_h}{2} \sum_{i=1}^{n_h} \omega_i K_E \left(\frac{\mathbf{y} - \mathbf{x}_i}{h} \right), \quad (\text{B.9})$$

donde,

$$\omega_i = \sum_{u=1}^m \delta[b(\mathbf{x}_i) - u] \sqrt{\frac{\hat{q}_u}{\hat{p}_u(\mathbf{y}_0)}} \quad (\text{B.10})$$

La aproximación realizada es de gran utilidad ya que se puede observar que para minimizar la distancia entre las dos distribuciones, se debe maximizar el segundo término en la ecuación (B.9), ya que el primer término es independiente de \mathbf{y} . Por otra parte, el segundo término representa la estimación de densidad calculada con el núcleo $K_E(\mathbf{y})$ en la imagen actual de la secuencia y con los datos ponderados mediante ω_i . La maximización del coeficiente de Bhattacharyya se puede obtener eficientemente mediante las iteraciones del vector de desplazamiento medio.

Apéndice C

Interfaz computacional

Para realizar el seguimiento visual de una persona, se desarrolló una interfaz gráfica que implementa el método de seguimiento propuesto. Con el fin de mostrar las herramientas que el sistema proporciona al usuario, en este apéndice se describe la interfaz así como su modo de operación.

Durante el seguimiento el sistema opera sobre secuencias de imágenes previamente grabadas en archivos con formato *JPEG* o en imágenes capturadas en tiempo real por una cámara conectada a la computadora. Las imágenes se capturan en formato *RGB* con dimensiones predefinidas de 320×240 píxeles.

La ventana principal del sistema se muestra en la figura C.1, como se observa la interfaz está dividida en tres zonas principales: en la parte superior se localiza la barra de menús, en la cual se encuentran agrupados los comandos del programa cuyas funciones permiten establecer si se va a trabajar con imágenes grabadas o con video, seleccionar los objetos de interés y determinar el tipo de seguimiento que se va a realizar (frontal o con profundidad); en la parte central se encuentra el área de trabajo donde se presentan las imágenes y finalmente en la parte inferior se tiene la barra de estado del sistema.

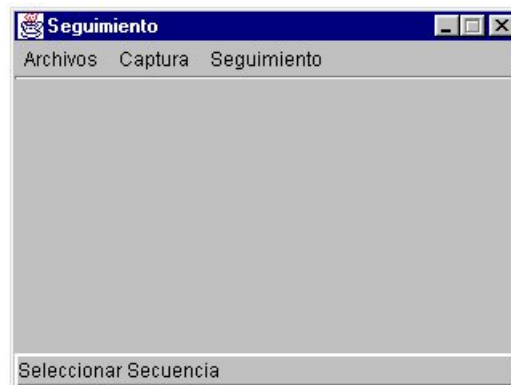


Figura C.1: Ventana principal de la interfaz computacional del sistema de seguimiento

A continuación se describe detalladamente la función de cada uno de los elementos de la barra de menús.

C.1 Archivos

En el menú *Archivos* (figura C.2) se encuentra el comando que permite abrir una secuencia de imágenes grabadas. Para abrir una secuencia se requiere generar un archivo con extensión *.txt* que contenga el nombre de las imágenes separadas por un salto de renglón. El archivo de texto se debe guardar en la misma carpeta en la que se encuentra la secuencia de interés, para no tener que especificar su directorio. Así, el archivo sólo contiene los nombres de las imágenes y su extensión, por ejemplo:

```
frame000.jpg
frame001.jpg
```

....

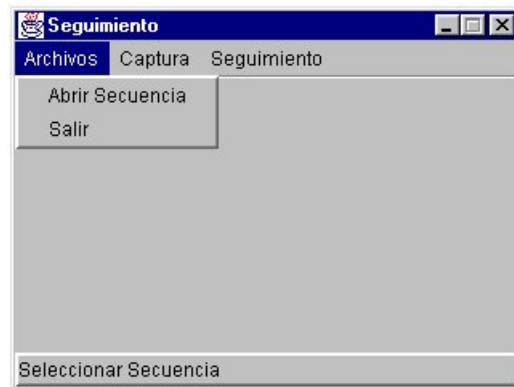


Figura C.2: Opciones del menú *Archivo*

Abrir secuencia: Esta función abre un cuadro de diálogo donde el usuario puede seleccionar el archivo con extensión *.txt* que contiene los nombres de la secuencia de interés.

Salir: Cierra las ventanas del sistema, deshabilita la conexión con la cámara, y sale del programa.

C.2 Captura

En el menú *Captura* cuyas opciones se muestran en la figura C.3, se tienen los comandos que permiten habilitar la captura de video.

Secuencias: Esta opción permite trabajar con secuencias grabadas y está habilitada al inicio del programa.

Figura C.3: Opciones del menú *Captura*

Video: Habilita la captura de imágenes de la cámara y muestra cada una de las imágenes en el área de trabajo de la ventana principal.

Terminar video: Finaliza la comunicación con la cámara y se activa la opción para trabajar con secuencias.

Grabar secuencia: Permite grabar una secuencia de imágenes en formato *JPEG* a partir de la información capturada por la cámara. Las imágenes se guardan con el nombre *frame#imagen.jpg* en el directorio *c:\secuencia*. Para terminar la grabación se deshabilita esta opción.

C.3 Seguimiento

En el menú *Seguimiento* (figura C.4) se tienen los comandos que permiten seleccionar las partes del cuerpo de interés e iniciar su seguimiento. Los comandos del menú de *Seguimiento* son válidos tanto para secuencias de imágenes guardadas en archivos como para imágenes capturadas por la cámara, la única diferencia es que en el caso de imágenes capturadas en tiempo real no se requiere el comando de *Iniciar Seguimiento*, ya que el seguidor comienza en el momento de terminar la selección de cada una de las ventanas de interés.

C.3.1 Seguimiento

Esta opción permite realizar el seguimiento de cualquier objeto u objetos sin considerar las restricciones de estructura del cuerpo humano. En la figura C.5 se muestran las funciones de este menú y un ejemplo de un objeto que se puede seguir con el seguimiento sin estructura (vehículo autónomo).

Seleccionar ventanas: Permite seleccionar los objetos de interés arrastrando el ratón de la computadora para generar la región de interés que se dibuja en la imagen mediante elipses.

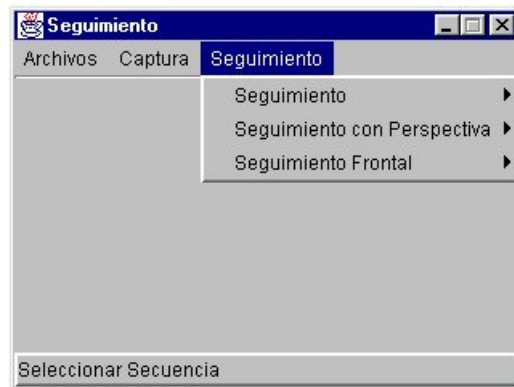


Figura C.4: Opciones del menú *Seguimiento*



Figura C.5: Opciones del menú *Seguimiento* sin consideraciones de estructura

Iniciar seguimiento: Esta opción inicia el seguimiento para el caso de imágenes grabadas en archivos.

C.3.2 Seguimiento con perspectiva

La opción de seguimiento con perspectiva permite realizar el seguimiento de una persona caminando y obtener su posición real a partir de un modelo de perspectiva. La figura C.6 muestra los comandos del menú y un ejemplo de la selección de las ventanas de interés.

Seleccionar ventanas: Permite determinar las regiones de interés por medio del ratón. En el caso de seguimiento con perspectiva se requieren tres ventanas (cabeza, tronco, pies). Al terminar de elegir la primera ventana correspondiente



Figura C.6: Opciones del menú *Seguimiento* con perspectiva

a la cabeza aparece el cuadro de diálogo que se muestra en la figura C.7, en el que se debe elegir la opción correspondiente y posteriormente se debe pulsar *Elegir otra ventana* para continuar con el torso y así hasta finalizar, en cuyo caso se debe seleccionar la opción de *Terminar*.

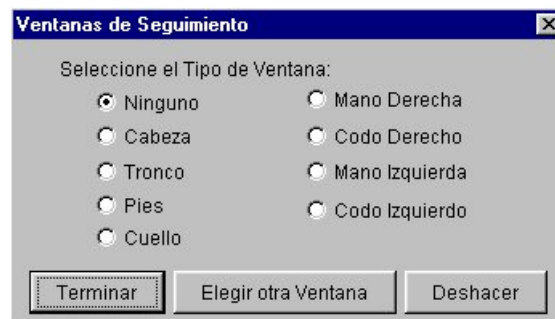


Figura C.7: Ventana de selección de las partes del cuerpo de interés

Iniciar seguimiento: Activa el seguimiento con perspectiva para el caso de secuencias grabadas.

Configuración de perspectiva: Esta función del menú muestra las ventanas correspondientes al ajuste de los parámetros de seguimiento con proyección de perspectiva. Enseguida se describe cada una de las ventanas de configuración:

a) **General:** En la ventana *General* que se muestra en la figura C.8 se activan las opciones para el seguimiento con perspectiva.

Aplicar perspectiva: Calcula la proyección en el ambiente de las posiciones obtenidas por el seguidor en la imagen, mediante el modelo matemático de perspectiva.

Aplicar jerárquica con 3 ventanas: Realiza el seguimiento aplicando las reglas de estructura para el seguimiento de la cabeza, el tronco y los pies.

Aplicar escala: Ajusta el tamaño de las ventanas de seguimiento aplicando el método de adaptación de escala.

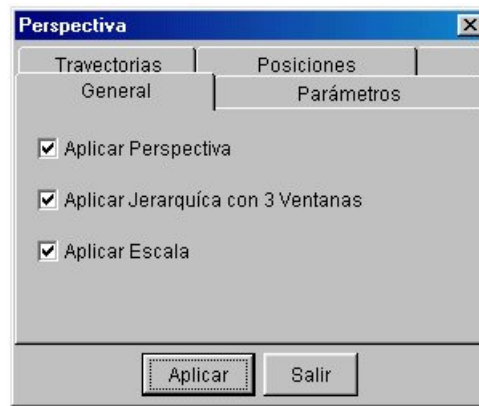


Figura C.8: Ventana de opciones para el seguimiento con profundidad

b) **Parámetros:** En la ventana *Parámetros* que se presenta en la figura C.9 se establecen las medidas del ambiente de trabajo.

Altura de la cámara: Altura a la que se encuentra la cámara a partir del nivel del piso.

Distancia al primer punto: Distancia del punto en el que se encuentra la cámara, proyectado perpendicularmente en el piso, al primer punto que se capta en la imagen.

Distancia al centro: Distancia del punto en el que se encuentra la cámara al punto en el ambiente correspondiente al centro de la imagen.

Ancho de la base: Distancia que existen entre los puntos en el ambiente real correspondientes a las esquinas inferior izquierda e inferior derecha de la imagen.

Altura de la persona: Altura de la persona que se va a seguir mediante el método de seguimiento propuesto.

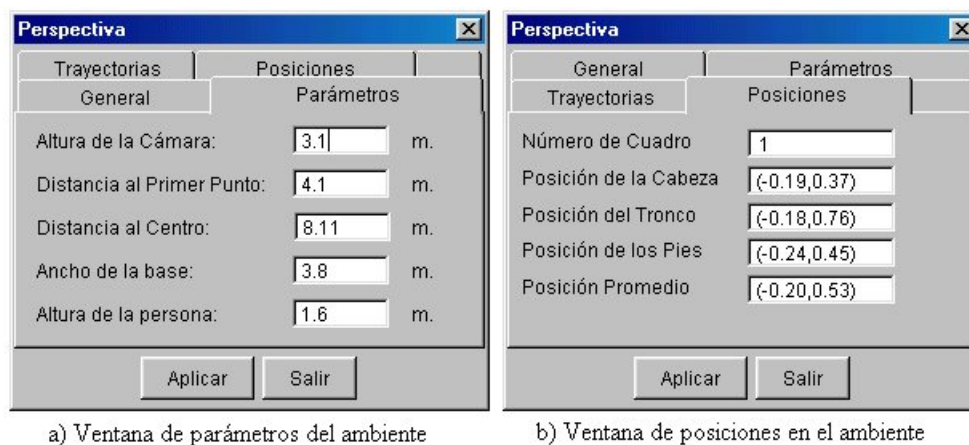
c) **Posiciones:** La ventana *Posiciones* que se presenta en la figura C.9b muestra la posición obtenida mediante la proyección de perspectiva sobre el plano del ambiente correspondiente a cada parte del cuerpo.

Número de cuadro: Indica el número de imagen en la que se obtuvieron las posiciones.

Posición de la cabeza: Determina la posición de la persona en el ambiente con base en la posición devuelta por la ventana de seguimiento correspondiente a la cabeza.

Posición del tronco: Indica la posición real de la persona con base en la información de posición obtenida mediante el seguimiento del torso de la persona.

Posición de los pies: Indica la posición real de la persona considerando los datos de la ventana de seguimiento de los pies.



a) Ventana de parámetros del ambiente

b) Ventana de posiciones en el ambiente

Figura C.9: Ventanas de configuración de parámetros y de posiciones en el ambiente

d) **Trayectorias:** En la ventana *Trayectorias* que se muestra en la figura C.10 se presentan las opciones que permiten ocultar o mostrar las gráficas correspondientes a las trayectorias descritas por cada uno de los planos de las diferentes partes del cuerpo, después de aplicar el modelo de perspectiva.

Graficar trayectoria real: Muestra la trayectoria con los puntos de prueba que se midieron en el ambiente real.

Graficar trayectoria de la cabeza: Muestra la trayectoria generada por el seguidor de la cabeza.

Graficar trayectoria del tronco: Muestra la trayectoria obtenida mediante el seguimiento del torso.

Graficar trayectoria de los pies: Muestra la trayectoria real correspondiente al plano de los pies.

Graficar trayectoria promedio: Muestra gráficamente el promedio de las trayectorias de la cabeza, el torso y los pies.

Guardar trayectorias: Esta opción guarda en un archivo de texto, las posiciones reales de la persona obtenidas a partir del mapeo de perspectiva de los planos de la cabeza, el torso y los pies respectivamente.



Figura C.10: Ventana de selección para mostrar las gráficas de trayectorias reales

C.3.3 Seguimiento frontal

El seguimiento frontal permite realizar el seguimiento de los movimientos de una persona frente a la cámara y su reproducción mediante un modelo virtual de la persona. En la figura C.11 se muestran las opciones del menú y un ejemplo de la selección de las partes del cuerpo de interés.

Seleccionar ventanas: Permite seleccionar las regiones de interés mediante el ratón. Para seleccionar las partes del cuerpo a seguir, es necesario que la persona adopte la posición de inicio que se muestra en la figura C.11 y permanezca estática en esa posición hasta terminar el proceso de selección. En el caso del seguimiento frontal se requieren seis ventanas: cabeza, cuello, codo derecho, mano derecha, codo izquierdo y mano izquierda, las cuales se seleccionan en el orden indicado, en la ventana de la figura C.7.

Iniciar seguimiento: Activa el seguimiento frontal de la persona para secuencias de imágenes grabadas en archivos.

Mostrar modelo: Abre una ventana con el modelo virtual de la persona como se muestra en la figura C.12.

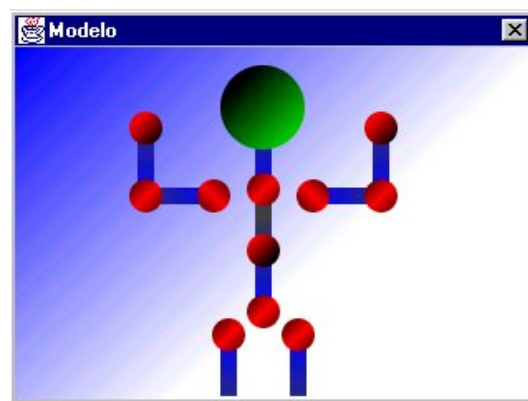
Figura C.11: Opciones del menú *Seguimiento* frontal

Figura C.12: Modelo virtual de la persona

Guardar posiciones del modelo: Genera un archivo de texto que contiene las posiciones en pixeles de cada una de las partes del cuerpo de interés.

Referencias

- [Aguilar 03] C. Aguilar. *Seguimiento de una persona en tiempo real usando métodos derivativos*. Tesis de Maestría, Instituto Tecnológico de Estudios Superiores de Monterrey, Campus Monterrey. Mayo, 2003.
- [Anisfield 00] N. Anisfield. *Ascension technology puts spotlight on DC field magnetic motion tracking*. HP Chronicle, Vol. 17, No. 9, 2000.
- [Bhattacharyya 43] A. Bhattacharyya. *On a measure of divergence between two statistical populations defined by their probability distributions*. Calcutta Math Society, Vol. 35, pp. 99–110, 1943.
- [Chaudhari 01] A. Chaudhari, R. Bragg, J. Alexander y T. P. Andriacchi. *A video-based, markerless motion tracking system for biomechanical analysis in an arbitrary environment*. Summer Bioengineering Conference, Vol. 50, 2001.
- [Cheng 95] Y. Cheng. *Mean shift, mode seeking and clustering*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 17, No. 8, pp. 790–799, 1995.
- [Comaniciu 99] D. Comaniciu y P. Meer. *Mean shift analysis and applications*. IEEE International Conference on Computer Vision, Vol. 2, pp. 1197–1203, 1999.
- [Comaniciu 00] D. Comaniciu, V. Ramesh y P. Meer. *Real-time tracking of non-rigid objects using mean shift*. IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2000), Vol. 2, pp. 142–149, 2000.
- [Fleet 92] D. Fleet. *Measurement of image velocity*. Kluwer Academic Publishers, 1992.
- [Fukunaga 75] K. Fukunaga y L. D. Hostetler. *The estimation of the gradient of a density estimation, with applications in pattern recognition*. IEEE Transactions on Information Theory, Vol. 21, pp. 32–40, 1975.

- [Herda 00] L. Herda, P. Fua, R. Plaenkers, R. Boulic y D. Thalmann. *Skeleton-based motion capture for robust reconstruction of human motion*. Proceedings of Computer Animation, 2000.
- [Isard 98] M. Isard y A. Blake. *Condensation - conditional density propagation for visual tracking*. International Journal of Computer Vision, Vol. 29, No. 1, pp. 5–28, 1998.
- [Lipton 98] A. Lipton, H. Fujiyoshi y R. Patil. *Moving target classification and tracking from real-time video*. IEEE Image Understanding Workshop, pp. 129–136, 1998.
- [Moeslund 01] T. Moeslund y E. Granum. *Survey of computer vision-based human motion capture*. Computer Vision and Image Understanding, Vol. 81, No. 3, pp. 231–268, 2001.
- [Olvera 00] M. Olvera. *Simulación de movimiento de estructuras articuladas basadas en seguimiento visual*. Tesis de Maestría, Instituto Tecnológico de Estudios Superiores de Monterrey, Campus Monterrey. Mayo, 2000.
- [Paradiso 97] J. Paradiso y F. Sparacino. *Optical tracking for music and dance performance*. Fourth Conference of Optical 3D Measurement Techniques, pp. 11–18, 1997.
- [Pers 00] J. Pers y S. Kovacic. *Computer vision system for tracking players in sports games*. First International Workshop on Image and Signal Processing and Analysis, pp. 81–86, 2000.
- [Roberts 02] T. J. Roberts, S. J. McKenna y I. W. Ricketts. *Adaptive learning of statistical appearance models for 3D human tracking*. British Machine Vision Conference, 2002.
- [Scott 92] D. W. Scott. *Multivariate density estimation*. Wiley, 1992.
- [Siebel 01] N. Siebel y S. Maybank. *Real-time tracking of pedestrians and vehicles*. IEEE Workshop on Performance Evaluation of Tracking and Surveillance, 2001.
- [Silverman 86] B. W. Silverman. *Density estimation for statistics and data analysis*. Chapman and Hall, 1986.
- [Song 00] Yang Song, Xiaolin Feng y Pietro Perona. *Towards detection of human motion*. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2000), pp. 810–817, 2000.
- [Stauffer 99] C. Stauffer y W. Grimson. *Adaptive background mixture models for real-time tracking*. IEEE Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. 246–252, 1999.

-
- [Tomasi 91] C. Tomasi y T. Kanade. *Detection and tracking of point features*. Technical Report CMU-CS-91132. Carnegie Mellon University School of Computer Science, 1991.
- [Vázquez 02] J. M. Vázquez. *Control de dirección de un vehículo autónomo*. Tesis de Maestría, Instituto Tecnológico de Estudios Superiores de Monterrey, Campus Monterrey. Mayo, 2002.
- [Wren 97] C. Wren, A. Azarbayejani, T. Darrell y Alex Pentland. *Pfinder: real-time tracking of the human body*. IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 780–785, 1997.

