

# INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE MONTERREY



**TECNOLÓGICO  
DE MONTERREY®**

## **Sistema de Seguimiento Visual para la Rehabilitación**

Presenta:

**Gildardo Azcárate Hernández**

Sometido al Programa de Graduados en Informática y Computación en cumplimiento parcial con los requerimientos para obtener el grado de:

**Maestro en Ciencias Computacionales**

Asesores:

**Dr. Luis Enrique Sucar Succar**

**Dr. Juan Frausto Solís**

Cuernavaca, Morelos. Diciembre de 2007

# **Sistema de Seguimiento Visual para la Rehabilitación**

Presentada por:

**Gildardo Azcárate Hernández**

Aprobada por:

---

**Dr. Luis Enrique Sucar Succar**

Profesor-Investigador del INAOE, Puebla  
Asesor

---

**Dr. Juan Frausto Solís**

Profesor-Investigador del Departamento de Computación  
ITESM Campus Cuernavaca  
Asesor

---

**Dr. Fernando Ramos Quintana**

Director del Programa de Graduados en Informática y Computación  
ITESM Campus Cuernavaca  
Sinodal

---

**Dr. Hugo Gustavo González Hernández**

Director de la Carrera de Ingeniería Mecatrónica  
Escuela de Ingeniería y Arquitectura, ITESM Campus Puebla  
Sinodal

*Dedicado a mi familia,  
y a Nallely.*

# Agradecimientos

Quiero agradecer en primer lugar a mis padres, Gildardo y Cristina, que a través de su gran amor, cariño, comprensión y honradez, han sido el mejor ejemplo de vida que cualquier hijo pudiera desear. Gracias por darme la vida y su amor. Gracias por su inmenso apoyo, que sin el no hubiese realizado esta maestra. Gracias, todos mis logros son suyos.

A mis hermanas, Beatriz, Cristina y Gisela que siempre me apoyaron y alentaron. Gracias por su cariño y comprensión.

A mi asesor Dr. Enrique Sucar por su invaluable orientación en esta tesis, quien siempre depositó su confianza en mí, por su gran apoyo que me brindó para terminar esta maestría.

Al Dr. Juan Frausto por su paciencia y comentarios, a mis sinodales Dr. Fernando Ramos, Dr. Hugo González, gracias por su tiempo y comentarios.

A mis amigos de la maestría, Deny Lizbeth y Juan Paulo, quien siempre conté con ellos en todo momento. Gracias por su apoyo y amistad.

A Nallely, quien siempre estuvo a mi lado aunque estuviera lejos de mí. Gracias por tus palabras que siempre me alentaron y reconfortaron en los días más grises. Gracias por la paciencia que me tuviste. Gracias por permanecer a mi lado y apoyarme en estos dos últimos años, por eso y por mucho más, gracias.

# Resumen

Avances en el campo de visión computacional están generando cambios novedosos y radicales en la forma de interactuar con la computadora. Recientemente el seguimiento visual del movimiento humano ha incrementado la atención por parte de los investigadores. El interés en este campo es motivado por su amplia gama de aplicaciones, incluyendo el análisis del rendimiento atlético y clínico, interfaces humano-computadora, sistemas de vigilancia, captura del movimiento humano para juegos virtuales y animación. Dentro del análisis clínico, ha surgido la inquietud de desarrollar sistemas que detecten el progreso de rehabilitación en pacientes con atrofias cerebrales. Tradicionalmente, las personas que han sobrevivido a un derrame cerebral toman fisioterapia con la ayuda de un fisioterapeuta que diagnostica si se está llevando la rehabilitación apropiadamente. Sin embargo, debido a los altos costos de las terapias y a los periodos prolongados de los ejercicios de rehabilitación, los pacientes no reciben suficiente tratamiento.

En esta tesis, se propone un sistema de seguimiento visual para apoyar el programa de rehabilitación para pacientes que han sufrido un derrame cerebral. El sistema de seguimiento se probó en dos escenarios diferentes: a) Interfaz visual para la terapia utilizando simulación, y b) Análisis del movimiento para evaluar el avance en la rehabilitación. En el primer escenario se realiza el seguimiento tridimensional de la mano de un individuo para interactuar con el software T-WREX, el cual permite interactuar con un ambiente virtual moviendo su brazo dañado, llevando a cabo diferentes tareas diseñadas para imitar situaciones de la vida real orientadas a la rehabilitación. En el segundo escenario, el sistema sigue el movimiento del brazo del paciente utilizando cintas de color adheridas al hombro, codo y muñeca, como marcas distintivas. Las marcas son capturadas por dos cámaras de video, y se estima la posición 3D de cada una de ellas para poder reconstruir el brazo en una estructura de alambre 3D. Se analiza el movimiento del brazo por medio de la extracción de características de movimiento relevantes, como los ángulos entre el brazo y el antebrazo. Para medir el grado de avance del movimiento del brazo en el proceso de rehabilitación, la información de movimiento se parametriza en un modelo probabilista, empleando los *Modelos Ocultos de Markov*, y se compara con el movimiento del brazo de una persona sana, previamente almacenado. La distancia (semejanza) entre ambas trayectorias se puede medir utilizando una modificación de la distancia de *Kullback - Leibler*. El sistema fue probado en pacientes reales, y el análisis del movimiento del brazo fue satisfactorio.

# Índice general

Índice de figuras	x
-------------------	---

Índice de tablas	xv
------------------	----

1. Introducción	1
1.1. Sistemas para la rehabilitación	1
1.2. Descripción del problema	3
1.3. Objetivo	4
1.4. Arquitectura del sistema propuesto	4
1.4.1. Descripción de la arquitectura	5
1.5. Pruebas y Resultados	6
1.6. Organización de la tesis	7
<b>2. Análisis del movimiento humano</b>	<b>8</b>
2.1. Introducción	8
2.2. Sistemas basados en visión con marcas	8
2.2.1. Qualisys	9
2.2.2. VICON	10
2.2.3. CODA	11
2.3. Sistemas basados en visión libres de marcas	11
2.3.1. Seguimiento 2D	12
2.3.2. Seguimiento 3D	15
2.4. Sistemas basados en sensores	17

## ÍNDICE GENERAL

2.5.	Sistemas basados en asistencia robótica . . . . .	18
2.6.	Sistemas para la rehabilitación . . . . .	19
2.7.	Resumen . . . . .	22
<b>3.</b>	<b>Modelos Ocultos de Markov</b>	<b>23</b>
3.1.	Descripción de los Modelos Ocultos de Markov. . . . .	23
3.1.1.	Definición . . . . .	24
3.1.2.	Elementos de un MOM. . . . .	25
3.1.3.	Tres problemas MOM básicos. . . . .	26
3.1.4.	Topologías de los MOM. . . . .	30
3.2.	Métricas para la comparación de trayectorias. . . . .	31
3.2.1.	Métricas basadas en trayectorias. . . . .	33
3.2.2.	Métricas basadas en MOM. . . . .	34
3.3.	Resumen . . . . .	36
<b>4.</b>	<b>Sistema de seguimiento visual</b>	<b>38</b>
4.1.	Introducción . . . . .	38
4.2.	Seguimiento visual de la mano: T-WREX. . . . .	39
4.2.1.	Calibración de las cámaras. . . . .	40
4.2.2.	Segmentación. . . . .	42
4.2.3.	Seguimiento 2D. . . . .	48
4.2.4.	Reconstrucción 3D. . . . .	50
4.3.	Seguimiento visual de las marcas: Análisis de gestos. . . . .	52
4.3.1.	Determinación de los gestos. . . . .	52
4.3.2.	Calibración de las cámaras. . . . .	53
4.3.3.	Segmentación. . . . .	53
4.3.4.	Seguimiento 2D. . . . .	55
4.3.5.	Reconstrucción 3D. . . . .	56
4.4.	Resumen . . . . .	56

## ÍNDICE GENERAL

<b>5. Análisis de gestos</b>	<b>58</b>
5.1. Introducción	58
5.2. Extracción de características de movimiento	59
5.2.1. Discretización	61
5.3. Análisis del movimiento	63
5.3.1. Topología de los MOM empleada	63
5.3.2. Entrenamiento de los MOM	64
5.3.3. Comparación de gestos empleando MOM	64
5.4. Resumen	64
<b>6. Resultados experimentales</b>	<b>66</b>
6.1. Seguimiento visual de la mano: T-WREX	66
6.1.1. Ambiente de trabajo	66
6.1.2. Calibración de las cámaras	67
6.1.3. Segmentación	68
6.1.4. Seguimiento 2D	69
6.1.5. Reconstrucción 3D	71
6.1.6. Implementación	72
6.2. Seguimiento visual de las marcas: Análisis de gestos	73
6.2.1. Ambiente de trabajo	73
6.2.2. Calibración de las cámaras	74
6.2.3. Segmentación	74
6.2.4. Seguimiento 2D	75
6.2.5. Reconstrucción 3D	75
6.2.6. Análisis de gestos	77
6.2.7. Entrenamiento de los Modelos Ocultos de Markov	77
6.2.8. Análisis del movimiento	77
6.3. Resumen	79



## ÍNDICE GENERAL

<b>7. Conclusiones</b>	81
7.1. Conclusiones . . . . .	81
7.2. Trabajo futuro. . . . .	82
<b>Referencias</b>	87

# Índice de figuras

1.1. Diagrama de bloques del sistema de seguimiento visual, a) Sistema de seguimiento de la mano, que interactúa con la herramienta T-WREX y, b) Análisis de gestos, que analiza el movimiento del brazo de un paciente. . . . .	6
2.1. Imagen de una persona caminando usando MLD en sus articulaciones. . . . .	9
2.2. Sistema <i>Qualisys</i> en operación. . . . .	10
2.3. Sistema <i>VICON</i> en funcionamiento. . . . .	10
2.4. Sistema <i>CODA</i> . . . . .	11
2.5. Demostración del sistema desarrollado por Wren. a) Imagen original, b) Segmentación, y c) Representación 2D de los <i>blobs</i> . . . . .	12
2.6. Secuencias de imágenes utilizadas por Baumberg y Hogg, que muestran los contornos activos para el seguimiento de un peatón. . . . .	12
2.7. Sistema de seguimiento propuesto por Schlegel. a) Localización de la persona, b) Segmentación de la camisa. . . . .	13
2.8. Imagen que muestra el seguimiento del rostro humano en el sistema propuesto por Bradski. La ventana de seguimiento localiza y sigue el rostro detectando el área de piel en la imagen. . . . .	13
2.9. Imagen que muestra la localización y seguimiento del rostro y mano de un individuo en el sistema propuesto por Avilés. Imagen izquierda: Localización del rostro, hombro y mano. Imagen derecha: Secuencia de imágenes que muestran el seguimiento de la mano. . . . .	14
2.10. Imagen que muestra la identificación de objetos de interés y su cercanía con la mano del usuario. . . . .	14
2.11. Figura de alambre del cuerpo humano. . . . .	15
2.12. Sistema de movimiento humano, propuesto por Black. . . . .	15
2.13. Imagen que muestra el sistema propuesto por Howe. Imagen superior: seguimiento en el espacio 2D empleando una sola cámara. Imagen inferior: Estimación de las articulaciones del cuerpo humano en el espacio 3D. . . . .	16

## ÍNDICE DE FIGURAS

2.14. Secuencias de imágenes mostrando la representación del cuerpo humano empleando un modelo de alambre en el método propuesto por Cheung . . . . .	17
2.15. Imagen del dispositivo MT9 . . . . .	17
2.16. Dispositivo CyberGlove . . . . .	18
2.17. Ilustración del ARM-Guide. Imagen izquierda: Un paciente interactuando con el dispositivo robótico, Imagen derecha: Prototipo del ARM-Guide . . . . .	19
2.18. La imagen muestra el dispositivo MT9 fijado al brazo de un individuo, monitoreando el ángulo formado entre el brazo y el antebrazo . . . . .	19
2.19. La imagen ilustra el sistema propuesto por Tao. Las marcas de color son fijadas en las articulaciones del brazo de un individuo para el seguimiento del mismo . . . . .	20
2.20. Segmentación de la mano en el sistema propuesto por Tao y Zhou. a) Imagen original, b) Imagen segmentada al aplicarse el algoritmo Camshift . . . . .	20
2.21. Vista del ambiente virtual del sistema propuesto por McNeil . . . . .	21
3.1. Cadena de Markov mostrando las probabilidades de transición y los símbolos de salida, en el ejemplo de lanzar una moneda. La secuencia de estados puede determinarse a partir de la secuencia de salida . . . . .	24
3.2. Modelo Oculto de Markov mostrando las probabilidades de transición y de salida, en el ejemplo de lanzar dos monedas. La secuencia de estados no puede determinarse a partir de la secuencia de salida . . . . .	25
3.3. Ilustración de los 3 tipos de topologías de los MOM. a) Modelo <i>ergódico</i> de 4 estados, b) Modelo <i>izquierda-derecha</i> de 4 estados, c) Modelo <i>izquierda derecha paralelo</i> de 6 estados . . . . .	31
3.4. Ejemplos de casos ambiguos de las métricas convencionales: a) Diferente duración, b) Igual duración pero diferente velocidad de captura . . . . .	32
3.5. Imagen que muestra la <i>distancia de cruce</i> entre dos MOM's. Las trayectorias $T_1$ y $T_2$ son parametrizadas en los modelos $\lambda_1$ y $\lambda_2$ , respectivamente . . . . .	36
4.1. Diagrama de bloques del sistema de seguimiento visual. Los recuadros de color muestran los módulos en que se conforma el seguimiento: Calibración, Segmentación, Seguimiento 2D y, Reconstrucción 3D . . . . .	39
4.2. Imágenes capturadas del software T-WREX. (a) Pantalla de inicio, (b) - (c) Videojuegos virtuales en un ambiente tridimensional . . . . .	40
4.3. La imagen muestra la vista lateral del sistema de seguimiento visual de la mano, así como el sistema de referencia del mundo . . . . .	41
4.4. La imagen muestra el modelo de proyección <i>punta de alfiler</i> . . . . .	42

## ÍNDICE DE FIGURAS

4.5.	La imagen muestra el sistema de referencia de la cámara con respecto al sistema de referencia del mundo. La cámara se encuentra Rotada (R) y Traslada (T) a un nuevo punto en el espacio 3D. . . . .	43
4.6.	Imagen que muestra el proceso de segmentación para la mano de una persona empleando color y movimiento. . . . .	44
4.7.	La imagen ilustra el proceso de segmentación de píxeles de piel. a) Imagen original, b) Imagen segmentada . . . . .	45
4.8.	La imagen muestra el proceso de segmentación basado en movimiento. . . . .	46
4.9.	Ejemplo de segmentación basado en fusión de color y movimiento. a) Imagen original, b) Segmentación de piel, c) Detección de movimiento y c) Fusión (AND) de color y movimiento. . . . .	47
4.10.	Localización inicial de la mano. a) Posición inicial, b) Ventana inicial de seguimiento . . . . .	47
4.11.	Diagrama de flujo del algoritmo Camshift utilizado para el seguimiento de la mano. . . . .	49
4.12.	Secuencia de tres imágenes que muestran el seguimiento de la mano aplicando el algoritmo Camshift. . . . .	49
4.13.	Cruce de las líneas de visión en el espacio tridimensional. . . . .	50
4.14.	La imagen muestra la vista superior del sistema de seguimiento y el sistema de referencia del mundo. Las marcas de color identifican las articulaciones. Amarillo: hombro, rojo: codo y azul: muñeca . . . . .	52
4.15.	Gesto de <i>flexión-extensión</i> . . . . .	53
4.16.	Gesto <i>circular</i> . . . . .	53
4.17.	La imagen ilustra el proceso de segmentación. a) Imagen original, b) Histograma 2D en el espacio de color <i>H-S</i> que muestra la distribución del color de la marca roja (codo), y c) Imagen resultante al aplicar el método de <i>Retroproyección de Histograma</i> . . . . .	54
4.18.	La imagen ilustra la posición inicial de la marca de color. a) Centro de gravedad del objeto, y b) Ventana de seguimiento inicial . . . . .	54
4.19.	Secuencia de imágenes (a - f) que ilustra el proceso de seguimiento de las marcas de color, al realizar el gesto <i>flexión - extensión</i> . . . . .	55
4.20.	La imagen ilustra la representación del brazo en el espacio tridimensional. Se trazan líneas de unión entre las articulaciones. . . . .	56
5.1.	Diagrama de bloques del sistema desarrollado. M1) Módulo de seguimiento visual, M2) Módulo de extracción de características, y M3) Módulo de análisis de movimiento . . . . .	59

## ÍNDICE DE FIGURAS

5.2.	La imagen ilustra la postura del brazo en un instante de tiempo, $\alpha$ indica el ángulo entre el brazo y el antebrazo, los ángulos $\theta$ y $\Phi$ representan las coordenadas esféricas del codo tomando como referencia al hombro. H, M y C representan las coordenadas tridimensionales (X, Y, Z) del hombro, codo y muñeca, respectivamente . . . . .	60
5.3.	Imagen que ilustra la trayectoria del gesto <i>circular</i> , a) Variación de los ángulos con respecto al tiempo y, b) Trayectoria del gesto en el espacio de características $[\alpha - \theta - \Phi]$ .	62
5.4.	La imagen muestra el proceso de agrupamiento al aplicar el algoritmo K-medias sobre el conjunto de datos de las trayectorias de los gestos, a) Conjunto total de datos de trayectorias y, b) Agrupamiento en 32 clases . . . . .	63
5.5.	Topología empleada en los MOM . . . . .	64
6.1.	Imagen que muestra la configuración de las cámaras 1 y 2 en el ambiente de trabajo para el seguimiento de la mano: T-WREX . . . . .	67
6.2.	Patrón de calibrado para obtener los parámetros intrínsecos de cada cámara (cámara 1 y 2) . . . . .	68
6.3.	Distribución de los pixeles en los modelos de histograma de piel y no-piel, en el espacio de color H-S . . . . .	68
6.4.	Imagen que muestra el clasificador bayesiano para la segmentación de regiones de piel. La imagen de la izquierda muestra la imagen original y, la imagen de la derecha muestra los pixeles en escala de grises, donde el valor de gris (0 - 255) indica la pertenencia a la clase piel . . . . .	69
6.5.	La imagen muestra el seguimiento de la mano en dos secuencias de imágenes capturadas por dos cámaras de video empleando el algoritmo Camshift . . . . .	70
6.6.	Ejemplos del sistema de seguimiento visual de la mano al interactuar con la herramienta T-WREX. La ventana de seguimiento visual se despliega en la parte superior izquierda de la pantalla . . . . .	72
6.7.	Ambiente de trabajo para el sistema de seguimiento de las marcas: análisis de gestos.	73
6.8.	Ejemplo del algoritmo de segmentación <i>Retroproyección de Histogramas</i> . a) Imagen original, y b)-d), segmentación de las marcas de color amarillo, rojo y azul, respectivamente . . . . .	74
6.9.	Secuencia de imágenes que muestran la ejecución del algoritmo Camshift en cada una de las marcas de color . . . . .	75
6.10.	Secuencia de imágenes que muestran el seguimiento del brazo en el espacio tridimensional. Las imágenes de las dos primeras columnas muestran el seguimiento de las marcas de color en el espacio 2D, capturadas por las dos cámaras de video. Las imágenes de la tercera columna muestran la reconstrucción del brazo en el espacio tridimensional . . . . .	76

## ÍNDICE DE FIGURAS

- 6.11. La imagen muestra las trayectorias del movimiento del brazo al ejecutar el gesto circular. La imagen de la izquierda muestra la trayectoria del brazo de una persona sana. La imagen de la derecha muestra la trayectoria del brazo del paciente. . . . . 77
- 6.12. La imagen ilustra la distancia de Kullback-Leibler para los gestos *deflexión - extensión* y *circular*, respectivamente, ejecutados por un paciente. La distancia converge en medida que la longitud de la secuencia de observaciones incrementa . . . . . 78
- 6.13. La imagen muestra las gráficas comparativas de la distancia de los gestos ejecutados por una persona sana y por un paciente. La gráfica de la izquierda muestra las distancias obtenidas al ejecutar el gesto *flexión-extensión* por una persona sana (gráfica de color rojo) y por un paciente (gráfica de color negro). La gráfica de la derecha muestra las distancias obtenidas al ejecutar el gesto *circular* por una persona sana (gráfica de color rojo) y por un paciente (gráfica de color negro). . . . . 79

# Índice de tablas

6.1. Resultados de la etapa de reconstrucción 3D (en centímetros). Las columnas $X_e$ , $Y_e$ , $Z_e$ muestran las posiciones de la mano estimadas por el sistema, las columnas $X_r$ , $Y_r$ , $Z_r$ muestran las posiciones reales medidas manualmente, y la columna <i>Error</i> muestra la diferencia entre la posición estimada y la posición real. . . . .	71
--	----

# Capítulo 1

## Introducción

En los últimos años, el seguimiento visual del movimiento humano ha incrementado la atención por parte de los investigadores. El interés en este campo es motivado por su amplia gama de aplicaciones, incluyendo el análisis del rendimiento atlético y clínico, interfaces humano-computadora, sistemas de vigilancia, captura del movimiento humano para juegos virtuales y animación. Dentro del dominio del análisis clínico, ha surgido la inquietud por desarrollar sistemas que ayuden en el proceso de rehabilitación en pacientes que han sufrido de un derrame cerebral. Los sistemas de seguimiento visual usados en esta área permitirán seguir y analizar el movimiento de las extremidades del cuerpo humano apropiadamente, además de permitir al paciente realizar ejercicios que los motiven en el proceso de rehabilitación. En esta tesis, se propone un sistema de seguimiento basado en visión computacional para el seguimiento de la mano y del brazo de un paciente, que le permita realizar ejercicios de movimiento del brazo proporcionando información de su progreso de rehabilitación.

### 1.1. Sistemas para la rehabilitación

Cada año en los Estados Unidos alrededor de 600,000 personas sobreviven a un derrame cerebral, mientras que en México se estima una cifra alrededor de 200,000 [2]. Aproximadamente el 80 % de las personas que sobreviven a un derrame pierden la habilidad en el movimiento del brazo y de la mano. Tradicionalmente, las personas que han sobrevivido a un derrame cerebral toman fisioterapia con la ayuda de un fisioterapeuta que diagnostica si se está llevando a cabo apropiadamente la rehabilitación. Sin embargo, debido a los altos costos de la terapia y a los periodos prolongados en los ejercicios de rehabilitación, los pacientes no reciben el suficiente tratamiento y son enviados a casa muy tempranamente. La rehabilitación resultante es, a menudo, autodirigida con información poco profesional. Por lo tanto, un objetivo importante para la ingeniería de la rehabilitación, es desarrollar una tecnología que permita a los individuos con derrame practicar intensivos movimientos de entrenamiento, sin estar en la presencia permanente de un fisioterapeuta.

Avances en el seguimiento visual del movimiento humano han permitido desarrollar sistemas enfocados en la rehabilitación [60,61,65]. Los sistemas de seguimiento visual existentes se pueden clasificar en dos categorías principales:

- **Sistemas de seguimiento visual basados en marcas.** Esta es una técnica que usa sensores ópticos (eg. cámaras), para el seguimiento del cuerpo humano, cuya imagen es capturada colocando



## 1.1. Sistemas para la rehabilitación

identificadores (*landmarks*) en las articulaciones del cuerpo. Este tipo de sistemas ha atraído la atención de investigadores en la ciencia médica, los deportes y la ingeniería. Su principal desventaja es el uso de marcas que impiden la movilidad de algunas partes del cuerpo humano, así como su alto costo comercial y su manejo especializado.

- **Sistemas de seguimiento visual libres de marcas.** Esta es una técnica que utiliza cámaras de video convencionales para la captura del movimiento humano. Estos sistemas son capaces de superar algunos problemas de oclusión y solo se concentran en los contornos, bordes o características relevantes del cuerpo humano. Sin embargo, requieren de un alto costo computacional.

Los métodos estándares para el análisis clínico del movimiento humano son los métodos de seguimiento basados en marcas. Existen varios sistemas comerciales de captura del movimiento humano basados en marcas que pueden ser empleados para el seguimiento del movimiento de los pacientes, como son Qualisys y CODA [51, 56]. Sin embargo, además de las dificultades del calibrado tanto de las cámaras como de las marcas, estos sistemas son demasiado costosos para ser usados por los pacientes desde su casa, y demasiado complicados para que el fisioterapeuta interprete los resultados de seguimiento del paciente.

Recientemente, se han realizado muchos intentos por diseñar sistemas de seguimiento libres de marcas para la captura del movimiento humano [1,19,64]. Los sistemas de seguimiento libres de marcas son más atractivos debido a que solo se necesitan cámaras de video convencionales en vez de cámaras y marcas especiales. Sin embargo, diseñar un sistema para el seguimiento del movimiento humano no es una tarea trivial. Existen varias dificultades [49,50], incluyendo ambigüedades en la profundidad, deformidades en la apariencia, complejidad en los modelos cinemáticos y oclusiones. Para simplificar los problemas del seguimiento del movimiento humano, la mayoría de los algoritmos de seguimiento emplean *modelos* tridimensionales de la forma de la persona, otros emplean múltiples cámaras para mejorar la robustez. Los modelos de la forma del sujeto varían desde un simple modelo de estructura de alambre [8], hasta modelos volumétricos sofisticados [9,15,26].

El programa de rehabilitación requiere de un sistema de seguimiento visual que sea de bajo costo, con cierta precisión y que pueda ejecutarse en tiempo real. Tanto los métodos de seguimiento visual basados en marcas y los métodos libres de marcas sólo pueden cumplir partes de estos requerimientos. En otras palabras, los sistemas de seguimiento basados en marcas pueden proporcionar la precisión de seguimiento requerida, pero son demasiado costosos. Por otro lado, los sistemas de seguimiento libres de marcas son relativamente baratos, pero su robustez y precisión necesitan ser mejoradas. Por lo tanto, es necesario crear sistemas de seguimiento visual que tomen las ventajas de ambos métodos de seguimiento.

Tao y otros [60], proponen un algoritmo de seguimiento basado en color para capturar el movimiento de partes del cuerpo humano enfocado en el proceso de rehabilitación en casa. Diferentes cintas de color son colocadas en las articulaciones de interés del cuerpo y seguidas en una secuencia de video. El rendimiento de su sistema es comparado con el sistema de seguimiento comercial Qualisys [56]. De manera similar Zhou y otros [61], proponen un sistema de seguimiento visual en tiempo real para la captura del movimiento del brazo de un individuo. Su sistema integra técnicas de visión computacional y sensores de inercia para seguir el brazo en un espacio 3D. Consideran que el uso de dos fuentes de información reduce el problema de oclusión presente en los sistemas que sólo utilizan técnicas de visión.

Reinkensmeyer y otros [44-46], desarrollaron el sistema T-WREX enfocado al proceso de rehabilitación después de un derrame cerebral. El sistema consiste en un dispositivo robótico que es

## 1.2. Descripción del problema

fijado al brazo del paciente para poder medir su movimiento en un espacio tridimensional. La información de movimiento del brazo es enviado al software T-WREX para interactuar con un ambiente virtual, llevando a cabo diferentes tareas diseñadas para imitar situaciones de la vida real orientadas a la rehabilitación. Sin embargo, la manufactura del dispositivo robótico tiene un alto costo, limitando así su accesibilidad. Además, el uso de un dispositivo fijado al brazo del paciente impide ejecutar movimientos de forma natural.

Es evidente de acuerdo a los trabajos de investigación recientes, el interés que existe en desarrollar sistemas computacionales que ayuden en el proceso de rehabilitación en pacientes que han sufrido un derrame cerebral. El uso de sistemas de seguimiento visual ofrece una alternativa para el proceso de rehabilitación, ya que son relativamente baratos y no se requieren mecanismos especiales para su uso. Por lo tanto, consideramos que es una gran motivación participar en el desarrollo de estos sistemas.

## 1.2. Descripción del problema

La rehabilitación es el proceso por medio del cual se ayuda al paciente a conseguir los máximos niveles posibles de independencia y calidad de vida, tanto en el aspecto físico como en los aspectos emocional y social. Para llevar a cabo el proceso de rehabilitación se requiere de un equipo de médicos fisioterapeutas que guíen al paciente a través de una serie de ejercicios para recuperar la movilidad en sus extremidades afectadas, como pueden ser los brazos. Sin embargo, el costo de las fisioterapias son elevadas y los pacientes no reciben el suficiente tratamiento.

Lo anterior nos hace pensar en la necesidad de desarrollar sistemas computacionales de bajo costo, que le permitan a un paciente realizar ejercicios de movimiento de brazo para apoyarlo en el proceso de rehabilitación, sin estar en la presencia permanente de un fisioterapeuta. Algunos investigadores han propuesto sistemas para apoyar en el proceso de rehabilitación, empleando para ello sensores de movimiento y dispositivos robóticos, que son fijados a las extremidades del cuerpo humano para monitorear su movimiento. Sin embargo, el uso de estos dispositivos limita la movilidad del paciente en forma natural, y el costo para implementar el sistema es elevado. Por otro lado, existen en el mercado sistemas para la captura del movimiento humano que son usados para el análisis clínico, pero su costo es elevado y requieren de un equipo especial, además de un manejo de operación complicado.

Emplear técnicas de seguimiento visual libre de marcas es una buena opción para el seguimiento de las extremidades de un individuo enfocado a la rehabilitación, debido a que son de bajo costo y no requieren de un equipo especial. En nuestro caso, se requiere de una metodología que permita el seguimiento del brazo de un paciente para apoyar en el proceso de rehabilitación.

Por lo tanto, el problema que se propone abordar en la presente tesis es la localización y el seguimiento del brazo derecho de un individuo a través de una secuencia de imágenes capturadas por dos cámaras de video, extraer información de la trayectoria de movimiento de su brazo para comparar dicha trayectoria con aquella realizada por una persona sana.

La localización y seguimiento del brazo de un individuo involucra diversas áreas de estudio. La elección de los movimientos del brazo a analizar deberán ser los utilizados en las sesiones fisioterapéuticas a favor de una mayor utilidad en el proceso de rehabilitación. La localización del brazo de una persona dentro de una imagen, así como su seguimiento y la extracción de características des-

### 1.3. Objetivo

criptivas de movimiento a través de una secuencia de imágenes, son problemas pertenecientes al área de visión computacional. Además, asociar ésta información con un conjunto de datos previamente almacenados, es un problema que corresponde al reconocimiento de patrones [3]. De esta forma, para lograr el análisis del movimiento del brazo de una persona en su proceso de rehabilitación, puede dividirse en las siguientes etapas principales:

- Localización del brazo en una imagen. El primer paso en los sistemas de seguimiento visual es la localización automática de las partes del cuerpo humano. Esto es un problema aun no resuelto dentro del campo de visión computacional, debido a que no se conoce con certeza qué características de los objetos son suficientes para diferenciarlas del resto de la imagen.
- Seguimiento del brazo y extracción de características de movimiento. Una vez localizadas las posiciones iniciales de las partes del cuerpo humano, el siguiente reto es el seguimiento a través de una secuencia de imágenes. Los métodos tradicionales suponen que el movimiento humano es un movimiento suave, con velocidad y aceleración constantes. Sin embargo, el movimiento humano varía de forma complicada. Además, extraer información de movimiento es otro problema dentro del análisis del movimiento. Para construir representaciones que capturen de manera confiable el movimiento de partes del cuerpo humano, las características de imágenes de personas deben ser cuidadosamente seleccionadas y extraídas.
- Comparación de trayectorias (evaluación). Finalmente, en la fase de seguimiento de un objeto es posible obtener trayectorias de movimiento espacio-temporales para su análisis. El tema principal en la evaluación de los resultados del seguimiento es una métrica de la distancia que determine la similitud entre dos trayectorias. Comparar la información de movimiento con otra previamente almacenada no es una tarea trivial. Existen varios problemas para comparar trayectorias de movimiento. Por ejemplo, las trayectorias necesitan tener la misma duración de muestreo, o bien, las trayectorias fueron tomadas en distintos instantes de tiempo.

### 1.3. Objetivo

El objetivo de la tesis es desarrollar un sistema de seguimiento basado en visión computacional que siga el movimiento del brazo de una persona dentro de una secuencia de imágenes capturadas por dos cámaras de video, mida las propiedades del movimiento de su brazo y evalúe el grado de avance en el proceso de rehabilitación.

### 1.4. Arquitectura del sistema propuesto

En esta tesis se propone el desarrollo de un sistema de seguimiento visual enfocado a la rehabilitación después de un derrame cerebral. El sistema desarrollado permite sustituir al dispositivo robótico del sistema T-WREX, propuesto por Reinkensmeyer y otros [44-46]. La localización y seguimiento visual de la mano en un espacio tridimensional le permite a un paciente interactuar con los videojuegos del software T-WREX, simulando tareas orientadas a la rehabilitación. Por otro lado, el sistema de seguimiento propuesto se aplica también para el análisis de gestos<sup>1</sup>, en el cual se

<sup>1</sup>El término gesto se puede interpretar como un acto visual que envía una señal a un espectador. En esta tesis, el gesto es considerado como los movimientos del brazo de un paciente en el proceso de rehabilitación

## 1.4. Arquitectura del sistema propuesto

analiza el movimiento del brazo de un paciente para determinar su grado de avance en el progreso de rehabilitación. El sistema detecta y sigue el brazo de un individuo en un espacio tridimensional. El análisis de gestos se realiza al extraer información de movimiento del brazo al realizar un gesto, dicha información se compara con aquella realizada por una persona sana. Para realizar lo anterior, se aplica en este trabajo un enfoque modular tal como se muestra en la arquitectura de la Figura 1.1. Una breve explicación de cada módulo se ofrece a continuación.

### 1.4.1. Descripción de la arquitectura

El sistema de seguimiento visual se aplica en dos escenarios diferentes<sup>2</sup> y se muestran encerrados por líneas punteadas en el diagrama de bloques de la Figura 1.1: **(a) Interfaz visual para la terapia utilizando simulación.** El cual tiene como objetivo que un paciente realice una serie de ejercicios de brazo al interactuar con videojuegos de la herramienta *T-WREX* [44-46], donde el sistema detecta y sigue la mano en un espacio tridimensional y, **(b) Análisis de gestos,** el cual tiene como objetivo analizar el movimiento del brazo de un paciente por medio de comparar gestos previamente definidos. El sistema detecta y sigue el brazo en un espacio tridimensional, utilizando cintas de color adheridas a las articulaciones. Los movimientos son comparados con gestos correctos de personas sanas previamente almacenados. La descripción de cada componente es la siguiente.

#### Escenario (a): Interfaz visual para la terapia utilizando simulación

- **Seguimiento de mano.** Este módulo se encarga de procesar la información visual de las secuencias de imágenes capturadas por dos cámaras de video. El sistema detecta y sigue la mano utilizando información de color y movimiento. Para la detección de color se emplea un algoritmo de segmentación de piel utilizando histogramas de color y un proceso de clasificación utilizando la regla de Bayes. Para la detección de movimiento se emplea un algoritmo de sustracción de imágenes consecutivas. Para determinar las coordenadas en el espacio 3D de la mano se emplea un algoritmo de intersección de líneas de visión de ambas cámaras.
- **T-WREX.** Este módulo se encarga de interactuar con la herramienta *T-WREX* [44-46]. Las coordenadas de la mano ( $X$ - $Y$ - $Z$ ) obtenidas en el módulo anterior son enviadas al software *T-WREX* para que el paciente interactúe con un ambiente virtual moviendo su brazo dañado, llevando a cabo diferentes tareas diseñadas para simular situaciones de la vida real orientadas a la rehabilitación.

#### Escenario (b): Análisis de gestos

- **Seguimiento de brazo.** Este módulo se encarga de procesar las secuencias de imágenes de dos cámaras de video. El sistema detecta y sigue cintas de color adheridas a las articulaciones del brazo utilizando información de color. Para la detección de las cintas se emplean histogramas de color. Cada cinta es de un color específico, representando una articulación del brazo (codo, hombro y muñeca). Para determinar las coordenadas en el espacio 3D de cada cinta de color se emplea de igual forma un algoritmo de intersección de líneas de visión (de ambas cámaras a las marcas). El brazo es representado en una estructura de alambre en el espacio 3D.

Lo cual da como resultado dos sistemas de cómputo diferentes

## 1.5. Pruebas y Resultados

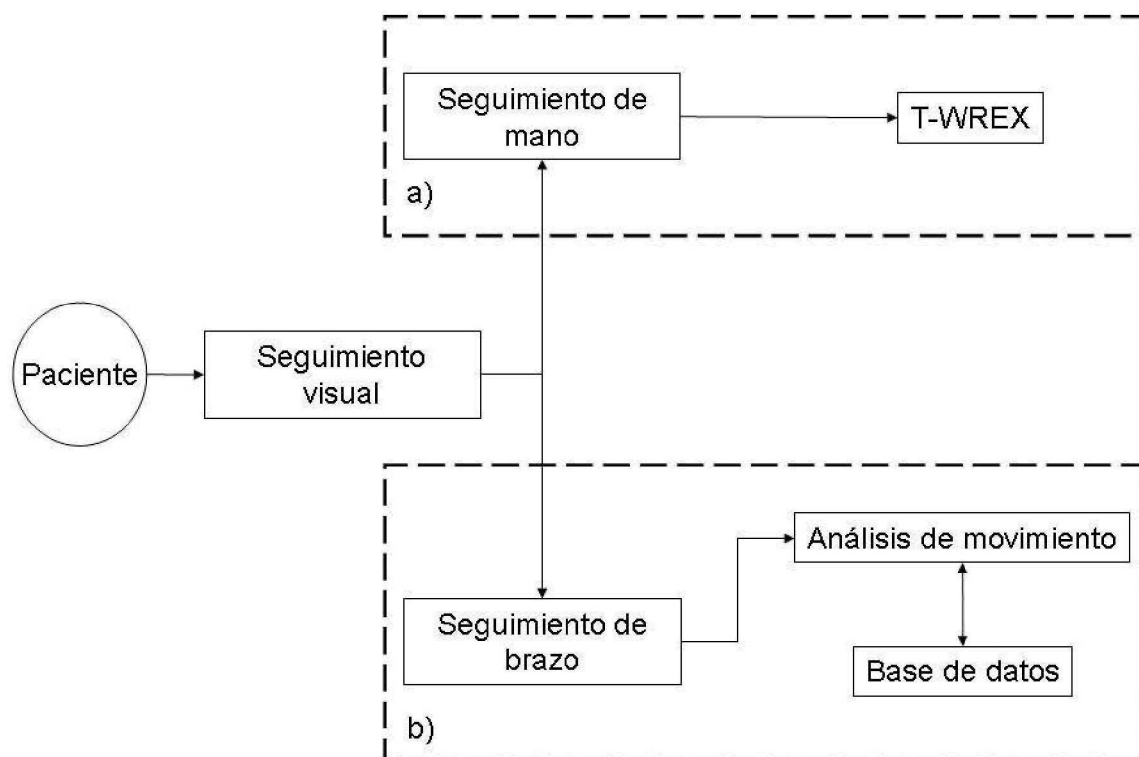


Figura 1.1: Diagrama de bloques del sistema de seguimiento visual, a) Sistema de seguimiento de la mano, que interactúa con la herramienta T-WREX y, b) Análisis de gestos, que analiza el movimiento del brazo de un paciente

- **Análisis de movimiento.** Este módulo se encarga de analizar los gestos del brazo. Se extrae información de las trayectorias del brazo al ejecutar un gesto. Los gestos del paciente son comparados con gestos almacenados en una base de datos realizados por una persona en su forma correcta. La comparación de ambas trayectorias se realiza usando un enfoque estadístico empleando *Modelos Ocultos de Markov*.

Las metodologías empleadas para el desarrollo de los dos sistemas ( a) Interfaz visual para la terapia utilizando simulación, y b) Análisis de gestos) se presentan en los capítulos 4 y 5, y son la propuesta de esta tesis.

## 1.5. Pruebas y Resultados

Las pruebas y resultados obtenidos en este trabajo se presentan en dos partes:

**Interfaz visual para la terapia utilizando simulación.** En el primer escenario, el sistema de seguimiento visual desarrollado fue implementado en el sistema T-WREX [44-46], en colaboración con la Universidad de Irvine California a cargo del Dr. David Reinkensmeyer. El objetivo del sistema de seguimiento visual fue reemplazar el dispositivo robótico para monitorear el

## 1.6. Organización de la tesis

movimiento de la mano de un paciente. El sistema localiza y sigue la mano de un paciente, enviando las coordenadas  $XYZ$  al software T-WREX, para manipular los videojuegos simulando tareas orientadas a la rehabilitación. El sistema fue probado con pacientes del Instituto Nacional de Neurología y Neurocirugía (INNN) [27], obteniéndose resultados satisfactorios.

- **Análisis de gestos.** En el segundo escenario, el sistema de seguimiento visual localiza y sigue el brazo de un individuo al realizar un gesto. Los gestos considerados fueron los movimientos del brazo utilizados en las sesiones de fisioterapia: *flexión-extensión* y movimiento *circular*. Para el entrenamiento de los Modelos Ocultos de Markov se empleó un promedio de 10 ejecuciones por gesto, los cuales se tomaron de dos personas sanas que ejecutaron los gestos en forma correcta. Para el entrenamiento de un paciente en proceso de rehabilitación, se realizaron 5 ejecuciones por cada gesto considerado. Los resultados experimentales muestran el comportamiento de la métrica basada en los Modelos Ocultos de Markov para comparar trayectorias de movimiento, dando resultados satisfactorios.

## 1.6. Organización de la tesis

El resto del documento está organizado de la siguiente manera. En el capítulo 2 se presentan los trabajos relacionados en el análisis del movimiento humano en general, y los sistemas para la rehabilitación después de un derrame cerebral en particular. En el capítulo 3 se describen los Modelos Ocultos de Markov, así como algunas técnicas para el análisis de trayectorias en el seguimiento de objetos. En el capítulo 4 se describe el desarrollo del sistema de seguimiento visual y su metodología de desarrollo propuesta para los dos escenarios: a) Seguimiento de la mano para interactuar con la herramienta *T-WREX* y, b) Análisis de gestos, en el cual se detecta y sigue el brazo empleando marcas de color; incluye los métodos utilizados para la localización y seguimiento. En el capítulo 5, se presenta el análisis de movimiento de brazo, incluye la extracción de características de movimiento del brazo de un individuo, así como el análisis de su movimiento utilizando Modelos Ocultos de Markov. Aquí también se describe la metodología propuesta para el análisis de gestos descrita en esta tesis. En el capítulo 6 se describen los resultados experimentales y un análisis sobre ellos. Finalmente, en el capítulo 7 se presentan las conclusiones y el trabajo futuro.

# Capítulo 2

## Análisis del movimiento humano

### 2.1. Introducción

El análisis del movimiento humano es un área de gran interés debido a la gran cantidad de aplicaciones en las que se puede utilizar (deportes, medicina, entretenimiento, etc.). En particular, el análisis clínico ha motivado a los investigadores a crear sistemas para ayudar en la rehabilitación de los pacientes. En los últimos años, se ha puesto especial interés en desarrollar sistemas automáticos para el seguimiento de las extremidades superiores de una persona que ha sufrido un derrame cerebral orientadas a la rehabilitación.

En la literatura se pueden encontrar sistemas para la rehabilitación que se pueden clasificar de acuerdo a las tecnologías utilizadas para la captura del movimiento humano [64]:

- Sistemas basados en visión con marcas.
- Sistemas basados en visión libres de marcas.
- Sistemas basados en sensores.
- Sistemas basados en asistencia robótica.

En las siguientes secciones se presentan algunos trabajos realizados en estas cuatro tendencias para la captura del movimiento humano en general, y para la rehabilitación de pacientes en particular.

### 2.2. Sistemas basados en visión con marcas

Muchos de los trabajos iniciales en la literatura de Visión Computacional enfocados en el problema de recuperar el movimiento de figuras articuladas en una imagen han sido inspirados en el trabajo realizado por el psicólogo Gunnar Johansson, que en 1975 presentó su famoso experimento *Moving Light Display* (MLD) [29]. En sus experimentos, colocó pequeñas luces en las extremidades de los individuos (ver Figura 2.1) y grabó sus videos realizando diferentes actividades como caminar, correr y bailar. Obervando el movimiento de las luces en las articulaciones, llevó a cabo la tarea

## 2.2. Sistemas basados en visión con marcas

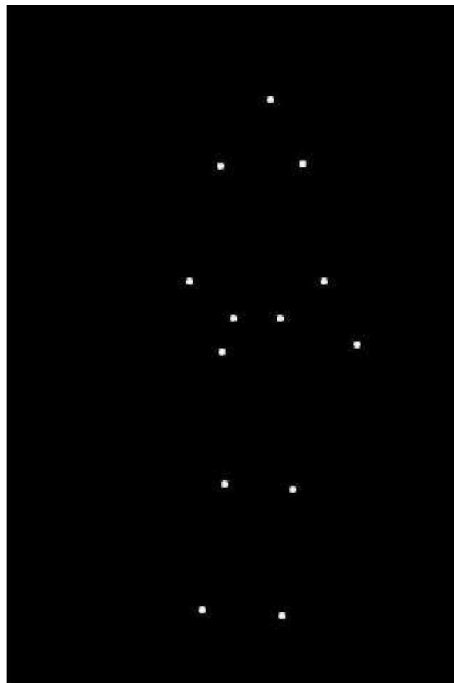


Figura 2.1: Imagen de una persona caminando usando MLD en sus articulaciones.

de clasificar las actividades y analizar ciertas características de movimiento, como el andar de una persona. También pudo observar que se podía determinar el género del individuo de acuerdo a la forma de moverse.

Si bien, el trabajo de Johansson estableció una sólida teoría sobre el seguimiento del movimiento humano, éste padecía de limitaciones como el costo computacional, oclusiones y procesamiento complicado. Sin embargo, los sistemas de seguimiento con marcas minimizan la incertidumbre de los movimientos de un individuo debido a que se enfocan sólo en la apariencia de las marcas. Consecuentemente, muchos sistemas de seguimiento basados en marcas están hoy en día disponibles en el mercado. A continuación se presentan algunos sistemas comerciales de seguimiento basados en marcas que pueden ser utilizados para sistemas de rehabilitación. Las principales desventajas de éste tipo de sistemas son la utilización de las marcas que limitan cierta movilidad del individuo, su alto costo comercial y su manejo especializado.

### 2.2.1. Qualisys

El sistema de captura de movimiento Qualisys (ver Figura 2.2), consiste de 1 a 16 cámaras, cada una emite un rayo o luz infrarroja [56]. Pequeñas marcas son adheridas al objeto o a la persona que va a ser seguida. La cámara lanza una luz infrarroja y las marcas la reflejan de regreso a la cámara. La cámara mide la posición en 2D del objeto reflejado y la combina con los datos 2D de las demás cámaras para calcular la posición 3D del objeto.

Los datos obtenidos por las cámaras pueden ser analizados por el Manejador de Movimiento Qualisys (QMM por sus siglas en inglés). La posición y orientación de cada segmento está determinado por tres o más puntos no-colineales. El modelo cinemático es calculado determinando las transformaciones del objeto seguido en el proceso de calibrado de las cámaras. Las desventajas prin-



## 2.2. Sistemas basados en visión con marcas



Figura 2.2: Sistema *Qualisys* en operación

principales son su alto costo comercial<sup>1</sup> y su complejidad para operar, ya que se requiere un gran número de cámaras para capturar el movimiento del cuerpo humano.

### 2.2.2. VICON

VICON es un sistema de seguimiento óptico en 3D (ver Figura 2.3), que está específicamente diseñado para el uso en ambientes virtuales [57], utilizando de tres a diez cámaras. Por su alta resolución y gran velocidad de procesamiento del software, el sistema es usado en ambientes para visualización militar y aeroespacial.

El sistema VICON calcula automáticamente el centro de las marcas adheridas a las articulaciones del cuerpo humano, reconstruyendo la posición en 3D y suministra información en 6 grados de libertad. Sin embargo, la configuración inicial del sistema requiere aproximadamente de una hora en la colocación de las marcas y del calibrado previo de las cámaras. Su alto costo comercial y su complejidad de operación de los dispositivos utilizados, son sus principales desventajas.

<sup>1</sup>16,000 euros en el año 2006



Figura 2.3: Sistema *VICON* en funcionamiento

## 2.3. Sistemas basados en visión libres de marcas



Figura 2.4: Sistema *CODA*

### 2.2.3. CODA

El sistema CODA (acrónimo de Cartesian Opto-electronic Dynamic Anthropometer) fue desarrollado inicialmente en 1974 en la Universidad Loughborough del Reino Unido, por David Mitchelson [51], mostrado en la Figura 2.4.

El sistema ha sido precalibrado para obtener mediciones en 3D. Se utilizan seis sensores que se adhieren al objeto para ser localizado y seguido. El sistema de seguimiento utiliza 360 grados de movimiento, útil en aplicaciones para animación virtual y en el análisis atlético. El cálculo de las marcas en coordenadas 3D se realiza en tiempo real con un retraso pequeño de 5 milisegundos. Sin embargo, su alto costo y su configuración hacen poco accesible el uso del sistema en ambientes en casa para el proceso de rehabilitación.

## 2.3. Sistemas basados en visión libres de marcas

En la sección anterior se mostraron las características de los sistemas basados en marcas, que son restrictivos en algunos grados de movilidad debido a las marcas adheridas al cuerpo humano. Una técnica menos restrictiva para la captura del movimiento humano son los sistemas libres de marcas, que son capaces de superar algunos problemas de oclusión y solo se concentran en los contornos, bordes o características del cuerpo humano. Esta es un área activa y prometedora pero desafiante, que se ha desarrollado en las últimas décadas. Sin embargo, el desarrollo de sistemas libre de marcas no es una tarea trivial, ya que existen varias dificultades [49, 50], incluyendo ambigüedades en la profundidad, deformidades en la apariencia, complejidad en los modelos cinemáticos y oclusiones.

Según Gavrilá [19], el análisis del movimiento del cuerpo humano se puede clasificar en dos enfoques dependiendo de la dimensionalidad del espacio de seguimiento: enfoque basado en 2 dimensiones (2D), y el enfoque basado en 3 dimensiones (3D).

### 2.3. Sistemas basados en visión libres de marcas

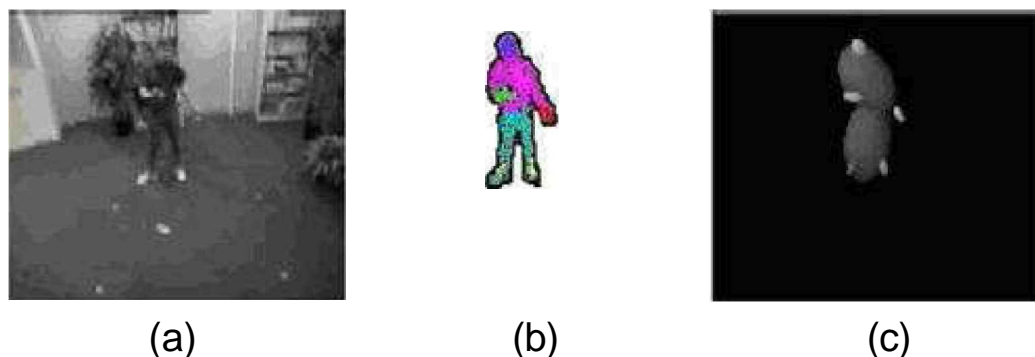


Figura 2.5: Demostración del sistema desarrollado por Wren. a) Imagen original, b) Segmentación, y c) Representación 2D de los *blobs*

#### 2.3.1. Seguimiento 2D

En el enfoque 2D sólo se analiza el movimiento del cuerpo en el plano bidimensional de la imagen. Sin embargo, algunos investigadores intentan proyectar una estructura 3D dentro del plano de la imagen para propósitos de procesamiento.

Debido a los movimientos arbitrarios y a la oclusión de ciertas partes del cuerpo, normalmente se usa conocimiento *apriori* acerca del movimiento humano en 2D usando segmentación. Por ejemplo, Wren [62] presenta un enfoque basado en regiones, considerando al cuerpo humano como un conjunto de *blobs* (manchas) que pueden ser identificadas por información espacial y distribución *Gaussiana* del color. En el proceso de inicialización, las regiones del primer plano de la imagen (*foreground*) pueden ser extraídas dado el modelo del fondo de la escena (*background*) previamente creado. Los blobs representan las manos, el rostro, las piernas, etc. Se realiza un análisis del contorno de la figura para identificar varias partes del cuerpo (ver Figura 2.5). Sin embargo, no se pueden identificar los puntos de unión en las extremidades de los brazos.

Baumberg y Hogg [6] utilizaron los modelos de contornos activos (Active Shape Models-ASM) para el seguimiento de peatones (ver Figura 2.6), y utilizaron *B-Splines* para representar los diferentes contornos; las regiones del primer plano de la imagen son extraídas por medio de la sustracción del fondo. Ellos aplicaron el filtro de Kalman para lograr las operaciones espacio-temporales, similar al trabajo desarrollado por Blake [10]. El sistema de Baumberg y Hogg genera automáticamente un modelo físico de la persona usando conjuntos de datos de entrenamiento [7]. Los modelos



Figura 2.6: Secuencias de imágenes utilizadas por Baumberg y Hogg, que muestran los contornos activos para el seguimiento de un peatón.

### 2.3. Sistemas basados en visión libres de marcas

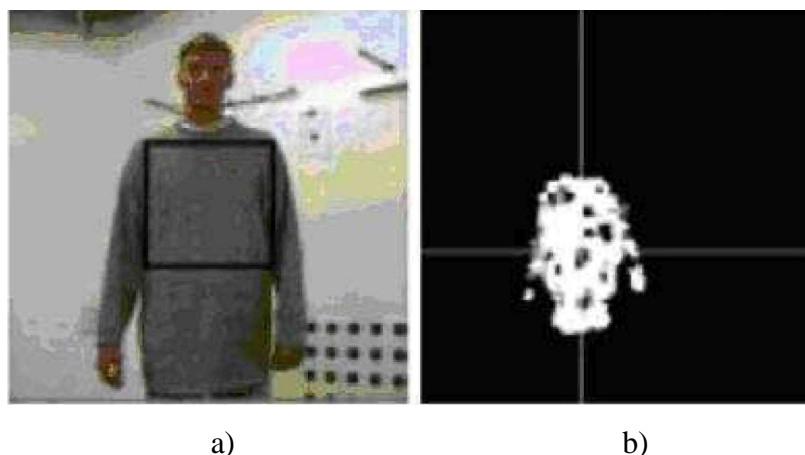


Figura 2.7: Sistema de seguimiento propuesto por Schlegel. a) Localización de la persona, b) Segmentación de la camisa.

del movimiento del cuerpo humano son limitados a simples contornos, donde las características descriptivas del movimiento no se pueden obtener fácilmente.

Schlegel [47] utiliza el seguimiento de una persona para interactuar con un robot móvil; utiliza un método de segmentación similar al método propuesto por Swain y Ballard [59], que consiste en la modelación de un objeto con base en un histograma de color multidimensional. El robot localiza a una persona identificando el color de su camisa dentro de una secuencia de imágenes capturadas por una cámara de video (ver Figura 2.7).

Bradski [13] propone un algoritmo llamado *CamShift* orientado al seguimiento del rostro humano para su uso en interfaces humano-computadora. Utiliza una ventana de búsqueda en donde identifica la mayor densidad de puntos pertenecientes al rostro. La ventana de seguimiento se posiciona en el centro de masa del objeto y se modifican sus dimensiones si éste cambia de tamaño (ver Figura 2.8). Este algoritmo ofrece una forma eficiente de seguimiento en tiempo real, sin embargo, tiene ciertos problemas en la fase de segmentación.

Avilés [3] propone un sistema de reconocimiento de gestos aplicado a la manipulación de robots móviles, el cual localiza el rostro humano, aplicando un algoritmo de rastreo radial [24] en la parte superior de la imagen por medio una segmentación de píxeles de color de la piel, basándose en un método del clasificador bayesiano propuesto por Jones [30]. Posteriormente el sistema localiza

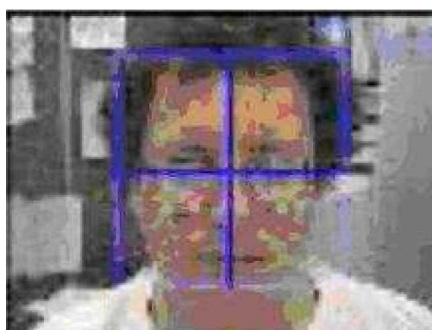


Figura 2.8: Imagen que muestra el seguimiento del rostro humano en el sistema propuesto por Bradski. La ventana de seguimiento localiza y sigue el rostro detectando el área de piel en la imagen.

### 2.3. Sistemas basados en visión libres de marcas



Figura 2.9: Imagen que muestra la localización y seguimiento del rostro y mano de un individuo en el sistema propuesto por Avilés. Imagen izquierda: Localización del rostro, hombro y mano. Imagen derecha: Secuencia de imágenes que muestran el seguimiento de la mano.

la mano derecha del usuario por medio de una búsqueda por regiones considerando medidas antropométricas del hombro de un individuo. Finalmente, el sistema sigue la mano derecha del individuo por medio de una ventana de búsqueda (ver Figura 2.9); los gestos son modelados basándose en la trayectoria descrita por el movimiento de la mano derecha, aplicando Modelos Ocultos de Markov (MOM). El sistema puede reconocer un total de cinco gestos: derecha, izquierda, acercar, atención y detenerse. Los gestos reconocidos son interpretados como comandos de movimiento para el robot móvil. El promedio ponderado de reconocimiento del sistema fue de 82.4%.

El mismo Avilés [4] propone un modelo para clasificar gestos similares. Emplea un clasificador Bayesiano *naive* en la ejecución de nueve gestos. Incluye información de postura de la mano con respecto al torso de un individuo. Afirma que el nuevo modelo tiene mayor grado de reconocimiento comparado con los *modelos ocultos de Markov*, además de un menor número de iteraciones en la etapa de entrenamiento del modelo.

Montero [38] propone un modelo computacional que integra técnicas de visión computacional y teoría de decisiones con el fin de reconocer los gestos manipulativos realizados por una persona en escenarios conocidos. Un escenario donde prueba el modelo es el de videoconferencia, en el cual detecta el rostro del expositor, así como la mano que manipula los objetos en base al color de piel. Detecta y sigue los objetos que conforman el entorno, como son libros, cuadernos, ratón de la compu-



Figura 2.10: Imagen que muestra la identificación de objetos de interés y su cercanía con la mano del usuario.



### 2.3. Sistemas basados en visión libres de marcas

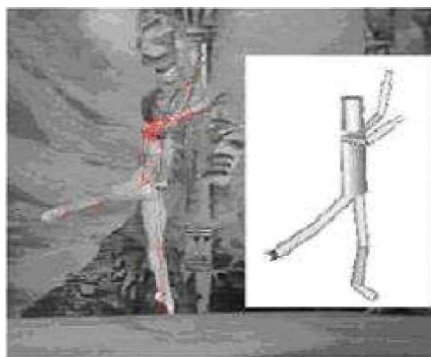


Figura 2.11: Figura de alambre del cuerpo humano.

tadora, etc. Se reconocen los gestos realizados por la persona y toma una decisión, la cual consiste en enfocar el área de interés dentro de la sala de videoconferencia (ver Figura 2.10).

#### 2.3.2. Seguimiento 3D

El seguimiento 3D intenta recuperar la posición articulada de un objeto a través del tiempo. La mayoría de los algoritmos de seguimiento 3D emplean *modelos* basados en la forma del objeto de interés, otros emplean múltiples cámaras para mejorar la robustez. Los modelos basados en la forma del objeto varían desde un simple modelo de estructura de alambre [8], hasta sofisticados modelos volumétricos en 3D [9,15,26].

El modelado de los movimientos del cuerpo humano a priori permite minimizar el problema del seguimiento. O'Rourke y Badler [39] encontraron que la predicción en el espacio de estados parece ser más estable que en el espacio de la imagen debido a la incorporación de conocimiento semántico previo. Su sistema de seguimiento consta de cuatro partes: predicción, síntesis, análisis de la imagen y estimación del estado. Esta estrategia ha sido utilizada en la mayoría de los sistemas de seguimiento existentes.

Las figuras de alambre (*stick*) es la representación de una estructura esquelética, que es normalmente considerada como una colección de segmentos y ángulos de articulaciones (ver Figura 2.11). Bharatkumar [8], por ejemplo, utiliza las figuras de alambre para modelar las extremidades inferiores del cuerpo humano, como la cadera y los tobillos.



Figura 2.12: Sistema de movimiento humano, propuesto por Black.

### 2.3. Sistemas basados en visión libres de marcas

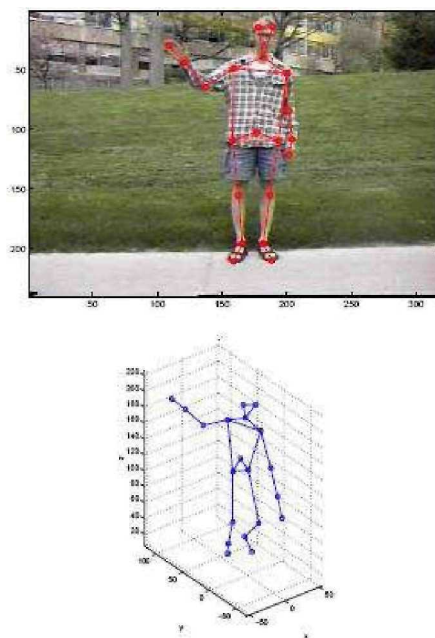


Figura 2.13: Imagen que muestra el sistema propuesto por Howe. Imagen superior: seguimiento en el espacio 2D empleando una sola cámara. Imagen inferior: Estimación de las articulaciones del cuerpo humano en el espacio 3D

Black [9] propone un sistema para el aprendizaje de modelos, parametrizados por la técnica del *flujo óptico* a través de una secuencia de imágenes (ver Figura 2.12). Una clase de movimiento está representada por un campo de flujo que es calculado por un conjunto de datos de entrenamiento usando la técnica de análisis de componentes principales (*PCA*). Los modelos de movimiento entrenados pueden ser usados para la estimación del flujo óptico o para el reconocimiento basado en modelos. Sin embargo, la desventaja del sistema es su alto costo computacional.

Howe [26] propone un método para reconstruir el movimiento del cuerpo humano en 3D usando un enfoque *Bayesiano*; utiliza conocimiento *a priori* de las posibles posiciones del cuerpo humano y las modela por medio de software especializado. Howe usa *Mezcla de Gaussianas* para modelar el movimiento; una vez entrenado el modelo, utiliza una sola cámara de video para el seguimiento 2D de las articulaciones en 40 partes del cuerpo. Finalmente, la información de la posición 2D de las articulaciones son tomadas como nueva información en un modelo *bayesiano* para obtener la estimación 3D del cuerpo humano (ver Figura 2.13). Aunque el método de Howe ofrece una forma de recuperar información 3D del cuerpo humano a partir de una sola cámara de video, se requiere de un proceso previo de modelado que es complicado, además de su alto costo computacional.

De manera similar al trabajo propuesto por Howe, Cheung y otros [15] proponen un método para el seguimiento completo del cuerpo humano utilizando información *deforma* y *color*. Su método, llamado por sus autores *Shape-from-Silhouette* (forma a partir de la silueta), utiliza el color para segmentar las partes del cuerpo en movimiento, basándose en información de imágenes anteriores y añadiendo restricciones para igualar el movimiento; de esta forma estiman las posiciones de las articulaciones en 3D. Ellos utilizan ocho cámaras calibradas para la captura del movimiento (ver Figura 2.14). Sin embargo, a pesar de los resultados satisfactorios, el uso de equipo especial y de la demanda computacional de este método, hacen poco deseable implementarlos en sistemas de tiempo real.

## 2.4. Sistemas basados en sensores

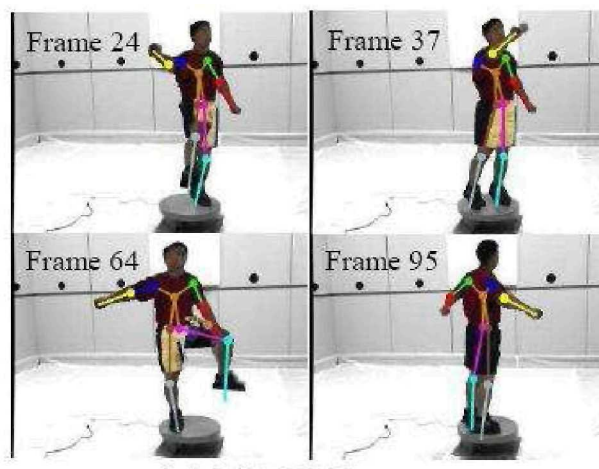


Figura 2.14: Secuencias de imágenes mostrando la representación del cuerpo humano empleando un modelo de alambre en el método propuesto por Cheung.

## 2.4. Sistemas basados en sensores

En los sistemas basados en sensores, los sensores son fijados al cuerpo humano para recolectar información de movimiento. Estos sensores son comúnmente clasificados como mecánicos, de inercia, acústicos, de radio o microondas, y magnéticos. Algunos de ellos tienen pequeñas marcas que pueden monitorear pequeñas amplitudes como el movimiento de los dedos.

Como parte de los sensores de inercia [32,63], los acelerómetros convierten aceleración lineal, angular o la combinación de ambas, en señales de salida [14]. Un ejemplo de acelerómetro es el dispositivo MTP [55] de la compañía *Xsens Motion Tech*, que mide la aceleración y campo magnético para obtener medidas 3D (ver Figura 2.15). Combinado con el software *MT9*, éste mecanismo provee información 3D en tiempo real en forma de ángulos de *Euler* y *Quaterniones*, a una frecuencia de 512 Hz. Debido a su tamaño compacto y a su alto rendimiento, el dispositivo MT9 ha sido fácilmente integrado en el campo de la bioinformática, robótica, animación, realidad virtual, etc. Sin embargo, el dispositivo (con seis unidades MT9) y el software de seguimiento MT9 tiene un costo muy elevado<sup>2</sup>.

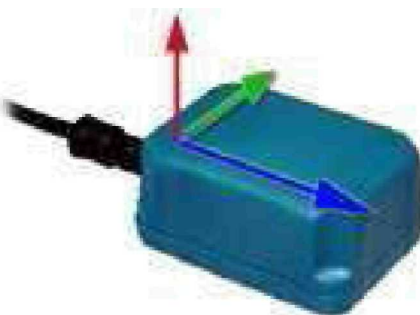


Figura 2.15: Imagen del dispositivo MT9

Desde finales de 1970's los investigadores han estudiado el uso de dispositivos de guantes para el análisis de los gestos realizados por la mano de un individuo. Los dispositivos basados en *Guantes* emplean sensores añadidos al guante que traducen las flexiones de los dedos y las convierten en señales eléctricas para determinar la postura de la mano (ver Figura 2.16). Los dispositivos de

<sup>2</sup> 15,000 euros en el año 2007.



## 2.5. Sistemas basados en asistencia robótica



Figura 2.16: Dispositivo CyberGlove

guante representativos son el *Dexterous Hand Master* [54], el *DataGlove* [53], el *CyberGlove* [52] y el *PowerGlove*. El *Dexterous Hand Master* y el *DataGlove* usan dispositivos electromagnéticos para el seguimiento de la mano, y el *PowerGlove* utiliza dos transmisores ultrasónicos montados en el guante y tres receptores montados en la pantalla de la computadora para determinar la localización de la mano dentro de 1/4 de pulgada. Sin embargo, estos sistemas requieren que el usuario use un dispositivo estorboso en la mano, y que la mano se dedique a la interfaz en lugar de que la interfaz se dedique al usuario. Además, la tecnología electromagnética para el seguimiento no funciona en lugares de trabajo donde estén presentes objetos de metal.

## 2.5. Sistemas basados en asistencia robótica

Recientemente, ejercicios repetitivos voluntarios administrados con asistencia de robots rehabilitadores, ha provisto una eficiente habilidad para el movimiento del brazo en personas que han sufrido un derrame cerebral [64]. Durante el curso de la rehabilitación, el movimiento humano es reflejado usando sensores adheridos al cuerpo, que consisten en sensores electromecánicos y electromagnéticos.

Reinkensmeyer y otros [44], desarrollaron un brazo robótico llamado *ARM-Guide* (Assisted Rehabilitation and Measurement Guide) para diagnosticar y tratar movimientos del brazo dañado después de un derrame cerebral y otras lesiones del cerebro (ver Figura 2.17). En sus experimentos, evalúan algunas habilidades motoras dañadas como la coordinación, la debilidad, etc. En su diseño, el antebrazo del individuo es amarrado a una tablilla especial que se mueve a lo largo del espacio de movimiento permitido. Un motor maneja los movimientos de la tablilla, y un codificador óptico montado en el motor indica la posición del brazo.

Aunque éste tipo de dispositivos administran información eficiente de las posiciones en las extremidades de los individuos, es notable su complejidad debido al costo que implica desarrollarlos, además de que limitan la movilidad del individuo.

## 2.6. Sistemas para la rehabilitación

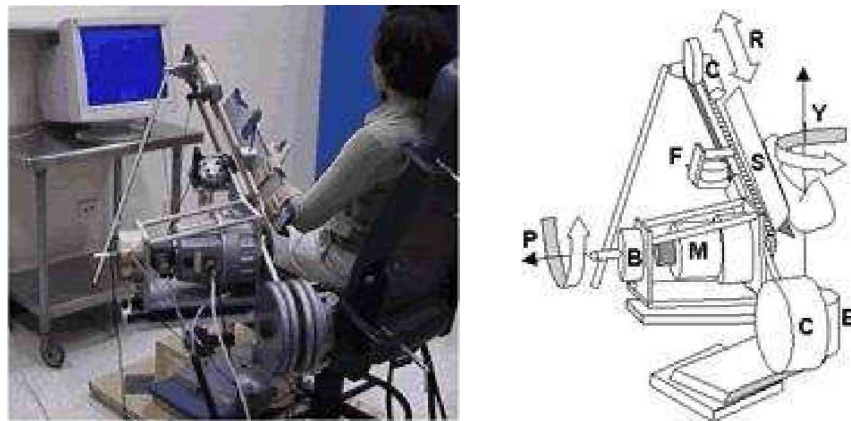


Figura 2.17: Ilustración del ARM-Guide. Imagen izquierda: Un paciente interactuando con el dispositivo robótico, Imagen derecha: Prototipo del ARM-Guide

## 2.6. Sistemas para la rehabilitación

El objetivo de la rehabilitación es permitir que las personas que han sufrido un derrame, obtengan un alto nivel de independencia para que puedan ser productivas en lo posible. Tradicionalmente, los pacientes llevan a cabo la rehabilitación con ayuda de un fisioterapeuta que diagnostica sus actividades durante la fisioterapia. Sin embargo, los altos costos hacen que los pacientes no reciban el suficiente tratamiento. Esto ha motivado a los investigadores a crear sistemas que auxilien en el proceso de rehabilitación, por medio del análisis en el movimiento de las extremidades. La mayoría de los sistemas propuestos para la rehabilitación emplean las técnicas mencionadas en las secciones anteriores.

Por ejemplo, Zhou y otros [65], proponen un sistema de monitoreo en tiempo real para medir el movimiento de las extremidades superiores del cuerpo humano orientado a la rehabilitación después de un derrame cerebral. Su sistema está basado en el uso del sensor de inercia comercial *MT9* [55] que es fijado en la muñeca del brazo del paciente (ver Figura 2.18). Por medio de un modelo de cinemática

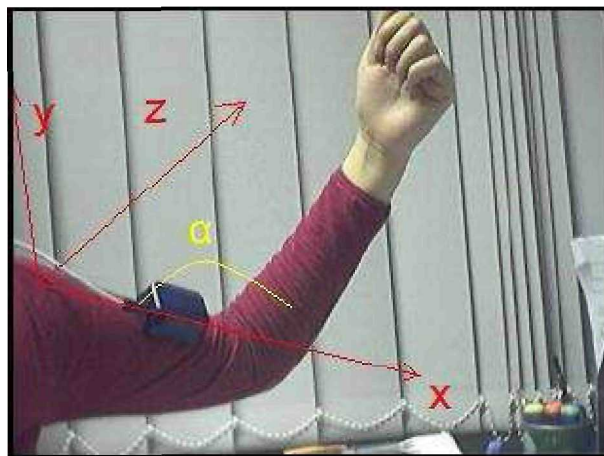


Figura 2.18: La imagen muestra el dispositivo MT9 fijado al brazo de un individuo, monitoreando el ángulo formado entre el brazo y el antebrazo.

## 2.6. Sistemas para la rehabilitación

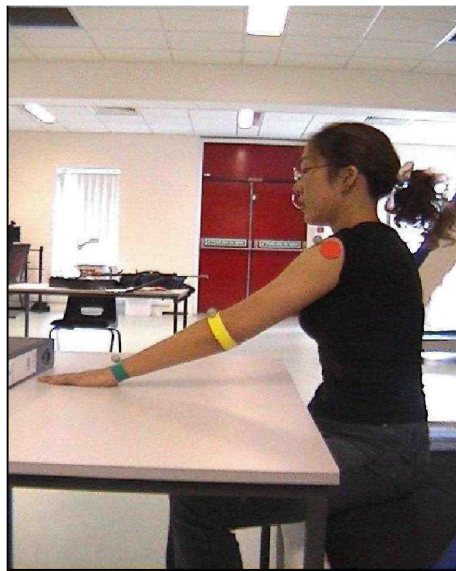


Figura 2.19: La imagen ilustra el sistema propuesto por Tao. Las marcas de color son fijadas en las articulaciones del brazo de un individuo para el seguimiento del mismo.

inversa, se obtienen los ángulos formados por el hombro y el codo, para medir el rendimiento en el movimiento del brazo del paciente. Los resultados de este sistema son comparados con el sistema comercial *Qualisys* [56], para verificar su rendimiento.

Tao y Hu [60], proponen un sistema para el movimiento del cuerpo humano basado en color para la rehabilitación en casa (Fig. 2.19). Este sistema tiene como finalidad seguir el brazo de una persona, combinando los dos enfoques de seguimiento del cuerpo humano basados en visión: sistemas basados en marcas y sistemas libres de marcas. Se adhieren pequeñas cintas de color en las principales articulaciones del brazo como son: muñeca, codo y hombro. Por medio de un algoritmo de seguimiento basado en color, detectan y siguen el brazo del paciente. Extraen información de los ángulos formados por el brazo y el antebrazo, y entre el hombro y el codo. Emplean el sistema comercial *Qualisys* [56], para medir la eficiencia de su sistema por medio de un algoritmo de correlación, que mide la similitud entre las trayectorias de los ángulos de ambos sistemas.

Tao y Zhou [61], proponen un sistema para la rehabilitación, integran técnicas de visión y de

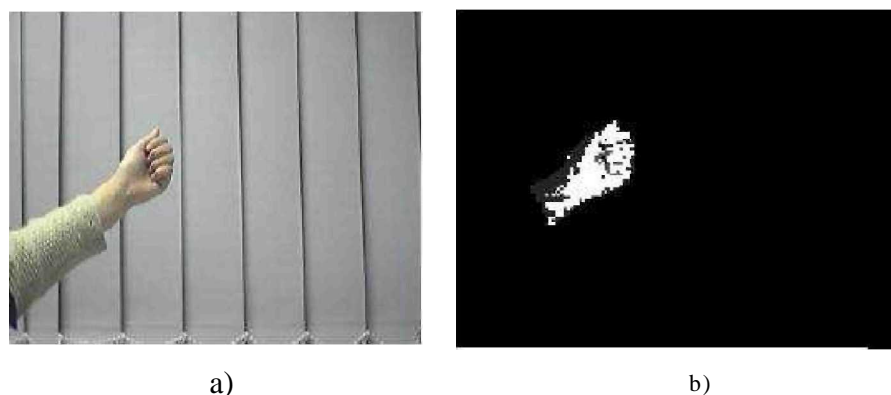


Figura 2.20: Segmentación de la mano en el sistema propuesto por Tao y Zhou. a) Imagen original, b) Imagen segmentada al aplicarse el algoritmo Camshift.

## 2.6. Sistemas para la rehabilitación

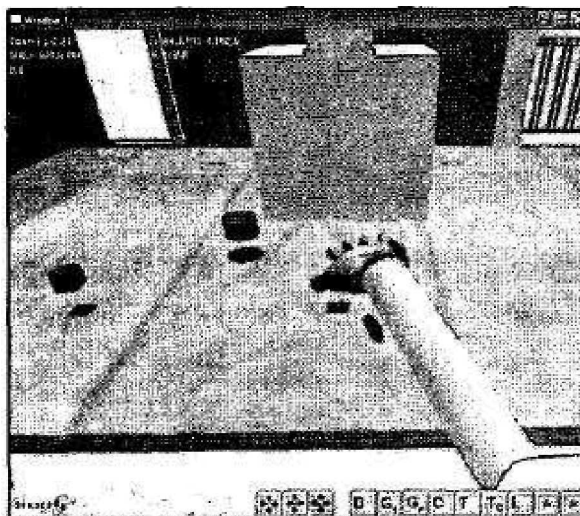


Figura 2.21: Vista del ambiente virtual del sistema propuesto por McNeil

sensores de inercia para seguir el movimiento del brazo de una persona en un espacio 3D en tiempo real (Fig. 2.20). Para el seguimiento basado en el sensor de inercia, utilizan el dispositivo *MT9*, que mide los ángulos de las articulaciones formadas por el brazo y el antebrazo. Para el seguimiento basado en visión utilizan el algoritmo *Camshift* [13]. La segmentación de la mano la realizan en base a histogramas de color en el espacio de color *HSV*. El sistema muestra buenos resultados en el seguimiento del brazo de una persona al integrar dos fuentes de información. Sin embargo, el uso del dispositivo de inercia impide realizar movimientos del brazo de forma más natural.

La tecnología de *Realidad Virtual* es una técnica prometedora que tiene la capacidad de crear ambientes interactivos en donde se pueden realizar ejercicios intensivos para manipular el entrenamiento de los movimientos de las extremidades superiores en pacientes que han sufrido un derrame cerebral [11]. McNeil y otros [37], proponen un sistema basado en realidad virtual para la rehabilitación (ver Figura 2.21). Su sistema está compuesto de tres sensores (acelerómetros) colocados en las articulaciones del brazo (hombro, codo y mano) para obtener medidas 3D de seguimiento, además de un *guante* para interactuar con los objetos en la escena. Afirman que el uso de ambientes virtuales es motivante para los pacientes en contraste con las terapias monótonas, dando como resultado una mejor rehabilitación.

La tecnología de *Realidad Mezclada* ha sido empleada en una variedad de sistemas de entrenamiento, arte, y en escenarios educacionales. Recientemente el desarrollo de sistemas de realidad mezclada ha provisto una alternativa para ambientes virtuales enfocados en el proceso de rehabilitación. La realidad mezclada son ambientes en donde los participantes pueden interactuar con dos fuentes de información: física (real) y digital (virtual), de una manera integrada. Pridmore y otros [42], proponen un sistema de *Realidad Mezclada* para la rehabilitación después de un derrame cerebral. Su sistema se enfoca en simular el escenario de tomar una taza caliente. Para esto, mezclan objetos reales (tazas) en ambientes virtuales. El paciente interactúa con las tazas físicas y el sistema las reproduce en un ambiente virtual. Para la localización y seguimiento emplean el método de *intersección de histogramas* propuesto por Swain y Ballard [5]. Aseguran que el uso de realidad mezclada tiene una mayor motivación en los pacientes en comparación con la *Realidad Virtual*, debido a dos aspectos principales: primero, es un sistema completo y no sólo se enfoca en un ambiente virtual, y segundo, la información con respecto al ambiente real es vista como un recurso, no como un problema.

### 2.7. Resumen

Los sistemas basados en visión con marcas, utilizan marcas adheridas a las articulaciones del cuerpo para reconstruirlo en un espacio 3D. Si bien, los sistemas existentes en el mercado como Qualisys, CODA y VICON, ofrecen un alto rendimiento en el seguimiento 3D, su alto costo comercial, así como su configuración y manejo complicado, hacen poco deseable implementarlos en sistemas para la rehabilitación de bajo costo y de operación no-especializada. Por otra parte, los sistemas basados en visión libres de marcas es un área activa de investigación, debido a que no es una tarea trivial, ya que se enfrentan a problemas tales como la oclusión, el procesamiento en tiempo real, ambigüedades en la profundidad, etc. Hacen uso de características tales como el color, la forma, los bordes, etc., para seguir el cuerpo humano. El seguimiento 2D sigue a un objeto de la escena 3D proyectado en el plano imagen, pero se ve limitado por falta de información espacial. El seguimiento 3D intenta recuperar la posición articulada de un objeto a través del tiempo, ya sea con o sin conocimiento previo del modelo a seguir. Sin embargo, el procesamiento es demandante computacionalmente. En los sistemas basados en sensores, los sensores son fijados al cuerpo humano para recolectar información de movimiento. Debido a su tamaño pequeño y a su alto rendimiento son utilizados en diversas áreas como la bioinformática, realidad virtual, animación, etc. Sin embargo, limitan la movilidad de las acciones de forma natural, además de ser costosos comercialmente. De manera similar, los dispositivos de *guante*, son colocados en las manos para medir las flexiones de los dedos y poder determinar la postura de la mano. No obstante, es un mecanismo estorboso para el usuario, además de que no funciona en ambientes cargados de metal. Los sistemas basados en asistencia robótica, utilizan dispositivos adheridos al brazo para controlar el movimiento y medir su posición 3D, como los brazos robóticos. Pero sus desventajas radican en que los movimientos naturales del paciente se ven limitados, y además de tener altos costos de desarrollo.

Se han propuesto varios sistemas para la rehabilitación que ayuden al paciente a realizar sus ejercicios fisioterapéuticos. Algunos están basados en sensores de inercia, o en técnicas de visión computacional, o en una combinación de ambas. También utilizan la tecnología de *guantes* junto con la Realidad Virtual para motivar a los pacientes a través de ejercicios para interactuar con objetos en la escena. Si bien, estos sistemas obtienen información de movimiento del brazo tales como ángulos, posiciones de la mano, etc., no se ha propuesto alguna métrica que permita evaluar automáticamente el grado de avance en el proceso de rehabilitación en base a la información del movimiento del brazo.

# Capítulo 3

## Modelos Ocultos de Markov

Durante el proceso de rehabilitación, los movimientos del brazo del paciente necesitan ser localizados y aprendidos para que los movimientos incorrectos puedan ser evaluados y así darle retroalimentación al paciente. De hecho, la rehabilitación es un proceso dinámico que usa técnicas disponibles para corregir cualquier comportamiento de movimiento no deseado para alcanzar uno esperado (por ejemplo, una posición ideal) [64]. Comparar la trayectoria del movimiento del brazo con un conjunto de datos previamente almacenados, no es una tarea trivial. Más aun, si se necesita comparar una trayectoria con otra existen diversos problemas [41], por ejemplo: a) las trayectorias necesitan tener la misma duración de muestreo, b) los datos de las trayectorias son tomados en diferentes instantes de tiempo, y c) si se normalizan las trayectorias para que tengan la misma duración y conjunto de datos, ésta normalización destruye las propiedades temporales de la trayectoria. Por lo tanto, se necesita de una técnica que pueda medir efectivamente las diferencias entre trayectorias sin considerar restricciones. En este capítulo se presentan algunas técnicas para medir las diferencias entre trayectorias basadas en Modelos Ocultos de Markov (MOM), que pueden capturar las propiedades dinámicas de la trayectoria. En la sección 3.1 se presentan los Modelos Ocultos de Markov. En la sección 3.2 se describen algunas técnicas para comparar MOM. Finalmente, en la sección 3.3 se presenta un resumen del capítulo.

### 3.1. Descripción de los Modelos Ocultos de Markov

Los Modelos Ocultos de Markov (MOM) son una herramienta útil para modelar series de datos en el tiempo. Los MOM han sido usados en la mayoría de los sistemas actuales para el reconocimiento de patrones y de Inteligencia Artificial (IA) tales como: reconocimiento de voz, biología molecular, compresión de datos, entre otros. En las últimas décadas los MOM han sido usados en aplicaciones de Visión Computacional tales como el modelado de secuencias de imágenes, el seguimiento de objetos, el reconocimiento visual de gestos, etc. Una razón para la popularidad de los MOM ha sido su habilidad para caracterizar con alto grado de precisión, datos que presentan una estructura secuencial en presencia de ruido y ligeras variaciones.

### 3.1. Descripción de los Modelos Ocultos de Markov

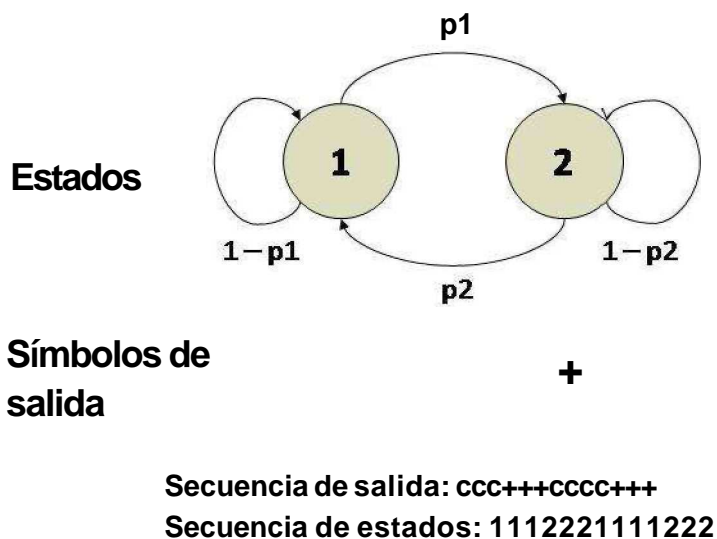


Figura 3.1: Cadena de Markov mostrando las probabilidades de transición y los símbolos de salida, en el ejemplo de lanzar una moneda. La secuencia de estados puede determinarse a partir de la secuencia de salida.

#### 3.1.1. Definición

Antes de describir los Modelos Ocultos de Markov, es necesario describir los *procesos de Markov*. Un proceso estocástico es llamado un proceso de Markov *de j-ésimo* orden si la densidad de la probabilidad condicional de un evento actual, dado todos los eventos presentes y pasados, depende solamente de los *j-ésimos* eventos más recientes. Si el evento actual en el tiempo  $t$  depende únicamente del evento anterior más reciente en el tiempo  $t-1$ , entonces este proceso es un proceso de Markov *de primer orden*, en ese caso, dado el valor del estado  $S_{t-1}$ , el estado actual  $S_t$  es independiente de todos los estados anteriores al instante  $t-1$  [20]. En otras palabras, el estado  $S_t$ , en algún tiempo  $t$ , encapsula todo lo que se necesita saber acerca de la historia del proceso para predecir el proceso futuro.

Un Modelo Oculto de Markov (denotado  $X$ ) es un proceso doblemente estocástico. La primera capa estocástica es un proceso de Markov de *primer orden*, que determina la transición entre los estados. La segunda capa es el conjunto de probabilidades de salida, que genera observaciones de salida para cada estado. Los estados no son determinados de manera directa a partir de las observaciones, por lo tanto, son ocultos. Por ejemplo, considérese el siguiente escenario. Usted se encuentra en un cuarto con una cortina que impide ver lo que sucede del otro lado del cuarto. Del otro lado de la cortina, se encuentra una persona que lanza una moneda (o varias monedas). Usted solo puede conocer el resultado del lanzamiento. Los posibles estados del sistema pueden ser Cara (c) o Cruz (+). Las probabilidades de transición entre los estados y las respuestas de salida se muestran en la Figura 3.1. En este caso, se puede observar que cada estado genera una única salida, esto es, la salida es determinista, de modo que es posible recuperar la secuencia de estados a partir de la secuencia de salida. Esto es un ejemplo de una cadena de Markov. Sin embargo, si la persona lanza dos monedas, esto es, si cada estado del sistema puede generar cualquiera de las dos salidas (Cara o Cruz), no es posible recuperar sin ambigüedad la secuencia de estados a partir de las observaciones, tal como se



### 3.1. Descripción de los Modelos Ocultos de Markov

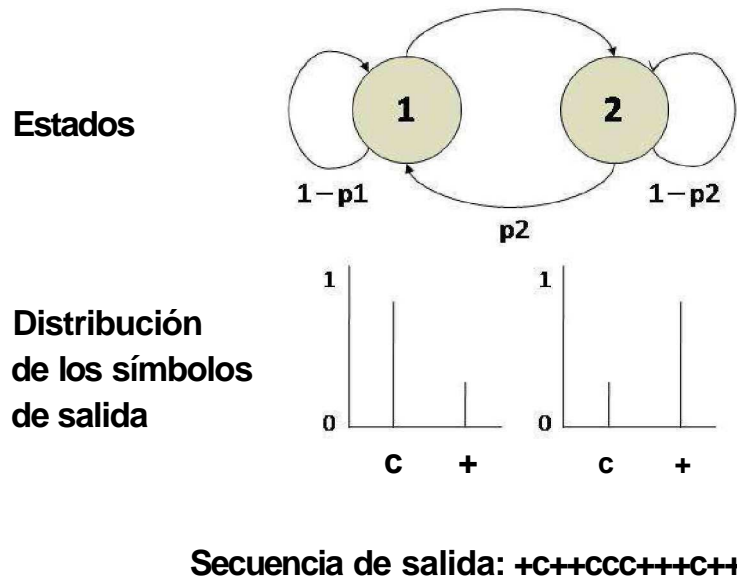


Figura 3.2: Modelo Oculto de Markov mostrando las probabilidades de transición y de salida, en el ejemplo de lanzar dos monedas. La secuencia de estados no puede determinarse a partir de la secuencia de salida.

muestra en la Figura 3.2, entonces los estados son ocultos. La definición formal de un Modelo Oculto de Markov con observaciones discretas se presenta a continuación.

#### 3.1.2. Elementos de un MOM

Formalmente, un Modelo Oculto de Markov *Discreto* de primer orden se caracteriza por los siguientes elementos [43]:

1.  $N$ , el número de estados en el modelo. El conjunto de estados individuales se denota por  $S = \{S_1, S_2, \dots, S_N\}$ , y un estado en el tiempo  $t$  como  $q_t$ .
2.  $M$ , el número de símbolos de observación por estado. Los símbolos de observación se interpretan como la salida física del sistema modelado. Los símbolos individuales se denotan como  $V = \{v_1, v_2, \dots, v_M\}$
3. La distribución de probabilidad de transición entre estados  $A = \{a_{ij}\}$ , donde:

$$a_{ij} = P[q_{t+1} = S_j | q_t = S_i], \quad 1 \leq i, j \leq N$$

y con las siguientes restricciones de probabilidad:

$$a_{ij} \geq 0, \quad 1 \leq i, j \leq N \quad y \quad \sum_{j=1}^N a_{ij} = 1, \quad 1 \leq i \leq N$$



### 3.1. Descripción de los Modelos Ocultos de Markov

4. La distribución de probabilidad de símbolos de observación en el estado  $j$ ,  $B = \{b_j(k)\}$ , donde:

$$b_j(k) = P[v_k \text{ en } t | q_t = S_j], \quad 1 \leq j \leq N, \quad 1 \leq k \leq M$$

con las siguientes restricciones de probabilidad:

$$b_j(k) \geq 0, \quad 1 \leq j \leq N, \quad 1 \leq k \leq M \quad \text{y} \quad \sum_{k=1}^M b_j(k) = 1, \quad 1 \leq j \leq N$$

5. La distribución de probabilidades iniciales para los estados  $\pi = \{\pi_i\}$ , donde:

$$\pi_i = P[q_1 = S_i], \quad 1 \leq i \leq N$$

con las siguientes restricciones de probabilidad:

$$\pi_i \geq 0, \quad 1 \leq i \leq N \quad \text{y} \quad \sum_{i=1}^N \pi_i = 1$$

Un Modelo Oculto de Markov se puede especificar de forma compacta por medio de la siguiente notación:

$$\lambda = (A, B, \pi)$$

Donde  $A$  representa la distribución de probabilidad de transición de estados,  $B$  representa la distribución de probabilidad de símbolos de observación y,  $\pi$  representa la distribución de probabilidades iniciales.

#### 3.1.3. Tres problemas MOM básicos

Existen tres problemas básicos de interés que pueden ser resueltos para que un modelo ( $X$ ) sea útil en aplicaciones reales. Estos problemas son descritos en [43], en la siguiente forma:

1. **Evaluación.** Dada una secuencia de observación  $O = O_1 O_2 \dots O_T$ , y el modelo  $X = \{A, B, \pi\}$ , ¿cómo podemos calcular eficientemente  $P(O | \lambda)$ , la probabilidad de la secuencia de observaciones dado el modelo?. Esto es, la probabilidad de que la secuencia observada sea generada por el modelo.
2. **Decodificación.** Dada una secuencia de observación  $O = O_1 O_2 \dots O_T$ , y el modelo  $X$ , ¿cómo podemos encontrar una secuencia de estados correspondiente  $Q = q_1 q_2 \dots q_T$  que haya generado dicha secuencia de observación?. Se trata de descubrir la parte oculta del modelo, esto es, encontrar la secuencia *más probable* de estados.
3. **Estimación.** ¿Cómo podemos ajustar los parámetros del modelo  $\lambda = \{A, B, \pi\}$  que maximice  $P(O | \pi)$ ?

Esta tesis se aboca principalmente a los problemas 1 y 3.

### 3.1. Descripción de los Modelos Ocultos de Markov

#### Solución al problema de *evaluación*

Se desea calcular la probabilidad de la secuencia de observación  $O = O_1O_2 \dots O_T$ , dado el modelo  $\lambda$ , ( $P(O|\lambda)$ ). La forma más simple (directa) de resolverlo es enumerando todas las posibles secuencias de estados de longitud  $T$  (el número de observaciones). Por ejemplo, considérese la secuencia de estados  $Q = q_1q_2 \dots q_T$ , donde  $q_1$  es el estado inicial. La probabilidad de la secuencia de observación  $O$  para ésta secuencia de estados es:

$$P(O|Q, \lambda) = \prod_{t=1}^T P(O_t|q_t, \lambda) \quad (3.1)$$

Suponiendo independencia estadística de las observaciones [431, se tiene:

$$P(O|Q, \lambda) = b_{q_1}(O_1) \cdot b_{q_2}(O_2) \cdot \dots \cdot b_{q_T}(O_T) \quad (3.2)$$

La probabilidad de la secuencia de estados puede ser escrita como:

$$P(Q|\lambda) = \pi_{q_1} a_{q_1q_2} a_{q_2q_3} \cdot \dots \cdot a_{q_{T-1}q_T} \quad (3.3)$$

La probabilidad conjunta de  $O$  y  $Q$ , esto es, la probabilidad que  $O$  y  $Q$  ocurran simultáneamente, es el resultado del producto de los términos anteriores:

$$P(O, Q|\lambda) = P(O|Q, \lambda)P(Q, \lambda) \quad (3.4)$$

La probabilidad de la secuencia de observaciones  $O$  (dado el modelo) se obtiene sumando la probabilidad conjunta sobre todas las posibles secuencias de estados  $q$ :

$$P(O|\lambda) = \sum_{\text{todas } Q} P(O|Q, \lambda)P(Q|\lambda) \quad (3.5)$$

$$= \sum_{q_1, q_2, \dots, q_T} \pi_{q_1} b_{q_1}(O_1) a_{q_1q_2} b_{q_2}(O_2) \cdot \dots \cdot a_{q_{T-1}q_T} b_{q_T}(O_T) \quad (3.6)$$

El cálculo de  $P(O|\lambda)$ , de acuerdo a la forma *directa*, es de un orden de complejidad  $2T \cdot N^T$  operaciones; esto se debe a que, por cada tiempo  $t = 1, 2, \dots, T$ , existen  $N$  posibles estados que pueden ser considerados (esto es, existen  $N^T$  posibles secuencias de estados), y por cada secuencia de estados se requieren alrededor de  $2^T$  operaciones en cada término en la sumatoria de la ecuación 3.6. Para propósitos prácticos, este cálculo es computacionalmente infactible. Existe un procedimiento más efectivo para resolver el problema de *evaluación*, el algoritmo de *Avance-Retroceso* (Forward - Backward) [43]. Considérese la variable  $\alpha_t(i)$  denominada *hacia adelante* (forward variable), definida como:

$$\alpha_t(i) = P(O_1O_2 \dots O_t, q_t = S_i|\lambda) \quad (3.7)$$

que corresponde a la probabilidad de una secuencia parcial de observación,  $O_1O_2, \dots, O_t$ , (hasta el tiempo  $t$ ) y el estado  $S_i$  en el tiempo  $t$ , dado el modelo  $X$ . Se puede resolver inductivamente a la variable  $\alpha_t(i)$  como sigue:

#### 1. Inicialización:

$$\alpha_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N \quad (3.8)$$

### 3.1. Descripción de los Modelos Ocultos de Markov

2. Inducción:

$$\alpha_{t+1}(j) = \left[ \sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(O_{t+1}), \quad 1 \leq t \leq T-1, \quad 1 \leq j \leq N \quad (3.9)$$

3. Terminación:

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i) \quad (3.10)$$

En el paso 1), se inicializa la variable *hacia adelante* como la probabilidad conjunta del estado  $S_i$  y la observación inicial  $O_1$ . En el paso 2), se calcula la variable *hacia adelante* inductivamente y a través del tiempo. Finalmente, en el paso 3), se calcula la probabilidad deseada  $P(O|\lambda)$  como la suma de las variables terminales  $T(I)$ .

El problema de evaluación también se puede resolver con el uso de la variable ( $\beta_t(i)$ ) denominada *hacia atrás* (backward variable), definida como [43]:

$$\beta_t(i) = P(O_{t+1}O_{t+2}\cdots O_T | q_t = S_i, \lambda) \quad (3.11)$$

que corresponde a la probabilidad de la secuencia parcial del tiempo  $t+1$  hasta  $T$ , dado el estado  $S_i$  en el tiempo  $t$  y el modelo  $X$ . De nuevo, se puede resolver a la variable  $\beta_t(i)$  inductivamente, de la siguiente manera:

1. Inicialización:

$$\beta_T(i) = 1, \quad 1 \leq i \leq N \quad (3.12)$$

2. Inducción:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \quad t = T-1, T-2, \dots, 1, \quad 1 \leq i \leq N \quad (3.13)$$

3. Terminación:

$$P(O|\lambda) = \sum_{i=1}^N \pi_i b_i(O_1) \beta_1(i) \quad (3.14)$$

#### Solución al problema de *decodificación*

Se desea encontrar la secuencia de estados *más probable*  $Q$  que haya generado una secuencia de observaciones dada  $O$ . Existe una técnica formal para encontrar ésta secuencia de estados, basada en métodos de programación dinámica, y es llamada el algoritmo de *Viterbi* [43]. Para hallar la mejor secuencia de estados,  $Q = \{q_1 q_2 \dots q_T\}$ , Para una secuencia de observaciones dada  $O = \{O_1 O_2 \dots O_T\}$ , se define la siguiente variable:

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P[q_1 q_2 \dots q_t = i, O_1 O_2 \dots O_t | \lambda] \quad (3.15)$$

### 3.1. Descripción de los Modelos Ocultos de Markov

es decir,  $S_t(i)$  es la probabilidad más alta a través de una ruta única, en el tiempo  $t$ , que cuenta para las primeras  $t$  observaciones y termina en el estado  $S_j$ . Por inducción se tiene:

$$\delta_{t+1}(i) = [\max_j \delta_t(i) a_{ij}] \cdot b_j(O_{t+1}) \quad (3.16)$$

Para mantener la secuencia de estados, es necesario mantener un seguimiento del argumento que maximice  $S_{t+1}(i)$ , para cada  $t$  y  $j$  por medio de un arreglo  $\Psi_t(j)$ . El proceso completo para hallar la mejor secuencia de estados es como sigue:

1. Inicialización:

$$\delta_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N \quad (3.17)$$

$$\psi_1(i) = 0 \quad (3.18)$$

2. Recursión:

$$\delta_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_j(O_t), \quad 2 \leq t \leq T, \quad 1 \leq j \leq N \quad (3.19)$$

$$\psi_t(j) = \arg \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}], \quad 2 \leq t \leq T, \quad 1 \leq j \leq N \quad (3.20)$$

3. Terminación:

$$P^* = \max_{1 \leq i \leq N} [\delta_T(i)] \quad (3.21)$$

$$q_T^* = \arg \max_{1 \leq i \leq N} [\delta_T(i)] \quad (3.22)$$

4. Recuperar la secuencia de estados:

$$q_t^* = \psi_{t+1}(q_{t+1}^*), \quad t = T-1, T-2, \dots, 1 \quad (3.23)$$

### Solución al problema de estimación

Se desea ajustar los parámetros del modelo  $(A, B, \Pi)$  para maximizar una secuencia de observación  $O$  dado el modelo  $X$ , esto es,  $P(O|\lambda)$ . Una forma de ajustar los parámetros del modelo  $\lambda$  es mediante el algoritmo de *Baum-Welch* [43].

Para describir el procedimiento de reestimación (actualización y mejoramiento iterativo) de los parámetros del MOM, se necesita definir primero la variable auxiliar  $\xi_t(i, j)$  [43], la cual define la probabilidad de estar en el estado  $S_i$  en el tiempo  $t$ , y en el estado  $S_j$  en el tiempo  $t+1$ , dado el modelo y la secuencia de observación. Esto es:

$$\xi_t(i, j) = P(q_t = S_i, q_{t+1} = S_j | O, \lambda) \quad (3.24)$$

La variable  $\xi_t(i, j)$  se puede definir en términos de las variables *Avance - Retroceso* (*Forward - Backward*), de la siguiente manera:

$$\begin{aligned} \xi_t(i, j) &= \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(O|\lambda)} \\ &= \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)} \end{aligned} \quad (3.25)$$

### 3.1. Descripción de los Modelos Ocultos de Markov

donde el numerador es justamente  $P(q_t = S_i, q_{t+1} = S_j, O|\lambda)$  y la división por  $P(O|\lambda)$  nos da la medida de la probabilidad deseada.

También es necesario definir la variable  $\gamma_t(i)$ , que corresponde a la probabilidad de estar en el estado  $S_i$  en el tiempo  $t$ , dada la secuencia de observación y el modelo. Esto es:

$$\gamma_t(i) = P(q_t = S_i | O, \lambda)$$

Se puede relacionar a la variable  $\gamma_t(i)$  con la variable  $\xi_t(i, j)$  si se suma sobre  $j$ , esto es:

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j)$$

Si se suma  $\gamma_t(i)$  sobre el tiempo  $t$ , se obtiene una cantidad que puede ser interpretada como el número de veces (sobre el tiempo) que el estado  $S_i$  es visitado. De manera similar, la suma de  $\xi_t(i, j)$  sobre  $i$  (desde  $t = 1$  hasta  $t = T - 1$ ) puede ser interpretada como el número esperado de transiciones del estado  $S_i$  al estado  $S_j$ . Esto es:

$$\sum_{t=1}^{T-1} \gamma_t(i) = \text{número de transiciones del estado } S_i \quad (3.28)$$

$$\sum_{t=1}^{T-1} \xi_t(i, j) = \text{número de transiciones del estado } S_i \text{ al estado } S_j \quad (3.29)$$

Usando las fórmulas anteriores, se puede obtener un método para reestimar los parámetros de un MOM. Un conjunto razonable de fórmulas de reestimación para  $\pi, A$ , y  $B$  son:

$$\begin{aligned} \bar{\pi}_i &= \text{número de veces en el estado } S_i \text{ en el tiempo } (t = 1) = \gamma_1(i) \quad (3.30) \\ \bar{a}_{ij} &= \frac{\text{número de transiciones del estado } S_i \text{ al estado } S_j}{\text{número de transiciones del estado } S_i} \end{aligned}$$

$$= \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad (3.31)$$

$$\begin{aligned} \bar{b}_j(k) &= \frac{\text{número de veces en el estado } j \text{ y haber observado el símbolo } v_k}{\text{número de veces en el estado } j} \\ &= \frac{\sum_{t=1}^T \gamma_t(j), O_t = v_k}{\sum_{t=1}^T \gamma_t(j)} \quad (3.32) \end{aligned}$$

Si iterativamente se utiliza el modelo  $\bar{\lambda} = (\bar{A}, \bar{B}, \bar{\pi})$  para repetir la reestimación, entonces se puede mejorar la probabilidad  $P(O|\lambda)$  hasta alcanzar algún óptimo local [43].

#### 3.1.4. Topologías de los MOM

Un MOM puede ser representado como un grafo dirigido, donde los nodos representan a los estados y las flechas representan las transiciones entre los estados. La topología específica que permita modelar de la mejor forma posible las propiedades observadas depende en gran medida de las características del problema. Las topologías más usadas son [43]:

1. *Completamente conectadas o ergódicas*: En las cuales cada estado del modelo puede ser alcanzado desde cualquier otro estado en un número finito de pasos (ver Figura 3.3a).

### 3.2. Métricas para la comparación de trayectorias

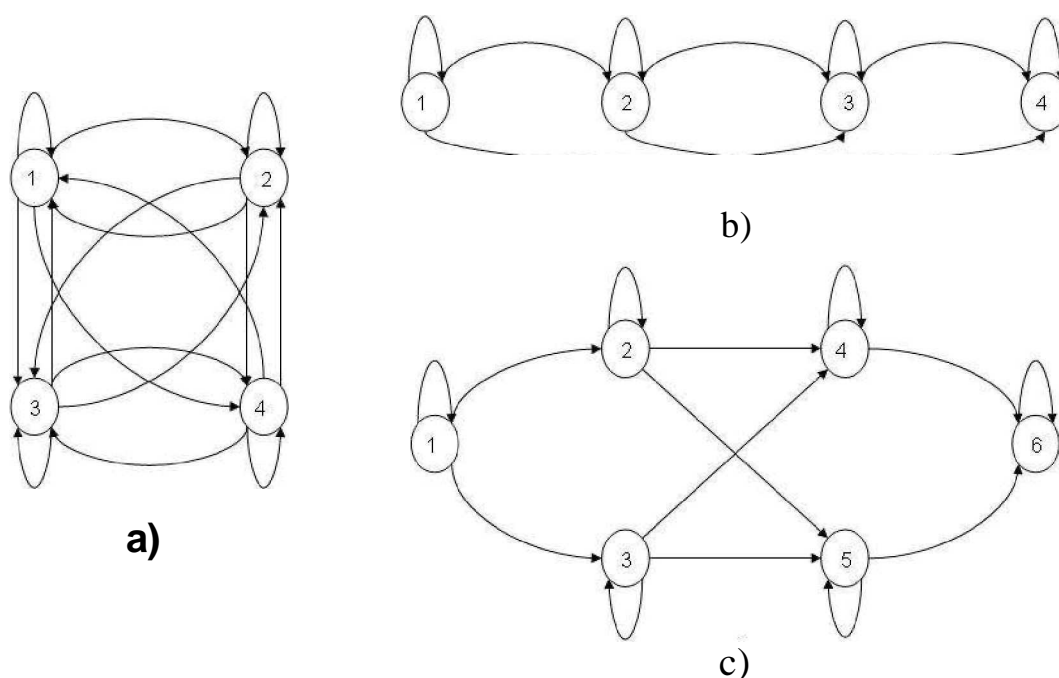


Figura 3.3: Ilustración de los 3 tipos de topologías de los MOM. a) Modelo *ergódico* de 4 estados, b) Modelo *izquierda-derecha* de 4 estados, c) Modelo *izquierda derecha paralelo* de 6 estados.

2. *Izquierda-derecha, hacia adelante o Bakis*: Las cuales tienen la propiedad de que en la medida que el tiempo crece se avanza en la secuencia de observación asociada  $O$  y en esa misma medida el índice que señala el estado del modelo permanece o crece, es decir, los estados del sistema van de izquierda a derecha (ver Figura 3.3b). En secuencias biológicas, en reconocimiento de la voz, y en el reconocimiento de gestos visuales [3, 58], estas topologías modelan bien los aspectos lineales de las secuencias.
3. *Izquierda-derecha paralelas*: Son dos topologías izquierda-derecha conectadas entre sí (ver Figura 3.3c).

En general, la topología de un MOM se define experimentalmente, variando ya sea el número de estados o las transiciones entre ellos, con el fin de obtener refinamientos sucesivos del modelo y mejorar su respuesta frente a las observaciones del evento representado [3,43,58].

## 3.2. Métricas para la comparación de trayectorias

Avances recientes en el seguimiento de objetos han hecho posible obtener trayectorias de movimiento espacio-temporales para su análisis. El tema principal en la evaluación de los resultados del seguimiento es una métrica de la distancia que determine la similitud entre las trayectorias [41]. La mayoría de las medidas existentes [28,48], calculan la distancia media entre los puntos correspondientes de dos trayectorias de igual duración. Estadísticas suplementarias como las distancias basadas en la varianza, la mediana, el mínimo, el máximo, son también propuestas para extender la descripción

### 3.2. Métricas para la comparación de trayectorias

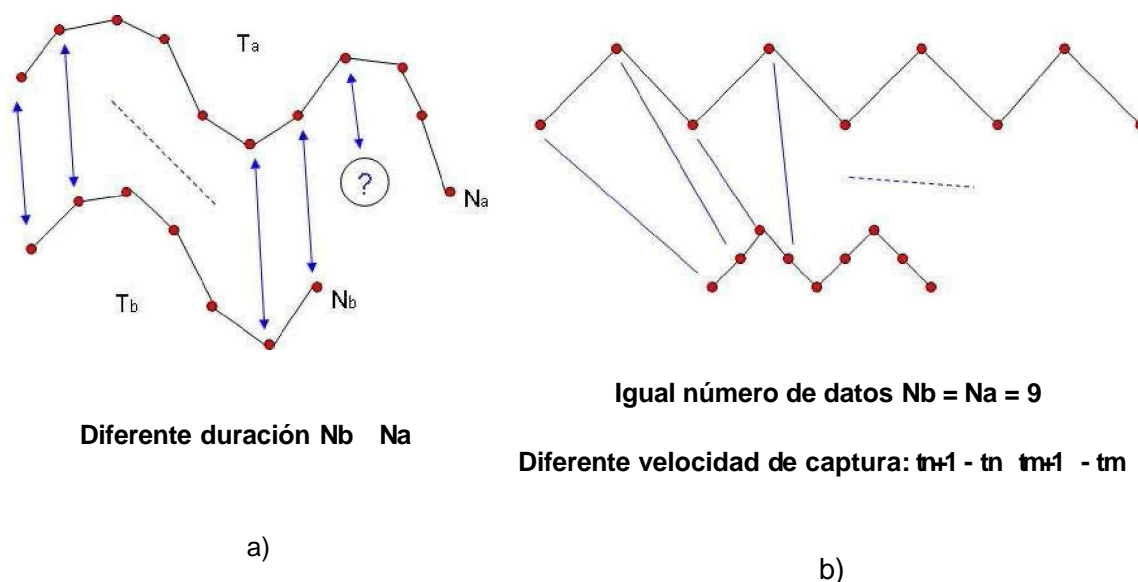


Figura 3.4: Ejemplos de casos ambiguos de las métricas convencionales: a) Diferente duración, b) Igual duración pero diferente velocidad de captura

de la similitud. Sin embargo, (como se mencionó al inicio del capítulo) existen ciertas limitaciones cuando se tratan de comparar las trayectorias [41]:

- Las trayectorias están limitadas a igual duración. Por *duración* se refiere al número de puntos coordenados que constituyen la trayectoria (ver Figura 3.4a).
- Las coordenadas de las trayectorias son tomadas en diferentes instantes de tiempo.
- Si se normaliza o parametriza la duración de la trayectoria para que pueda ser comparada, esta normalización destruye las propiedades temporales de la trayectoria.
- La velocidad de captura de los datos de seguimiento deben ser iguales (ver Figura 3.4b). Este es un problema común especialmente en el seguimiento de objetos en *tiempo real* que procesan el flujo de datos de las cámaras de video.

Por lo tanto, es necesario utilizar una técnica que pueda medir efectivamente las diferencias entre trayectorias sin restricciones. Una alternativa es utilizar los MOM que, como se verá más adelante, pueden capturar las propiedades dinámicas de las trayectorias. Existen varias técnicas para obtener una métrica que pueda medir la similitud entre trayectorias basadas en los MOM, a continuación se presentan las más relevantes. También se presentan algunas métricas convencionales utilizadas en la comparación de trayectorias con las limitaciones que se acaban de mencionar.

## 3.2. Métricas para la comparación de trayectorias

### 3.2.1. Métricas basadas en trayectorias

Una trayectoria, en el contexto del seguimiento de objetos, es una secuencia de coordenadas en el tiempo, representadas por el movimiento de un objeto sobre su duración (tiempo de vida), esto es, el número de cuadros de video que el objeto existe. Estas coordenadas corresponden a las posiciones del objeto en cuadros consecutivos. Las posiciones indican generalmente, el centro de masa del objeto (en coordenadas cartesianas), los ángulos en líneas de intersección para objetos articulados (en coordenadas polares o esféricas), o cualquier otra representación que determine la posición de un objeto en el espacio. Por lo tanto, es posible ver la trayectoria como una colección de valores que puedan abstraer la forma del objeto. En notación, una trayectoria se puede determinar como [41]:

$$T = \{(v_1, t_1), (v_2, t_2), \dots, (v_N, t_N)\} \quad (3.33)$$

donde  $v_i$  es un vector que representa la posición del objeto,  $t_i$  es el índice del tiempo, y  $N$  es la duración de la trayectoria.

Existen métricas sencillas basadas en técnicas estadísticas para comparar dos trayectorias de igual duración. La métrica más simple para calcular la distancia entre dos trayectorias  $T_1$  y  $T_2$  es la distancia media de las coordenadas, definida como [41]:

$$D_m(T_1, T_2) = \quad (3.34)$$

donde  $d_n^2$  indica el desplazamiento entre las posiciones, definido como:

$$d_n^2 = [(v_{1n}^1 - v_{1n}^2)^2 - (v_{2n}^1 - v_{2n}^2)^2 - \dots - (v_{mn}^1 - v_{mn}^2)^2]^{\frac{1}{2}} \quad (3.35)$$

donde  $V$  es el vector que representa la posición del objeto en la trayectoria  $T_i$ . Se puede notar que la métrica de la distancia media hace tres suposiciones críticas:

- La duración de ambas trayectorias son iguales:  $N^1 = N^2 = N$ .
- Las coordenadas están sincronizadas:  $t_n^1 = t_n^2$ , esto es, las posiciones en ambas trayectorias son tomadas en el mismo instante de tiempo.
- La velocidad de captura es constante.

Por consiguiente, es evidente que la media de las distancias es muy sensible a las correspondencias parciales, y no toma en cuenta las distorsiones en el tiempo.

Para proveer mayor información descriptiva, se puede incorporar estadísticas de segundo orden como la distancia basadas en la mediana, la varianza, el mínimo o el máximo [41]. La *varianza* está definida como:

$$D_v(T_1, T_2) = \frac{1}{N} \sum_{n=1}^N (d_n - D_m(T_1, T_2))^2 \quad (3.36)$$

donde  $d_n$  es el desplazamiento en la ecuación 3.35, y  $D_m(T_1, T_2)$  es la media de la distancia en la ecuación 3.34.



### 3.2. Métricas para la comparación de trayectorias

Para calcular la *mediana*, los desplazamientos son ordenados con respecto a sus magnitudes como  $d_n \rightarrow d_m$ , para que el valor del componente intermedio de la lista sea asignado:

$$D_{med} = \begin{cases} d_{\frac{N+1}{2}} & N \text{ impar} \\ \frac{1}{2}(d_{\frac{N}{2}} + d_{\frac{N+1}{2}}) & N \text{ par} \end{cases} \quad (3.37)$$

El *mínimo* y el *máximo* de dos trayectorias se definen como [41]:

$$D_{min}(T_1, T_2) = \text{mínd}_n \quad (3.38)$$

$$D_{max}(T_1, T_2) = \text{máxd}_n \quad (3.39)$$

Si bien, estas medidas estadísticas proveen información extra, heredan el defecto de la métrica ordinaria basada en la media de la distancia  $D_m(T_1, T_2)$  de la ecuación 3.34. Además, ninguna de las métricas es suficiente para hacer una evaluación precisa de la similitud [41].

#### 3.2.2. Métricas basadas en MOM

En general, las métricas basadas en MOM tratan de proyectar la trayectoria  $T$  (o un conjunto de trayectorias), dentro del espacio del modelo  $X$ , que es caracterizado por los parámetros del MOM, esto es,  $X = \{A, B, P\}$ . De esta forma, se pretende reemplazar la información de la trayectoria como los símbolos de observación, y la matriz de transición de estados captura las propiedades transitivas de la trayectoria [41]. Por lo tanto, para comparar dos trayectorias, se necesita comparar los modelos  $\lambda_1$  y  $\lambda_2$  que representan a ambas trayectorias, respectivamente. En un caso más general, se puede entrenar un modelo  $X$  que represente a una clase de trayectorias, de esta forma, se puede determinar la similitud de una nueva trayectoria con respecto al modelo entrenado. En las siguientes secciones se presentan algunas métricas para comparar pares de MOM, que pueden ser utilizadas como una alternativa en la comparación de trayectorias.

#### Distancia de Levinson o distancia euclidiana

La medida para comparar pares de MOM que aparece primero en la literatura fue propuesta por Levinson [35], para MOM's con observaciones discretas. La distancia, que es una *distancia Euclidiana* sobre la matriz de las probabilidades de observación de estados, es definida como:

$$D_e(\lambda_1, \lambda_2) = \sqrt{\frac{1}{MN} \sum_{j=1}^N \sum_{k=1}^M \|b_{jk}^1 - b_{jk}^2\|^2} \quad (3.40)$$

donde  $b_{jk}^i$  es la matriz de probabilidad de observación de estados en el modelo  $X$ . La métrica de la ecuación 3.40 es llamada *la medida para la estimación del error* en [35] y fué usada para caracterizar la estimación del error presente en el proceso de reestimación. Sin embargo, esta medida de la distancia no toma en cuenta la estructura temporal representada en la *cadena de Markov*, y es fácil encontrar algún par de modelos  $\lambda_1$  y  $\lambda_2$  donde la distancia tienda a cero pero las medidas de probabilidad generadas  $P(O_1|\lambda_1)$  y  $P(O_2|\lambda_2)$  sean completamente diferentes [18].

### 3.2. Métricas para la comparación de trayectorias

#### Distancia de Kullback-Leibler

Juang y Rabiner [31], proponen una medida de la distancia para comparar pares de MOM. Esta medida está basada en la distancia de Kullback-Leibler [34] entre dos funciones de distribución de probabilidad, que sigue los conceptos de la divergencia, entropía cruzada, o información de discriminación. La medida de la distancia entre dos MOM's  $\lambda_1$  y  $\lambda_2$  esta definida como:

$$D_k(\lambda_1, \lambda_2) = \frac{1}{T} [\log(P[O_T^1|\lambda_1]) - \log(P[O_T^1|\lambda_2])] \quad (3.41)$$

donde  $O_T^1$  es la secuencia de observaciones generada por el modelo  $\lambda_1$  y  $T$  es la longitud de la secuencia de observaciones. Esta medida se puede interpretar como qué tan bien el modelo  $\lambda_2$  coincide con las observaciones generadas por el modelo  $\lambda_1$ , relativo a que tan bien el modelo  $\lambda_1$  coincide con las observaciones generadas por si mismo [43]. Se puede notar que la medida de la distancia de la ecuación 3.41 *no es simétrica*. Una extensión de esta medida es su versión *simétrica*, definida como [31]:

$$D_{sk}(\lambda_1, \lambda_2) = \frac{D(\lambda_1, \lambda_2) + D(\lambda_2, \lambda_1)}{2} \quad (3.42)$$

que es el promedio de las dos distancias *no-simétricas*. La medida  $D_s(\lambda_1, \lambda_2)$  es simétrica con respecto a  $\lambda_1$  y  $\lambda_2$  y representa la medida de la dificultad de discriminar  $P(O|\lambda_1)$  y  $P(O|\lambda_2)$ .

Un problema en la medida de la distancia de la ecuación 3.41 es su convergencia. En [31] se realizan una serie de pruebas experimentales y concluyen que la distancia converge en medida que el tamaño de la secuencia de observaciones  $T$  incrementa. En general, la medida  $D(\lambda_1, \lambda_2)$  es sensible a los parámetros del modelo  $\lambda$ :  $M$  (el número de símbolos de observación por estado) y  $N$  (el número de estados).

#### Distancia de cruce

Porikli [41] propone una medida de la distancia para comparar trayectorias de un objeto basadas en la representación de MOM. La información de la trayectoria se parametriza en un modelo  $\lambda$ , donde los símbolos de observación sustituyen las posiciones del objeto, y la matriz de transición de estados captura las propiedades dinámicas de la trayectoria. La medida, llamada *distancia de cruce*, esta definida como:

$$D_c(T_1, T_2) = |L(T_1; \lambda_1) + L(T_2; \lambda_2) - L(T_1; \lambda_2) - L(T_2; \lambda_1)| \quad (3.43)$$

que corresponde al cruce de las trayectorias con el otro modelo, como se ilustra en la figura 3.5. Donde  $T_i$  es la secuencia de observación (trayectoria) y  $L(T_i; \lambda_i) = P(O_i|\lambda_i)$  es la probabilidad de la secuencia de observaciones dado el modelo. Los términos  $L(T_1; \lambda_1)$  y  $L(T_2; \lambda_2)$  indican la similitud de las trayectorias de pertenecer a su propio modelo. Los términos de cruce  $L(T_1; \lambda_2)$ ,  $L(T_2; \lambda_1)$  revelan la similitud de la trayectoria generada por los otros modelos de trayectorias. En otras palabras, si dos trayectorias son idénticas, los términos de cruce tendrán un valor máximo, así la ecuación 3.43 será igual a cero. Por otro lado, si las trayectorias son diferentes, su similitud de ser generadas por el otro modelo será pequeña, esto es, la distancia será grande.

La medida de la distancia de la ecuación 3.43 está pensada para modelar una sola trayectoria en un modelo  $X$ . Por lo tanto, para comparar dos trayectorias, la información de cada trayectoria es

### 3.3. Resumen

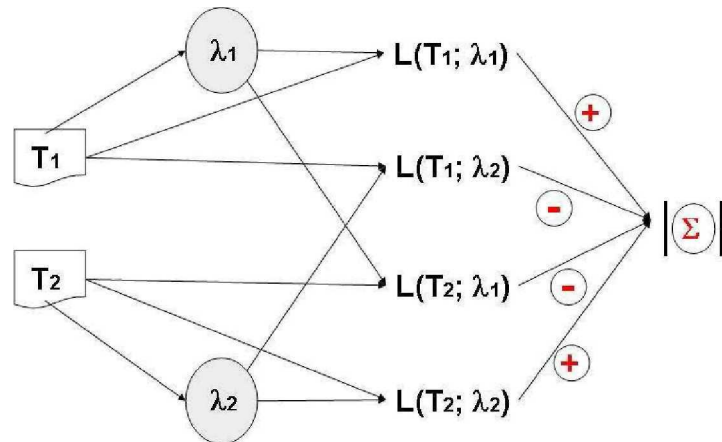


Figura 3.5: Imagen que muestra la *distancia de cruce* entre dos MOM's. Las trayectorias  $T_1$  y  $T_2$  son parametrizadas en los modelos  $\lambda_1$  y  $\lambda_2$ , respectivamente.

utilizada para entrenar el modelo. Así, se tendría un modelo por cada trayectoria que se pretenda comparar. Además, se utilizan MOM de densidad continua, esto es, se utilizan *mezclas de gaussianas* para modelar los símbolos de observación en el modelo.

### 3.3. Resumen

En este capítulo se presentaron los Modelos Ocultos de Markov (MOM), así como algunas técnicas para la comparación de trayectorias basada en MOM. Los MOM son una herramienta útil para modelar series de datos en el tiempo. Un MOM está compuesto de dos procesos estocásticos: un proceso de Markov de primer orden que determina la transición entre los estados y no es observable, y un proceso que genera observaciones de salida para cada estado. Existen tres problemas básicos de interés que pueden ser resueltos para que un MOM se útil en aplicaciones del mundo real: a) evaluación, que determina la probabilidad de la secuencia de observaciones dado el modelo, b) decodificación, que consiste en encontrar la secuencia de estados más probable que haya generado una secuencia de observaciones dada, y c) entrenamiento, que consiste en ajustar los parámetros del MOM para maximizar la probabilidad dada una secuencia de observaciones y el modelo. Un MOM se puede ver como un grafo dirigido, donde los nodos representan los estados, y las flechas representan las transiciones entre los estados. Las topologías más comunes son: ergódicas o completamente conectadas, izquierda-derecha o hacia adelante o Bakis, e izquierda-derecha paralelas. La topología depende en gran medida de las características del problema a modelar.

Como se puede apreciar, los MOM pueden representar la dinámica de un proceso, por lo tanto, se pueden utilizar como una alternativa en la comparación de trayectorias, donde la información de movimiento, que puede ser la posición de un objeto, se utiliza como las secuencias de observación de estados, y la matriz de transición de estados captura las propiedades dinámicas de la trayectoria. Utilizar MOM para representar trayectorias es una buena elección, debido a que permite una representación apropiada a secuencias de datos espacio-temporales. Además, existen diversos problemas con las técnicas convencionales para comparar trayectorias, debido a que suponen que las trayectorias

### 3.3. Resumen

son de igual duración, que la velocidad de captura de los datos es constante, y los datos son tomados en el mismo instante de tiempo. Esto es un problema en aplicaciones de seguimiento de objetos en tiempo real, donde las cámaras de video difieren en el proceso de captura. La distancia de Levinson se aplica sobre la matriz de las probabilidades de observación de estados, sin embargo, no considera todos los parámetros del modelo, por lo que se puede encontrar un par de MOM's, donde la distancia sea cercana a cero, pero sus probabilidades dada una secuencia de observación, sean completamente diferentes. La distancia de Kullback-Leibler, es una buena aproximación para comparar dos MOM, ya que considera todos los parámetros de un MOM; está basada en los conceptos de la divergencia, entropía cruzada, o información de discriminación. Un problema de esta distancia es la convergencia, (esto es, la distancia se mantiene estable) debido a que depende de la longitud de la secuencia de observaciones generada por el modelo. La distancia de cruce esta diseñada especialmente para comparar trayectorias de movimiento de objetos, donde la información de la trayectoria se reemplaza como los símbolos de observación del modelo, y los estados capturan las propiedades espacio-temporales de la trayectoria. Sin embargo, se requiere un modelo por cada trayectoria individual, lo que hace complicado modelar varias trayectorias de una misma clase. El seguimiento visual del brazo y el análisis de gestos se explican en los siguientes capítulos.

# Capítulo 4

## Sistema de seguimiento visual

En este capítulo se van a describir las fases empleadas en el sistema de seguimiento visual aplicado a los dos escenarios: a) T-WREX y, b) Análisis de gestos. En la sección 4.1 se muestra la introducción del capítulo, describiendo brevemente las fases en que se conforma el sistema de seguimiento visual. En la sección 4.2 se describe el proceso de seguimiento visual de la mano al interactuar con la herramienta T-WREX. En la sección 4.3 se describe el proceso de seguimiento visual de las marcas para el análisis de gestos. Finalmente, en la sección 4.4 se presenta el resumen del capítulo.

### 4.1. Introducción

El sistema de seguimiento visual se emplea en dos escenarios diferentes. En el primer escenario, el sistema de seguimiento visual se enfoca en la localización y seguimiento de la mano del paciente, obteniendo las coordenadas en el espacio 3D para interactuar con la herramienta T-WREX [44-46]. El segundo escenario se emplea para el análisis de gestos, en el cual se localizan y siguen tres marcas de color adheridas a las articulaciones del brazo del paciente para ser representado en un modelo en estructura de alambre en el espacio 3D. El sistema de seguimiento visual se puede representar como un diagrama de bloques secuencial mostrado en la figura 4.1. El sistema está dividido en cuatro módulos principales:

- **Calibración.** Este módulo se encarga del calibrado de las cámaras. Se obtienen los parámetros *intrínsecos* y *extrínsecos* que son utilizados para obtener la posición 3D de la mano y de las marcas de color.
- **Segmentación.** Este módulo se encarga de la localización del objeto (mano y marcas de color) dentro de una imagen. En el proceso de segmentación de la mano se aplica información de color y movimiento, mientras que en la segmentación de las marcas se emplean histogramas de color.
- **Seguimiento 2D.** Este módulo tiene como objetivo el seguimiento de un objeto a través de una secuencia de imágenes de video. El seguimiento se aplica en las secuencias de imágenes de dos cámaras de video.
- **Reconstrucción 3D.** Este módulo es el encargado de obtener las coordenadas 3D del objeto

## 4.2. Seguimiento visual de la mano: T-WREX

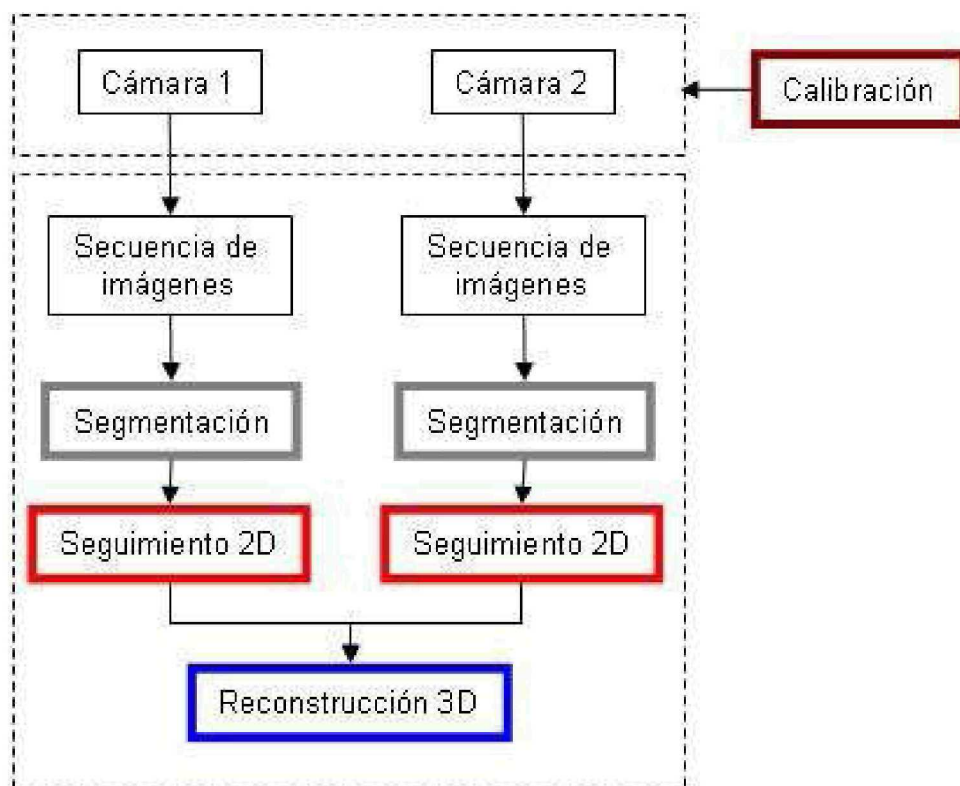


Figura 4.1: Diagrama de bloques del sistema de seguimiento visual. Los recuadros de color muestran los módulos en que se conforma el seguimiento: Calibración, Segmentación, Seguimiento 2D y, Reconstrucción 3D

en seguimiento (mano o marcas de color). Se aplica un algoritmo de intersección de líneas de visión de ambas cámaras al objeto para obtener una estimación de su posición 3D.

La arquitectura mostrada en la Figura 4.1 se aplica en los dos escenarios. Mientras que en el primer escenario el seguimiento se aplica a un solo objeto (la mano del paciente), en el segundo escenario se aplica a tres objetos (las marcas de color). La diferencia en el seguimiento visual de la mano y de las marcas es la fase de segmentación. A continuación se describe el seguimiento visual aplicado a los dos escenarios.

## 4.2. Seguimiento visual de la mano: T-WREX

La herramienta T-WREX [44-46] es un sistema desarrollado en la universidad *UC Irvine* de California por David Reinkensmeyer [44], enfocado a la rehabilitación después de un derrame cerebral. El sistema consiste en un brazo robótico que es fijado al brazo del paciente que le permite realizar movimientos asistidos en un espacio 3D. El dispositivo captura las coordenadas en el espacio tridimensional de la mano de un paciente. Dichas coordenadas, son enviadas al software T-WREX

## 4.2. Seguimiento visual de la mano: T-WREX

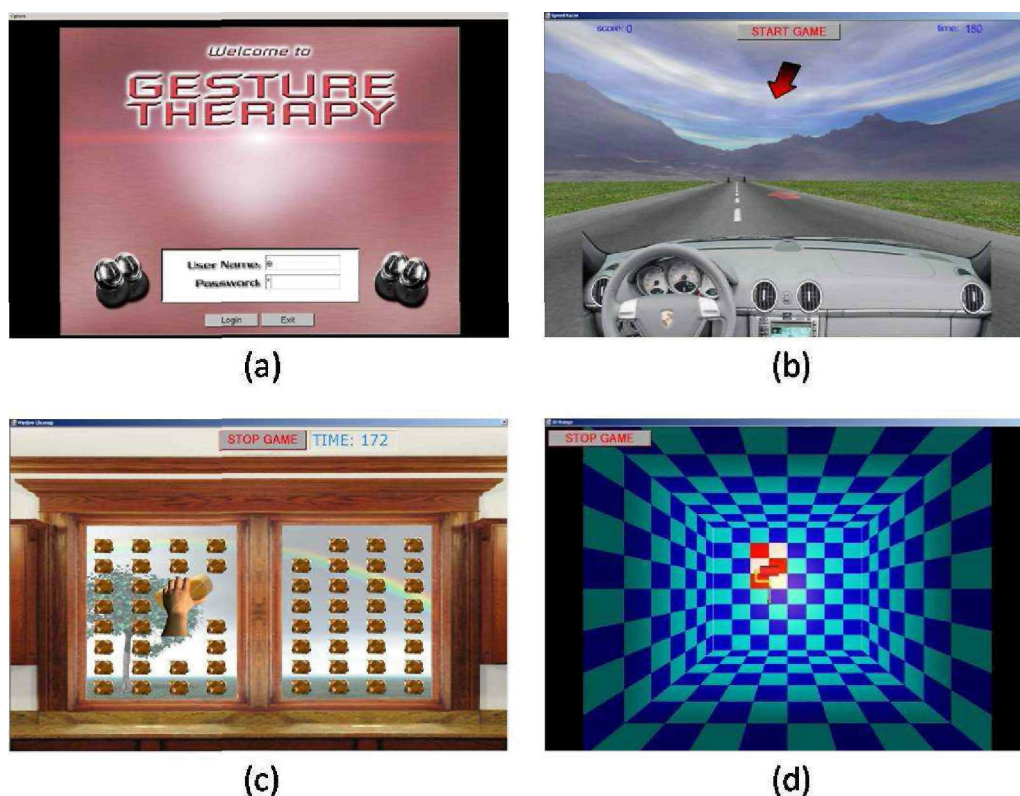


Figura 4.2: Imágenes capturadas del software T-WREX. (a) Pantalla de inicio, (b) - (c) Videojuegos virtuales en un ambiente tridimensional.

para interactuar con un ambiente virtual moviendo su brazo dañado, llevando a cabo diferentes tareas diseñadas para imitar situaciones de la vida real orientadas a la rehabilitación (ver Figura 4.2). Sin embargo, la manufactura del dispositivo robótico tiene un alto costo<sup>1</sup>, limitando así su accesibilidad. En esta tesis se propone un sistema de visión de bajo costo para el seguimiento de la mano de una persona, sustituyendo así el uso del dispositivo robótico. El sistema de seguimiento visual emplea un método que detecta la presencia de la mano del paciente en una imagen mediante el aprendizaje del color de la piel. El algoritmo utilizado se basa en un clasificador basado en la regla de Bayes, que distingue regiones de piel de una persona de las que no lo son. El seguimiento se aplica en dos cámaras posicionadas perpendicularmente. Mediante un algoritmo de intersección de líneas de visión de las cámaras a la mano del usuario, se estima la posición en el espacio tridimensional. A continuación se presenta una explicación de cada módulo en el proceso de seguimiento visual de la mano (ver Fig. 4.1).

### 4.2.1. Calibración de las cámaras

Para obtener una estimación de la posición de la mano en el espacio 3D, un paso previo al seguimiento visual es la calibración de las cámaras de video. La calibración de la cámara es un proceso que relaciona un modelo ideal con el dispositivo físico y determina la posición y orientación de ésta con respecto al sistema de referencia del mundo [25].

Se necesita conocer el modelo de proyección, así como la posición y orientación de cada cámara-

<sup>1</sup>\$4000 USD en el año 2007

## 4.2. Seguimiento visual de la mano: T-WREX

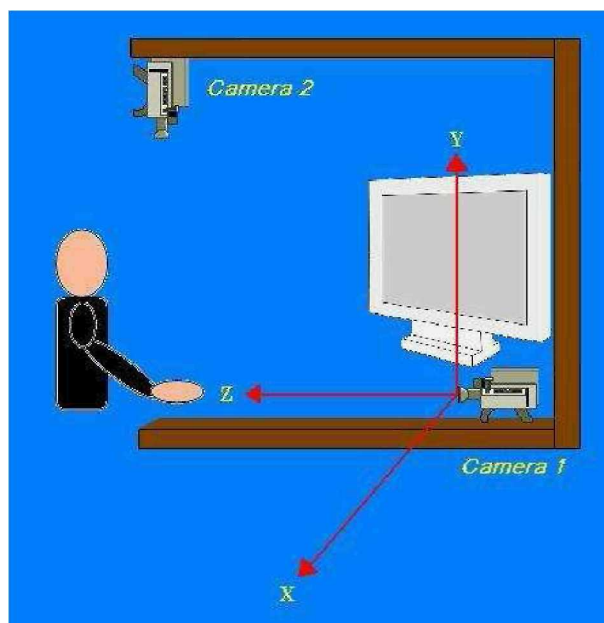


Figura 4.3: La imagen muestra la vista lateral del sistema de seguimiento visual de la mano, así como el sistema de referencia del mundo.

ra en el sistema de referencia del mundo. Estos dos grupos son los llamados parámetros *intrínsecos* y *extrínsecos* de la cámara [40]. La figura 4.3 muestra el sistema de referencia del mundo y la colocación de las cámaras. Las cámaras se colocan en forma perpendicular, a una distancia determinada. La cámara 1, esta colocada en forma cenital, la cámara 2 se encuentra posicionada frontalmente al usuario.

El modelo utilizado es el modelo *punta de alfiler* o *pin-hole* (ver Figura 4.4), que está basado en el principio de colinealidad, donde cada punto en el espacio del objeto es proyectado en una línea recta a través del centro óptico (centro de proyección) del plano de la imagen [25].

La relación entre un punto  $M$  en 3D y su proyección en la imagen  $m$  esta dada por la fórmula:

$$m = ARtM \quad (4.1)$$

donde  $A$  es la matriz intrínseca de la cámara, definida como:

$$A = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (4.2)$$

donde  $(c_x, c_y)$  son las coordenadas del centro óptico.  $(f_x, f_y)$  son las longitudes focales sobre los ejes  $(x, y)$ . Los parámetros intrínsecos describen la geometría interna y características ópticas de los lentes y del dispositivo de proyección de la imagen. La longitud focal es la distancia entre el lente y el plano de la imagen. El centro óptico es el centro de la imagen en coordenadas de píxeles.

$(R, t)$  son los parámetros extrínsecos que describen la posición y orientación de la cámara en el sistema de referencia del mundo. Relacionan el sistema de coordenadas del mundo con el sistema



## 4.2. Seguimiento visual de la mano: T-WREX

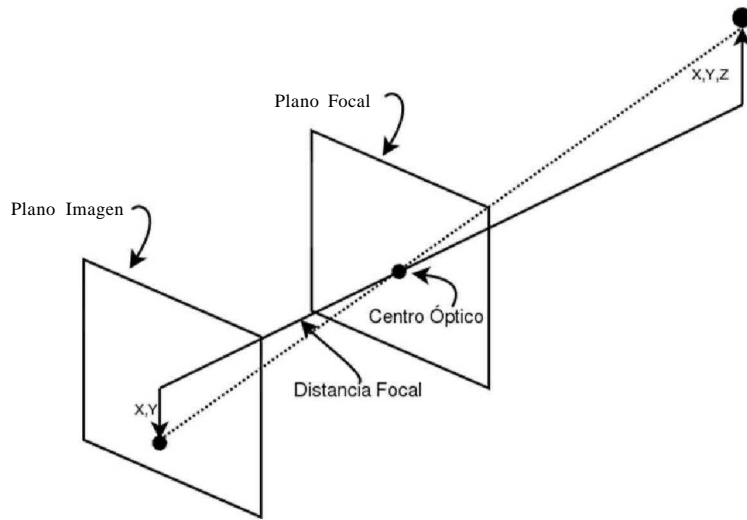


Figura 4.4: La imagen muestra el modelo de proyección *punta de alfiler*.

de coordenadas de la cámara (ver Fig. 4.5).  $R$  es la matriz de rotación y  $t$  el vector de translación, definidos como:

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}, \quad t = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} \quad (4.3)$$

Un punto  $P^C$  en el marco de referencia de la cámara  $[X^c, Y^c, Z^c]^T$  está relacionado con su correspondiente punto  $P^w$  en el marco de referencia del mundo  $[X^w, Y^w, Z^w]^T$  por:

$$P^C = {}_R P^w - t \quad (4.4)$$

Para obtener los parámetros intrínsecos de las cámaras, se utiliza un patrón de calibrado que consiste en un tablero de ajedrez. El patrón se posiciona frente a la cámara a una distancia determinada, y se detecta las esquinas del tablero. Se utiliza la librería de visión computacional *OpenCV* (*Open Source Computer Vision Library*) desarrollada por *Intel* [36]. Los parámetros extrínsecos se obtienen manualmente midiendo la posición y orientación de las cámaras con respecto al origen del sistema de referencia del mundo.

### 4.2.2. Segmentación

En nuestro caso, la segmentación, es el proceso mediante el cual se calcula qué píxeles de la imagen pertenecen al objeto de interés. El resultado de este proceso es una imagen binaria<sup>2</sup>: un

<sup>2</sup>También se puede sustituir por una imagen en escala de grises que recoge el grado de pertenencia de cada píxel al objeto considerado, de esta forma, se mantiene la mayor cantidad de información posible para la aplicación de algoritmos de seguimiento.

## 4.2. Seguimiento visual de la mano: T-WREX

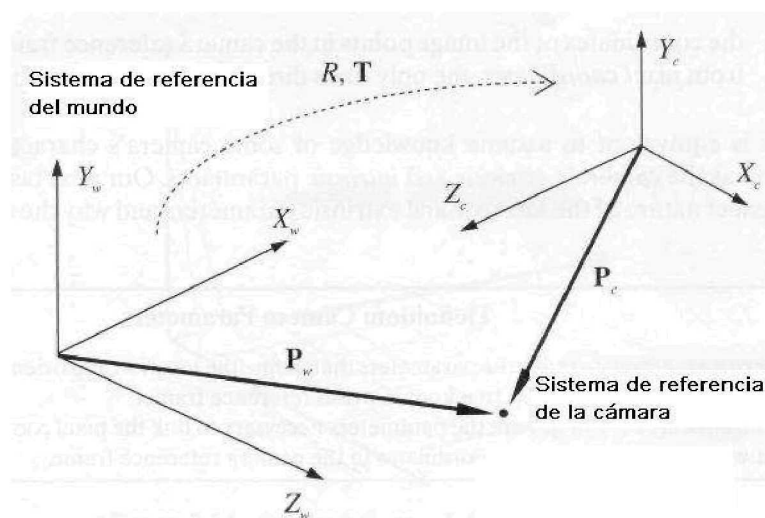


Figura 4.5: La imagen muestra el sistema de referencia de la cámara con respecto al sistema de referencia del mundo. La cámara se encuentra Rotada (R) y Traslada (T) a un nuevo punto en el espacio 3D.

*pixel* tiene valor uno si pertenece al objeto y en otro caso, su valor es cero. La segmentación es una operación en el límite entre el procesamiento de bajo nivel de imágenes y las operaciones que analizan la forma de los objetos. Básicamente se pueden establecer tres clases de segmentación: *a) métodos a nivel de pixel*, que utilizan los valores de color de cada elemento de la imagen, *b) métodos basados en la detección de bordes*, que detectan cambios de intensidad en la frontera de los objetos y extraen los bordes del mismo, y por último, *c) métodos basados en regiones*, que analizan los valores de color en grandes áreas de la imagen. Nosotros utilizaremos el primero.

El primer paso para el seguimiento de la mano consiste en su localización dentro de la imagen. Para alcanzar este objetivo se propone un algoritmo que combina información de color y movimiento. Para la detección de color se aplica un procedimiento para detectar regiones de piel utilizando un clasificador basado en la regla de Bayes. Para la detección de movimiento se aplica un algoritmo de substracción de imágenes consecutivas. El proceso de segmentación utilizado en este trabajo se ilustra en la Figura 4.6, a continuación se presenta una explicación de cada bloque.

### Segmentación basada en color

El color ha sido ampliamente usado en sistemas de seguimiento visual para tareas tales como la segmentación de imágenes, reconocimiento de objetos y seguimiento. Ofrece varias ventajas significativas como son: simplicidad computacional, robustez ante oclusiones parciales, rotación, escalado, y cambios en la resolución de la imagen. Aunque los métodos de segmentación basados en color han probado ser eficientes en una variedad de aplicaciones de visión, existen varios problemas asociados con estos métodos en donde la consistencia del color es uno de los más importantes [33]. Para solventar este problema, es preciso considerar un modelo de color que sea robusto ante los cambios de iluminación. El modelo de color utilizado es el *HSV* [22], el cual permite separar el brillo o luminosidad del color puro, utilizando las componentes *H-S* (croma y saturación), e ignorando la componente *V* del modelo.

Para segmentar la mano del usuario se emplea un clasificador probabilístico basado en una

## 4.2. Seguimiento visual de la mano: T-WREX

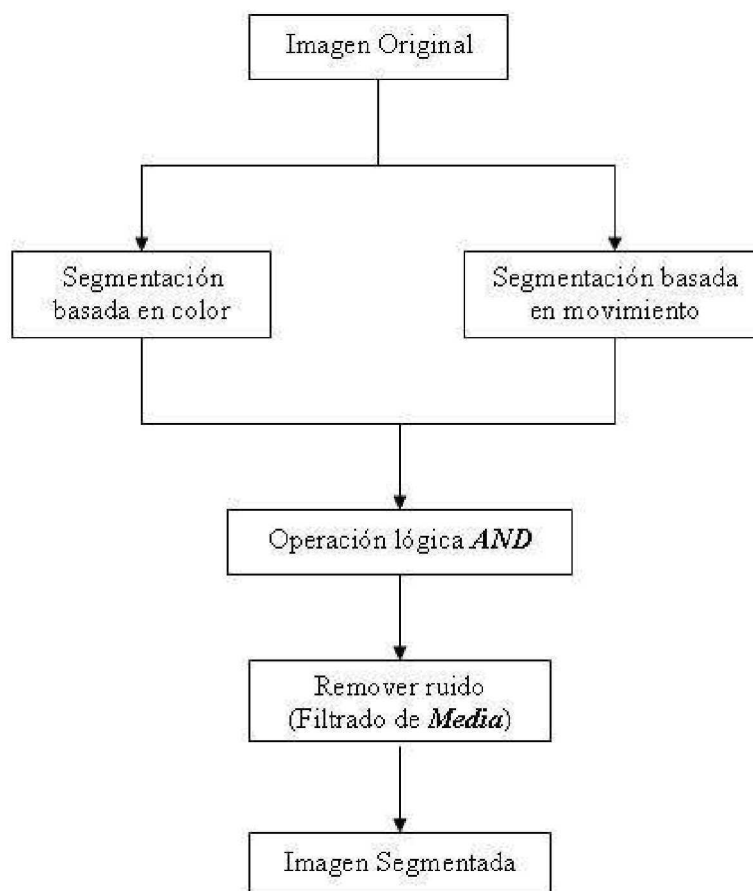


Figura 4.6: Imagen que muestra el proceso de segmentación para la mano de una persona empleando color y movimiento.

modificación de la regla de Bayes propuesta por Jones [30]. El clasificador identifica dos tipos de clases: a) píxeles que pertenecen a la clase *piel* y, b) píxeles que pertenecen a la clase *no-piel*. Para establecer la distribución de probabilidad de las clases, se emplean histogramas de color en el espacio *H-S*. Los histogramas se entrenan con una base de datos conteniendo 3,354,974 píxeles *de piel* y 3,459,358 píxeles *de no-piel*.

Para la construcción de los histogramas, se debe considerar el número de combinaciones por cada canal de color. En el modelo *H-S* se consideran 256 intervalos de clases por cada canal de color, esto es,  $(256)^2 = 65536$  intervalos de clase. El número de combinaciones se puede mejorar al reducir los intervalos de clase del histograma a 32, con lo que queda un total de  $(32)^2 = 1024$  intervalos de clase. Cada intervalo almacena el número de veces que un valor *H-S* de color -o color de pixel- ocurre en una base de datos de  $n$  imágenes utilizadas para el entrenamiento. El histograma se convierte en una distribución discreta de probabilidad  $P(.)$  en la forma usual:

$$P(HS) = \frac{c[HS]}{T_c} \quad (4.5)$$

Para calcular la probabilidad de que un valor de color *HS* pertenezca a la clase *piel* o *no-piel* se aplican las siguientes ecuaciones [30]:

## 4.2. Seguimiento visual de la mano: T-WREX

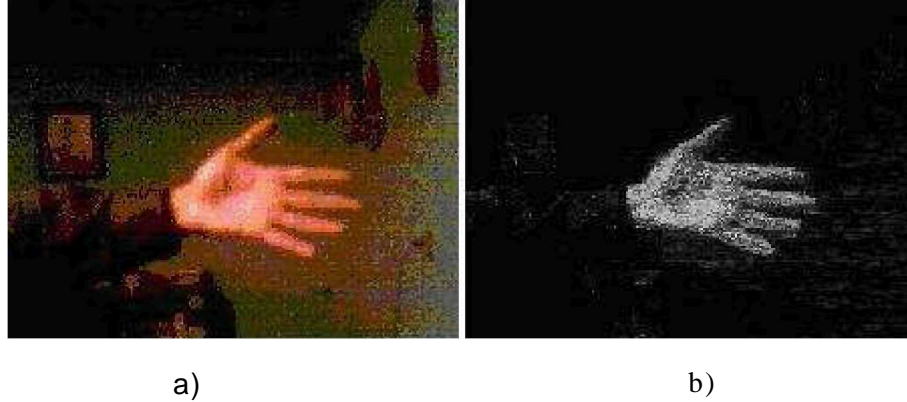


Figura 4.7: La imagen ilustra el proceso de segmentación de píxeles de piel. a) Imagen original, b) Imagen segmentada

$$P(HS|piel) = \frac{s[HS]}{T_s}, \quad P(HS|\neg piel) = \frac{n[HS]}{T_n} \quad (4.6)$$

De esta forma, se aplica el clasificador propuesto por Jones [30] basado en la siguiente regla de Bayes:

$$P(piel|HS) = \frac{P(HS|piel)P(piel)}{P(HS|piel)P(piel) + P(HS|\neg piel)P(\neg piel)} \quad (4.7)$$

donde  $P(piel|HS)$  es la probabilidad de que un píxel sea piel dado que toma un valor  $HS$ ; la probabilidad *a priori*  $P(piel)$  de que cualquier píxel sea piel se define como:

$$P(piel) = \frac{T_s}{T_s + T_n} \quad (4.8)$$

y la probabilidad *a priori*  $P(\neg piel)$  de que cualquier píxel sea no-piel se define como:

$$P(\neg piel) = \frac{T_n}{T_s + T_n} \quad (4.9)$$

De esta forma, podemos transformar una imagen de color a una imagen en escala de grises, tanto que el valor de gris de cada píxel muestre la probabilidad de pertenecer a la clase *piel*. La Figura 4.7 muestra un ejemplo del clasificador de píxeles de piel. La imagen resultante es una imagen en escala de grises, donde la intensidad del gris exhibe el grado de pertenencia del píxel a la clase *piel*. Con este proceso mantenemos la mayor información posible para el algoritmo de seguimiento que será analizado más adelante.

### Segmentación basada en movimiento

La simple substracción de imágenes adquiridas en diferentes instantes de tiempo puede detectar la presencia del movimiento, suponiendo que la cámara está estacionaria y que la iluminación es

## 4.2. Seguimiento visual de la mano: T-WREX

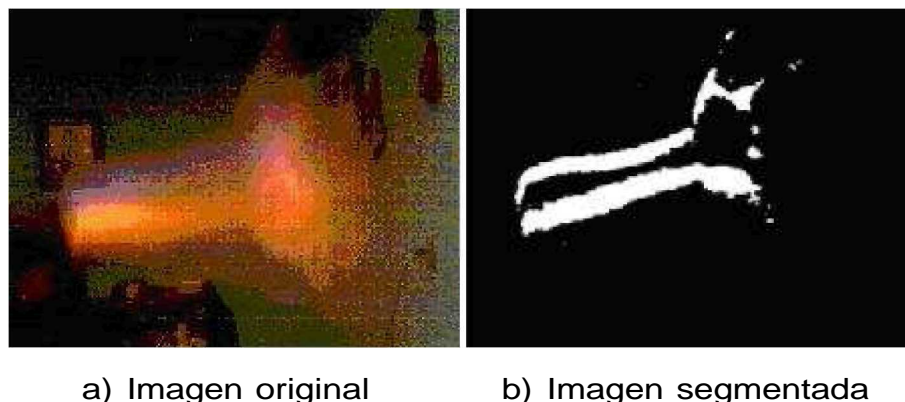


Figura 4.8: La imagen muestra el proceso de segmentación basado en movimiento

constante. Una imagen de diferencias  $d(i,j)$  es una imagen binaria donde los valores de  $I$  representan áreas de la imagen con movimiento, es decir, áreas donde hay una diferencia sustancial entre los niveles de intensidad de las imágenes consecutivas  $f_1(i,j)$  y  $f_2(i,j)$  [21]:

$$d(i,j) = \begin{cases} 0 & \text{si } f_1(i,j,t_1) - f_2(i,j,t_2) \leq \epsilon \\ 1 & \text{de otro modo} \end{cases} \quad (4.10)$$

Donde  $\epsilon$  es un umbral predeterminado. La imagen de diferencias puede basarse en características más complejas de las imágenes, tales como nivel medio de intensidad en alguna vecindad, texturas locales, etc. Es claro que con esta aproximación se puede detectar el movimiento de cualquier objeto distinto de su fondo. La Figura 4.8 muestra el proceso de segmentación del brazo de una persona basado en la substracción de imágenes.

### Segmentación basada en fusión de color y movimiento

El proceso de segmentación basada en la fusión de color y movimiento es el siguiente. La segmentación de la mano se aplica para ambos algoritmos (color y movimiento), después, se toman las imágenes resultantes y, mediante una operación lógica *AND* a nivel de píxel, se segmentan aquellos píxeles que pertenezcan al color de piel y que además presenten movimiento. Posteriormente se aplica un filtrado de suavizamiento basado en la media de la vecindad de un píxel, con el fin de eliminar pequeñas regiones consideradas como ruido. De esta manera se puede determinar la posición de la mano del usuario.

La Figura 4.9 muestra el proceso de segmentación basado en la fusión de color y movimiento. En (a) se muestra la imagen original de la persona en la escena. En (b) se realiza la segmentación basada en color de la piel. En (c) se presenta la detección del movimiento de la mano del usuario. Finalmente en (d) se presenta la imagen al unir las imágenes resultantes de los dos procesos anteriores, por medio de la operación *AND*.

Aplicando el algoritmo de segmentación de color y movimiento se obtiene la posición inicial de la mano. Una vez segmentada la mano, su posición inicial se obtiene por medio de calcular el centroide de la mancha segmentada, por medio de obtener los momentos de orden 0 y 1:

## 4.2. Seguimiento visual de la mano: T-WREX

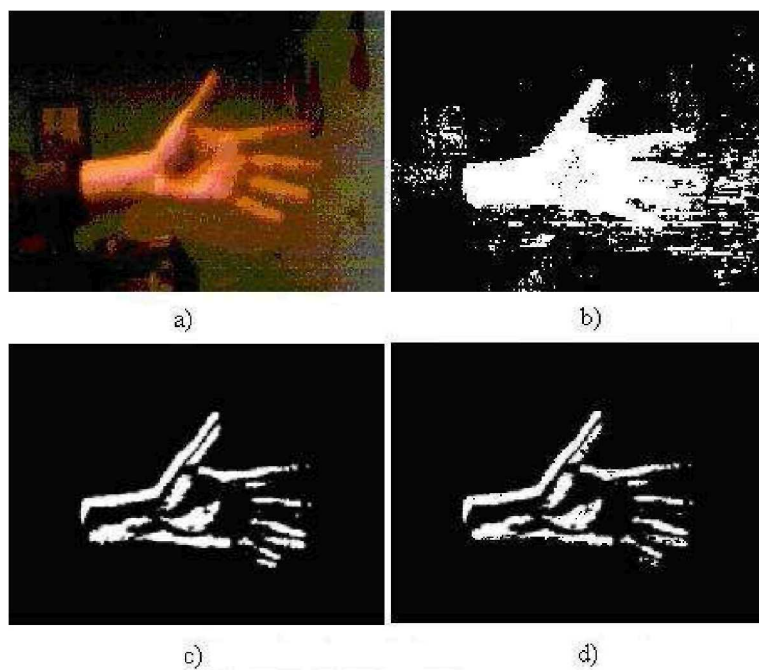


Figura 4.9: Ejemplo de segmentación basado en fusión de color y movimiento, a) Imagen original, b) Segmentación de piel, c) Detección de movimiento y c) Fusión (AND) de color y movimiento

$$\begin{aligned}
 M_{00} &= \sum_x \sum_y I(x,y) & (4.11) \\
 M_{10} &= \sum_x \sum_y xI(x,y) \\
 M_{01} &= \sum_x \sum_y yI(x,y)
 \end{aligned}$$

Donde el momento de orden cero corresponde al área del objeto, y los momentos de orden uno son los momentos para los ejes  $x$  y  $y$  de la imagen entera. Así, los momentos de orden cero y uno determinan el centro de gravedad del objeto (posición inicial de la mano):



Figura 4.10: Localización inicial de la mano, a) Posición inicial, b) Ventana inicial de seguimiento

## 4.2. Seguimiento visual de la mano: T-WREX

$$X_c = \frac{M_{10}}{M_{00}}, \quad Y_c = \frac{M_{01}}{M_{00}} \quad (4.12)$$

Donde  $X_c$  y  $Y_c$  es la posición inicial, sobre la cual se localizará la ventana de tamaño  $N \times M$ . Esta ventana servirá para reducir el cálculo de procesamiento para la segmentación de píxeles de piel, para evitar el cálculo en toda la imagen. El tamaño máximo de la ventana de seguimiento es fijado a un tamaño considerable para abarcar el área de la mano, tomando en cuenta la distancia de la cámara al usuario. La Figura 4.10 muestra el proceso de detección de la mano y la localización inicial de la ventana de seguimiento. El punto azul muestra la ubicación de la mano al aplicarse el método de fusión de color y movimiento. La ventana inicial es fijada después de haber localizado la posición inicial.

### 4.2.3. Seguimiento 2D

Después de la localización de un objeto en una imagen, uno de los problemas más importantes dentro del análisis del movimiento es el seguimiento del objeto a través de una secuencia de imágenes. El problema de seguimiento puede definirse como sigue: dada la localización de un objeto en un tiempo  $t$ , determinar la posición más probable de ese objeto en un tiempo  $t + T$  [16].

Existen dos enfoques básicos para resolver el problema de seguimiento: *basado en modelos* y *basado en características*. El seguimiento *basado en modelos* parte de la idea de aproximar un objeto conocido *a priori* -o modelo- con el objeto a seguir presente en la imagen. La desventaja de este enfoque es su alto costo computacional. El seguimiento *basado en características* se enfoca en detectar alguna característica del objeto a seguir (textura, borde, color, etc), siendo de bajo costo computacional, pero menos robusto ante problemas de ruido como la oclusión o la distracción de objetos con características semejantes. En este trabajo se utiliza el algoritmo *Camshift* [13], que se basa en características. Este algoritmo utiliza el color como característica discriminante, en nuestro caso, el color de la piel. Después de localizar la posición inicial de la mano, *camshift* sólo utiliza la segmentación basada en color para localizar la mano del usuario a través de una secuencia de imágenes, esto con el fin de reducir el cálculo en el proceso de seguimiento.

*Camshift* [13] se basa en ideas de estadística robusta y distribuciones de probabilidad. Los estadísticos robustos son aquellos que tienden a ignorar datos atípicos en la muestra (puntos que caen lejos de la región de interés), por lo tanto ayudan a compensar el ruido y las distracciones en los datos de visión (objetos que no son el seguido). De esta forma podemos encontrar la zona con mayor densidad de puntos pertenecientes a los datos, y dado que los datos son los píxeles que pertenecen a la piel, encontrar la mayor mancha en la imagen del color de la piel y sus dimensiones. El algoritmo *Camshift* trabaja como se muestra en la Figura 4.11:

1. Elegir el tamaño de la ventana de búsqueda. Este tamaño es determinado de forma heurística en relación de la distancia de la mano con la cámara de video.
2. Elegir la posición inicial de la ventana de búsqueda. La posición inicial se determina de acuerdo a lo expuesto en la sección anterior.
3. Desplazar la media (una o varias iteraciones), guardar el momento de orden 0. La media se aplica a una imagen en escala de grises, en nuestro caso, obtenida por medio del algoritmo de



## 4.2. Seguimiento visual de la mano: T-WREX

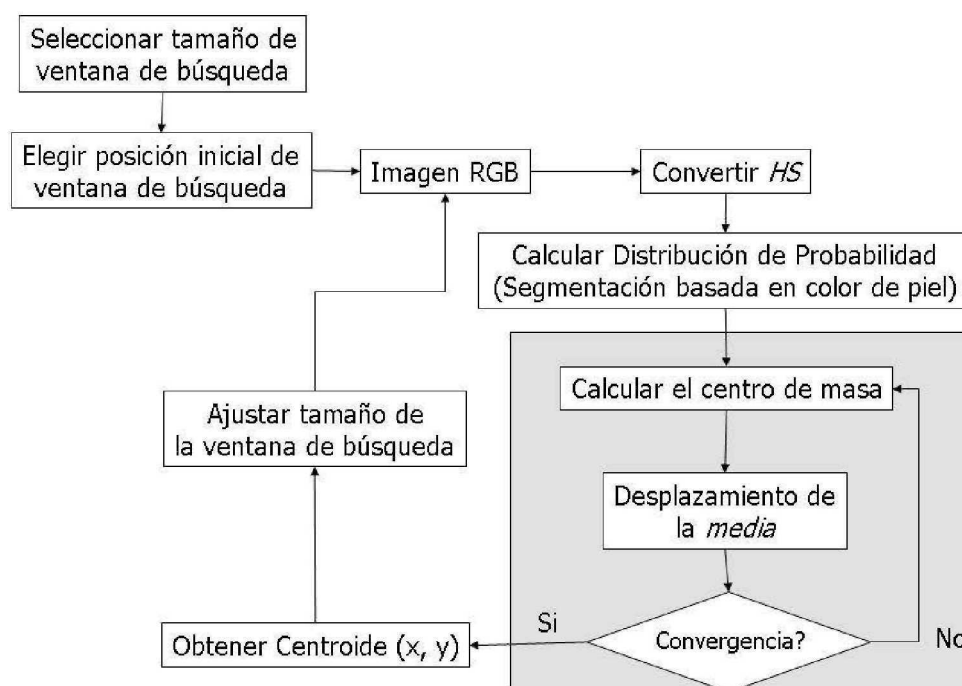


Figura 4.11: Diagrama de flujo del algoritmo Camshift utilizado para el seguimiento de la mano.

segmentación de piel. Los momentos de orden 0 y 1 se obtienen por medio de las fórmulas 4.11 y 4.12 (ver pag. 47).

4. Establecer el tamaño de la ventana de búsqueda igual a una función del momento de orden 0 encontrado en el paso 3.
5. Repetir los pasos 3 a 4 hasta la convergencia (la posición de la media se mueve menos que un umbral predeterminado). En este caso, el número de iteraciones no sobrepase a 10, o en caso contrario, el desplazamiento en el eje x y y, no exceda el tamaño de la ventana de búsqueda inicial.

En la figura 4.12 se muestra una secuencia de tres imágenes en donde se aplica el algoritmo de seguimiento de la mano de una persona.



Figura 4.12: Secuencia de tres imágenes que muestran el seguimiento de la mano aplicando el algoritmo Camshift



## 4.2. Seguimiento visual de la mano: T-WREX

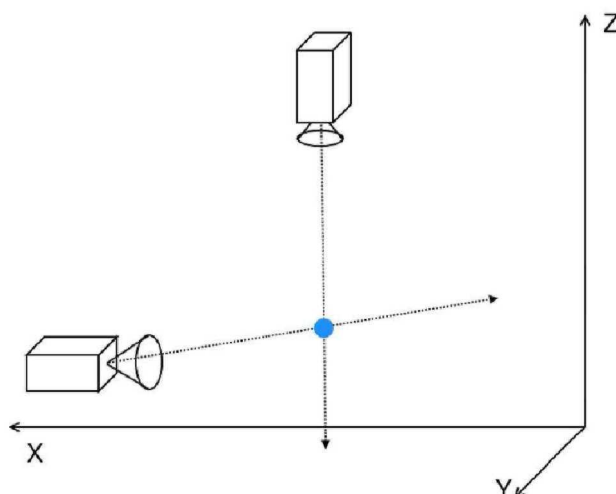


Figura 4.13: Cruce de las líneas de visión en el espacio tridimensional.

### 4.2.4. Reconstrucción 3D

Después de obtener la posición en el espacio 2D de la mano en ambas cámaras, el siguiente paso es estimar su posición en el espacio 3D. Casi todos los enfoques para el seguimiento del cuerpo humano (o partes del cuerpo) en el espacio 3D se basan en enfoques *basados en modelos*. Si bien, estas técnicas ofrecen una precisión en la reconstrucción del cuerpo humano y permiten modelar su dinámica, requieren de un alto costo computacional. Una alternativa es estimar las posiciones de los objetos (marcas) en 3D, usando el enfoque basado en *múltiples cámaras* [1], donde cada cámara obtiene la posición 2D del objeto en el plano de la imagen, y los combina con los datos de las demás cámaras.

Los sistemas de visión basados en una única cámara adolecen de indeterminación en cuanto a la distancia a la que se encuentran los objetos detectados. Dicha ambigüedad de profundidad se puede reducir añadiendo información al sistema sobre el tamaño de los objetos. Con ello se puede realizar una estimación de la distancia a la que se encuentran partiendo de su tamaño en la imagen. Otra posible forma de obtener la profundidad es emplear más de una cámara para observar la misma escena. En este caso no es necesario introducir información externa al sistema acerca del tamaño de los objetos. Al tener dos cámaras se puede realizar un proceso de triangulación detectando en ambas imágenes algún punto de interés de los objetos buscados, en nuestro caso, el centroide de la ventana de seguimiento de la mano.

El método para estimar la posición 3D de la mano de una persona esta basado en el método propuesto por González y Cañas [23], el cual consiste en proyectar las líneas de visión de dos cámaras al objeto buscado, según un modelo geométrico. La Figura 4.13 ilustra la posición tridimensional del objeto en las que se intersectan las rectas de proyección, o en su defecto su punto de cruce más cercano. El método supone que las cámaras se encuentran calibradas, y conocer su posición y orientación en el sistema de referencia del mundo, lo cual resulta imprescindible para deshacer correctamente la proyección.

Una vez obtenidas las coordenadas 2D de la mano en cada imagen (ver pág. 49), para obtener su posición 3D se reconstruye la línea de visión de cada cámara. Esta línea contiene todos los puntos tridimensionales que se proyectan en esas coordenadas en la imagen.

## 4.2. Seguimiento visual de la mano: T-WREX

Reconstruir la línea de visión buscada consistirá en invertir el proceso de proyección y la rotación/traslación, para esto, es necesario conocer los parámetros *intrínsecos* y *extrínsecos* (ver pág. 42). Para reconstruir la línea de visión es necesario conocer dos puntos por los que pase. El primero será el *centro óptico* de la cámara, que es igual al vector  $t$  de la ecuación 4.3. Para el otro punto tomamos el centro de masa de la mano, obtenido en la fase de seguimiento 2D ( $w, v$ ). Esto es, tomando en cuenta la relación que existe entre las coordenadas de la imagen en pixeles, y las espaciales:

$$\frac{1}{Z} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = A \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (4.13)$$

donde  $A$  es la matriz que define los parámetros intrínsecos, tomada de la ecuación 4.2. Queda la expresión:

$$\begin{aligned} u &= \frac{X}{Z} f_x + c_x \\ v &= \frac{Y}{Z} f_y + c_y \end{aligned} \quad (4.14)$$

Despejamos  $(u, v)$  (la posición del objeto), y para simplificar puede tomarse  $z = 1$ . De este modo se obtiene la expresión:

$$\begin{aligned} x &= \frac{(u - c_x)}{f_x} \\ y &= \frac{(v - c_y)}{f_y} \\ z &= 1 \end{aligned} \quad (4.15)$$

Este nuevo punto define la posición del objeto visto desde la cámara ( $P^C$ ), por lo que se debe considerar la nueva posición y orientación de la cámara. Tomando los *parámetros extrínsecos* de la ecuación 4.3, y la ecuación 4.4, que relaciona un punto en el marco de referencia de la cámara con respecto al marco de referencia del mundo, el nuevo punto 3D del objeto queda definido por la siguiente expresión:

$$P^W = P^C R^{-1} + t \quad (4.16)$$

Una vez calculado las rectas de proyección de cada cámara con respecto al objeto, el corte de las dos líneas de visión, o en su defecto, el punto más cercano de cruce [12], marcará la estimación de la posición del objeto en tres dimensiones. De esta forma se obtiene las coordenadas  $(X, Y, Z)$  de la mano, que son enviadas al software *T-WREX* en el cual el paciente puede interactuar con el ambiente virtual.

### 4.3. Seguimiento visual de las marcas: Análisis de gestos

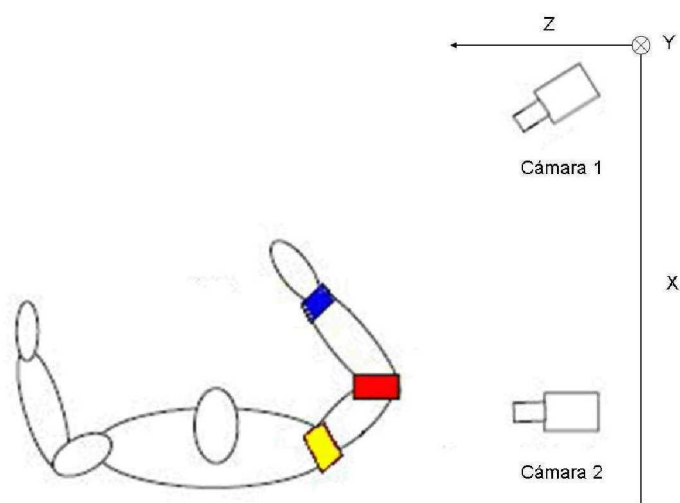


Figura 4.14: La imagen muestra la vista superior del sistema de seguimiento y el sistema de referencia del mundo. Las marcas de color identifican las articulaciones. Amarillo: hombro, rojo: codo y azul: muñeca

### 4.3. Seguimiento visual de las marcas: Análisis de gestos

El escenario de *Análisis de gestos* tiene como objetivo determinar el grado de avance del paciente al ejecutar un gesto predefinido. En este trabajo, se propone un sistema de seguimiento visual que permite el seguimiento del brazo con suficiente precisión y ejecución en tiempo real. Para facilitar la localización del brazo, se colocan marcas en las articulaciones que lo conforman, las cuales son: hombro, codo y muñeca. Las marcas son cintas de un color específico cada una. La marca de color amarillo representa el hombro, la marca de color rojo representa el codo, y la marca de color azul representa la muñeca del brazo (ver Figura 4.14).

El sistema detecta y sigue las marcas en cada imagen capturadas por las dos cámaras de video. Posteriormente, se estima la posición 3D de cada marca y se construye una figura de alambre que representa el brazo del paciente en el espacio 3D, lo cual permite la extracción de características de movimiento. A continuación se presenta el conjunto de gestos considerados en el análisis de gestos, así como las fases en el seguimiento visual de las marcas.

#### 4.3.1. Determinación de los gestos

El término *gesto* se puede interpretar como un acto visual que envía una señal a un espectador [3]. Los gestos se pueden clasificar en dos clases debido a la inclusión de movimiento: *gestos estáticos* y *gestos dinámicos*. Los gestos estáticos solo requieren de alguna postura del cuerpo. Por otro lado, los gestos dinámicos son aquellos que involucran movimiento en su realización. Aunque el objetivo principal del gesto es la comunicación de información, algunas personas o sistemas inteligentes pueden observar el gesto y analizarlo con la finalidad de interpretar su correcta ejecución. En esta tesis, el gesto es considerado como los movimientos del brazo del paciente en el proceso de rehabilitación, que son tomados de las sesiones de fisioterapia [17].

### 4.3. Seguimiento visual de las marcas: Análisis de gestos



Figura 4.15: Gesto *deflexión-extensión*

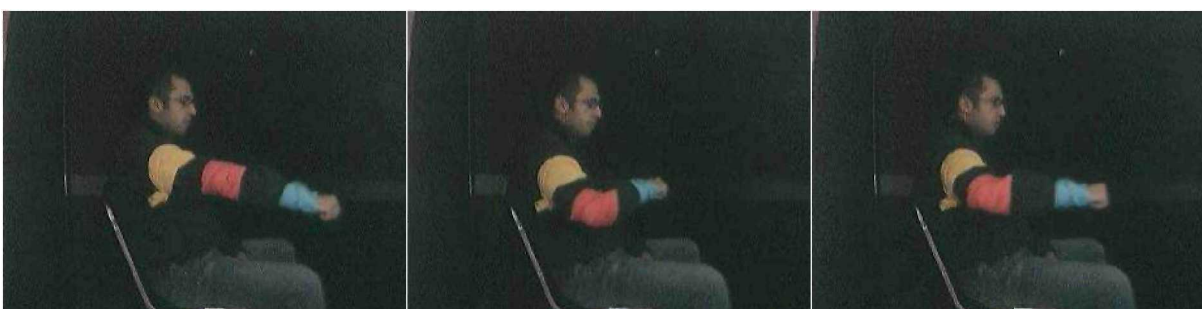


Figura 4.16: Gesto *circular*

En esta tesis se consideran dos gestos para su análisis: a) *flexión-extensión* y b) *circular*. Para realizar el gesto de *flexión-extensión*, el usuario deberá tener inicialmente el brazo extendido y flexionarlo en dirección paralela al torso, y regresarlo a la posición inicial (ver Fig. 4.15). En el gesto *circular*, el usuario tiene el brazo extendido frontalmente con la palma de la mano apuntando hacia abajo, y mueve la mano en forma circular en sentido inverso a las manecillas del reloj, regresando a su posición inicial (ver Fig. 4.16).

#### 4.3.2. Calibración de las cámaras

El proceso de calibrado de las cámaras es el mismo expuesto en la sección 4.2.1. Los parámetros *intrínsecos* se obtienen por medio del patrón de calibrado (tablero de ajedrez). La Figura 4.14 muestra el sistema de referencia del mundo y la colocación de las cámaras. Las cámaras se colocan en forma paralela sobre el eje X, a una distancia determinada. La cámara 1, se encuentra girada con respecto al eje Y, y el paciente se encuentra posicionado frontalmente a ellas.

#### 4.3.3. Segmentación

El método para localizar las marcas de color adheridas al brazo, se basa en el método propuesto por Swain y Ballard, llamado *Retroproyección de Histograma* [59], el cual sugiere la representación de los objetos por medio de histogramas de color, los cuales son almacenados en una base de datos. El objeto que se busca debe estar representado por un histograma  $M$ . También, se obtiene el histograma de la imagen  $I$  que se está analizando, y se calcula un tercer histograma  $R$ , resultado de dividir el histograma del modelo entre el histograma de la imagen. Es este histograma  $R$ , el que se mapea sobre la imagen, esto es, los valores de la imagen son sustituidos por los valores indexados por  $R$ . Así, el

### 4.3. Seguimiento visual de las marcas: Análisis de gestos

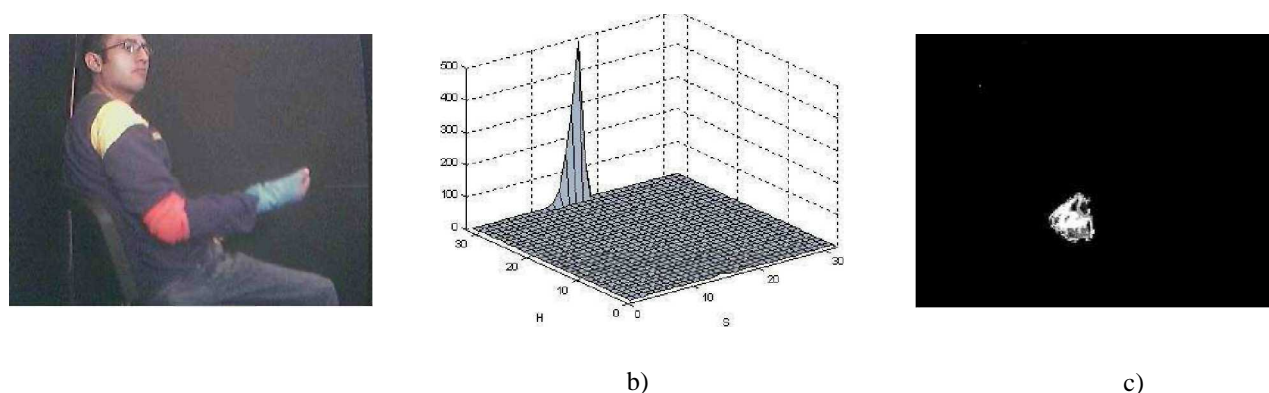


Figura 4.17: La imagen ilustra el proceso de segmentación, a) Imagen original, b) Histograma 2D en el espacio de color  $H-S$  que muestra la distribución del color de la marca roja (codo), y c) Imagen resultante al aplicar el método de *Retroproyección de Histograma*.



Figura 4.18: La imagen ilustra la posición inicial de la marca de color. a) Centro de gravedad del objeto, y b) Ventana de seguimiento inicial.

proceso para segmentar es el siguiente:

- Para cada objeto (marca de color) en la escena, se toman muestras del color y se crea un histograma bidimensional  $M$  en el espacio de color  $(H-S)$ , donde  $M(H, S)$  especifica el número de píxeles con el valor de la dupla  $(H, S)$ .
- Calcular el histograma de la imagen a analizar  $I$ .
- Obtener el histograma  $i$ ?

$$R = \frac{M}{I} \quad (4.17)$$

- Para cada píxel de la imagen  $I(x, y)$  se toma el valor del color  $(H, S)$ , y el píxel de la imagen resultante se sustituye por:

$$I(x, y) = \begin{cases} 255 & \text{si } R(H, S) > 1 \\ 128 \cdot \text{mín}(R(H, S), 1) & \text{de otro modo} \end{cases} \quad (4.18)$$

### 4.3. Seguimiento visual de las marcas: Análisis de gestos

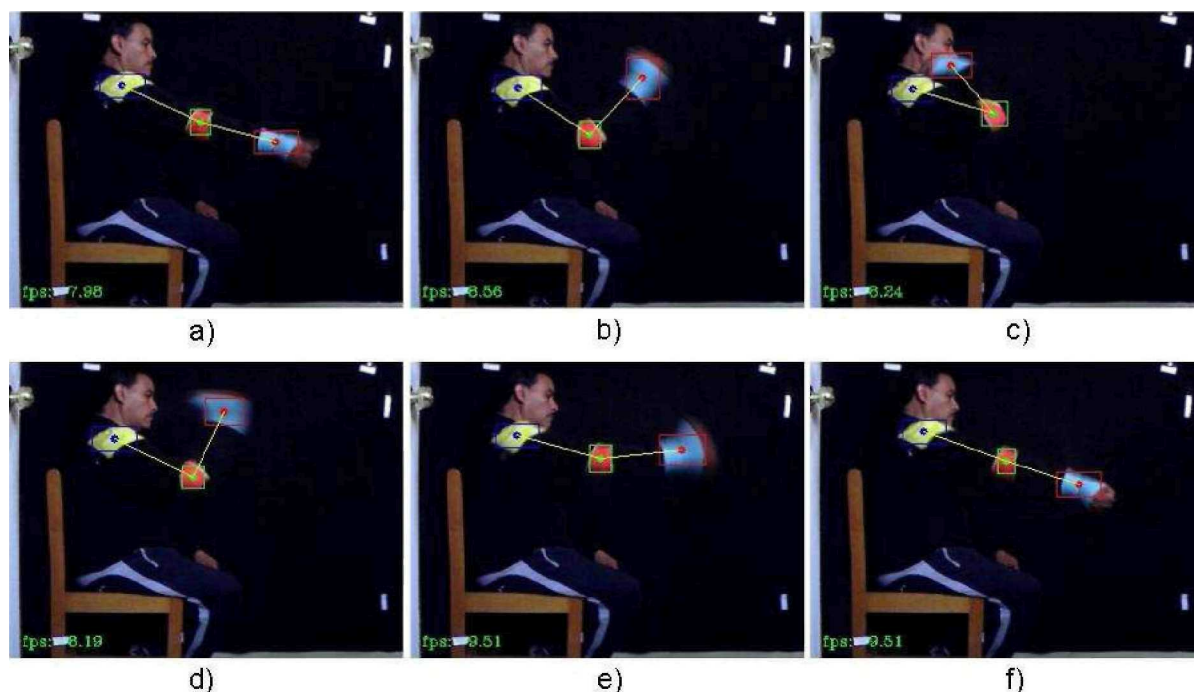


Figura 4.19: Secuencia de imágenes (a - f) que ilustra el proceso de seguimiento de las marcas de color, al realizar el gesto *flexión - extensión*

De esta forma, podemos transformar una imagen de color a una imagen en escala de grises, tanto que el valor de gris de cada pixel muestre el grado de pertenencia con el modelo de color del objeto a buscar. Este proceso se aplica en cada imagen capturadas por las dos cámaras de video, buscando un objeto (marca de color) a la vez. La figura 4.17 muestra el histograma 2D en el espacio de color ( $H - S$ ), que representa la distribución de color de la marca roja (codo del brazo), y la imagen resultante al aplicar el método *Retroproyección de Histograma*.

Después de segmentar las marcas de color, el siguiente paso es obtener las posiciones iniciales de cada una, a partir de las imágenes segmentadas. Para ello, se utilizan los momentos de orden cero y uno [21] (ver pág. 47). Este procedimiento, nos proporciona las coordenadas de la ventana que será utilizada en la fase de seguimiento, con el fin de reducir el cálculo, explicado en la siguiente sección. La Figura 4.18 muestra la posición inicial del codo del usuario, encerrada en un rectángulo. Las marcas son detectadas en los primeros cinco cuadros de la secuencia de imágenes.

#### 4.3.4. Seguimiento 2D

Después de la localización de las marcas de color por medio del método explicado en la sección anterior, el siguiente paso es su seguimiento a través de una secuencia de imágenes, para esto, se aplica el algoritmo *Camshift* [13]. El proceso de seguimiento se aplica sobre la ventana de seguimiento definida con anterioridad. Se aplica el algoritmo de *Retroproyección de Histograma* para mantener la segmentación de las marcas de color durante el seguimiento.

De esta forma, podemos obtener las posiciones de las marcas de color de cada articulación del brazo. El proceso se aplica en las secuencias de imágenes capturadas por las dos cámaras de video. En la Figura 4.19 se muestra una secuencia de imágenes en donde se aplica el algoritmo de



#### 4.4. Resumen

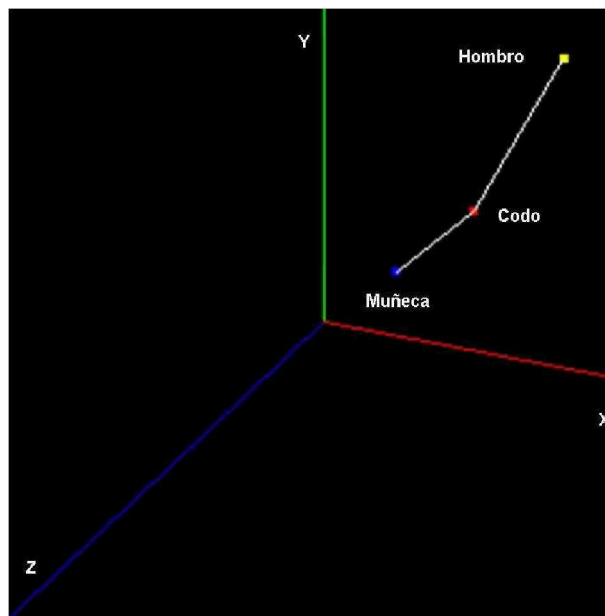


Figura 4.20: La imagen ilustra la representación del brazo en el espacio tridimensional. Se trazan líneas de unión entre las articulaciones.

seguimiento de las tres marcas de color. El punto en las marcas, especifica el centroide de la ventana de seguimiento, y por lo tanto, la posición de la marca en el espacio 2D de la imagen.

#### 4.3.5. Reconstrucción 3D

Después de obtener las posiciones de cada marca en el plano 2D de las imágenes capturadas por las dos cámaras de video, el siguiente paso es estimar su posición en el espacio 3D, para esto, se emplea el procedimiento de intersección de líneas de visión expuesto en la sección 4.2.4.

El proceso de seguimiento se aplica para cada marca de color, esto es, se proyectan tres líneas de visión de cada cámara a las marcas de color adheridas al brazo. De esta forma, se obtiene la posición tridimensional de cada marca con respecto al sistema de referencia del mundo, lo cual nos permite tener una representación del brazo por medio de una estructura de alambre (*stick*), al trazar líneas de unión entre cada articulación (ver Figura 4.20).

Con la representación del brazo en una estructura de alambre en el espacio 3D, se obtiene información de movimiento, como son los ángulos formados por las articulaciones. En el siguiente capítulo se presenta el *Análisis de gestos*, por medio de extraer información de movimiento.

#### 4.4. Resumen

En este capítulo se presentó el sistema de seguimiento visual aplicado a los dos escenarios: a) Seguimiento visual de la mano y, b) Seguimiento visual de las marcas. El seguimiento visual está conformado por cuatro etapas principales: Calibración, Segmentación, Seguimiento 2D y Reconstrucción 3D. En la fase de Calibración se obtienen los parámetros intrínsecos y extrínsecos de la cámara para ser usados en la etapa de reconstrucción 3D. La fase de segmentación se encarga de localizar la

#### 4.4. Resumen

posición inicial del objeto (mano y marcas de color). Para la segmentación de la mano, se emplea información de color y movimiento. En la segmentación de las marcas se utilizan histogramas de color. En la fase de Seguimiento 2D se aplica el seguimiento del objeto a través de una secuencia de imágenes de video, obteniéndose la posición en el espacio 2D del objeto en cada cámara. Finalmente, en la fase de Reconstrucción 3D se estima la posición del objeto en el espacio 3D aplicando un algoritmo de intersección de líneas de visión. El seguimiento visual de la mano tiene como objetivo interactuar con la herramienta *T-WREX*, por medio de enviar las coordenadas (X - Y - Z). En este escenario el paciente controla el ambiente virtual, llevando a cabo diferentes tareas diseñadas para imitar situaciones de la vida real orientadas a la rehabilitación. Por otro lado, el seguimiento de las marcas tiene como objetivo analizar los gestos realizados por el paciente. El brazo es representado en una estructura de alambre en el espacio 3D. La información de movimiento, así como el análisis de gestos y su entrenamiento son descritos en el siguiente capítulo.



# Capítulo 5

## Análisis de gestos

En este capítulo se describe el procedimiento para analizar el movimiento del brazo del paciente dentro del escenario de Análisis de gestos. En la sección 5.1 se presenta la introducción del capítulo, describiendo las fases en que se divide el Análisis de gestos. En la sección 5.2 se presenta la extracción de características de movimiento como resultado del seguimiento de las marcas. En la sección 5.3 se describe el procedimiento para analizar el movimiento del brazo de un individuo. Finalmente, en la sección 5.4 se presenta el resumen del capítulo.

### 5.1. Introducción

El análisis de movimiento del brazo de un individuo se muestra en un diagrama de bloques en la Figura 5.1. La descripción de cada bloque se muestra a continuación:

- Módulo de seguimiento visual. Este módulo (M1) se encarga de procesar las secuencias de imágenes capturadas por dos cámaras de video para localizar y seguir el brazo, empleando marcas de color, expuesto en el capítulo anterior.
- Módulo de extracción de características. En este módulo (M2) se extraen las características de movimiento del brazo como resultado del módulo de seguimiento visual. Estas características se basan en ángulos entre las articulaciones del brazo y el antebrazo, y entre el brazo y torso, para representar la postura del brazo en el espacio 3D. Posteriormente, se aplica un algoritmo de discretización para preparar los datos que serán utilizados en el módulo de análisis de movimiento.
- Módulo de análisis de movimiento. Este módulo (M3) se encarga de comparar las trayectorias de movimientos (gestos) del brazo del paciente con los movimientos correctos de personas sanas. Los datos discretos, generados en el módulo anterior, son utilizados para modelar y comparar las trayectorias de movimiento utilizando Modelos Ocultos de Markov.

## 5.2. Extracción de características de movimiento

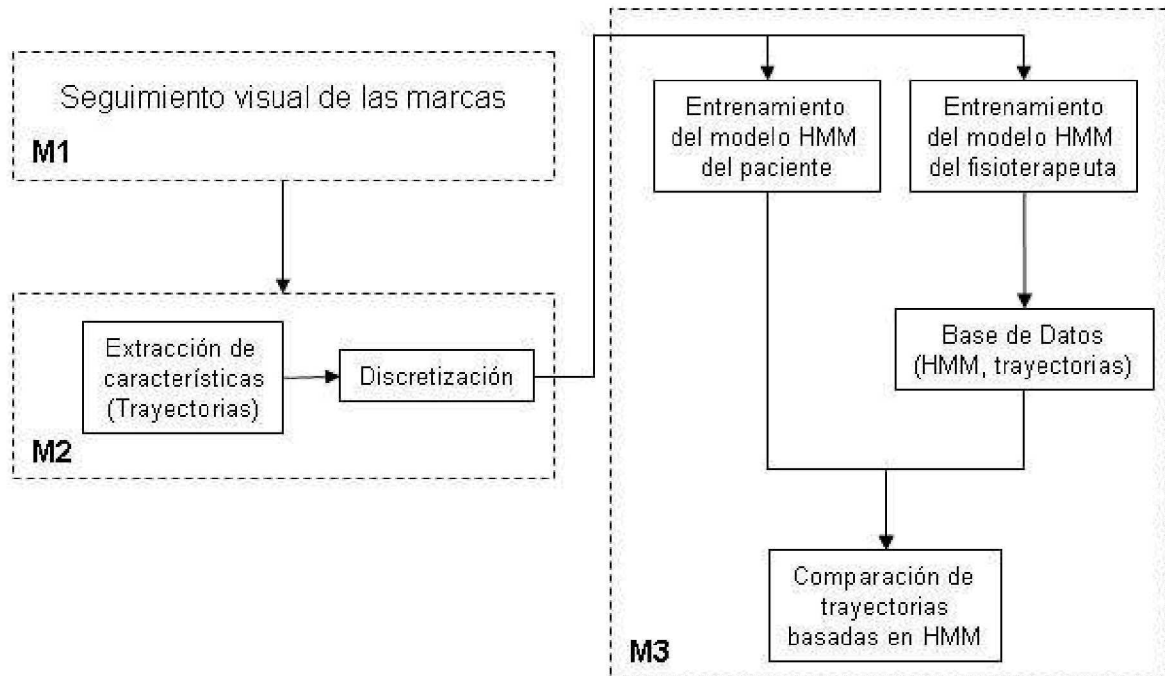


Figura 5.1: Diagrama de bloques del sistema desarrollado. M1) Módulo de seguimiento visual, M2) Módulo de extracción de características, y M3) Módulo de análisis de movimiento

## 5.2. Extracción de características de movimiento

Una vez que se determinó el proceso de segmentación y seguimiento del brazo en un espacio tridimensional, el siguiente paso es determinar qué características son convenientes para analizar el movimiento del brazo. Los métodos más comunes para representar las características de movimiento están basados en la trayectoria descrita por el objeto a seguir. Sin embargo, debido a que el brazo está representado por un conjunto de tres objetos (marcas de color), el uso de trayectorias para representar el movimiento provocaría una carga de variables, además de su manejo complicado, ya que se tendría tres variables ( $X$ ,  $Y$ ,  $Z$ ), por cada articulación que conforma el brazo.

Una alternativa para modelar el movimiento del brazo, es la representación de la postura en cada instante de tiempo, por medio de ángulos entre las articulaciones y el torso del usuario. Los ángulos son invariantes a translaciones y escalamiento, y son independientes de las características físicas de la persona.

En esta tesis se proponen tres grados de libertad para representar la postura del brazo (ver Figura 5.2). El primer ángulo ( $\alpha$ ) está formado entre el brazo y el antebrazo. Para el segundo y tercer ángulo ( $\theta$ ,  $\Phi$ ), se considera la posición del brazo con respecto al torso. Para esto, se toma como nuevo punto de referencia al hombro y se obtienen las coordenadas esféricas del codo.

Por lo tanto, el conjunto de características que describe la trayectoria del brazo al realizar los gestos, está formada por  $V = [\alpha - \theta - \Phi]$ . La obtención de estos valores se muestra a continuación.

Las posiciones tridimensionales de cada articulación están definidas como:

## 5.2. Extracción de características de movimiento

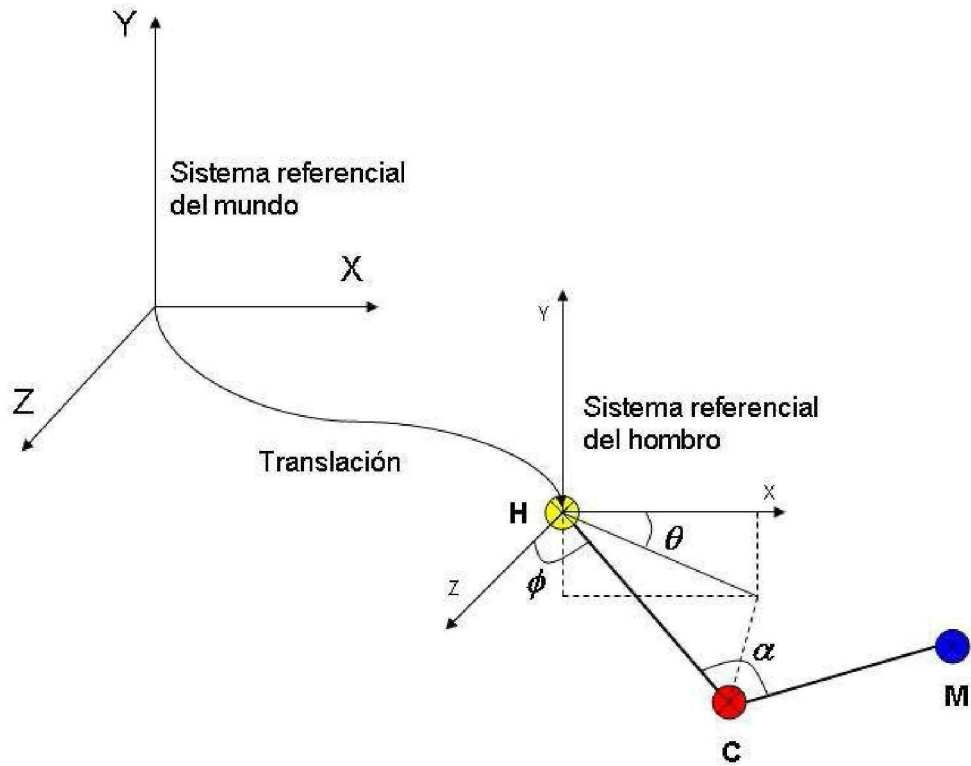


Figura 5.2: La imagen ilustra la postura del brazo en un instante de tiempo  $\alpha$  indica el ángulo entre el brazo y el antebrazo, los ángulos  $\theta$  y  $\Phi$  representan las coordenadas esféricas del codo tomando como referencia al hombro. H, M y C representan las coordenadas tridimensionales (X, Y, Z) del hombro, codo y muñeca, respectivamente.

$$H = \begin{bmatrix} X_H \\ Y_H \\ Z_H \end{bmatrix}, C = \begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix}, M = \begin{bmatrix} X_M \\ Y_M \\ Z_M \end{bmatrix} \quad (5.1)$$

donde H, M y C representan las posiciones 3D del hombro, codo y muñeca, respectivamente.

El ángulo entre el brazo y el antebrazo ( $\alpha$ ) se calcula a partir del ángulo que forman dos líneas en el espacio [5], dadas por las siguientes ecuaciones:

$$\begin{aligned} Q &= M - C \\ R &= C - H \\ \alpha &= \cos^{-1} \left( \frac{\|R \cdot Q\|}{\|R\| \cdot \|Q\|} \right) = \cos^{-1} \left( \frac{X_R X_Q + Y_R Y_Q + Z_R Z_Q}{\sqrt{X_R^2 + Y_R^2 + Z_R^2} \sqrt{X_Q^2 + Y_Q^2 + Z_Q^2}} \right) \end{aligned} \quad (5.2)$$

Los ángulos entre el brazo y el torso ( $\theta$ ,  $\Phi$ ), se obtienen por medio de las coordenadas esféricas del codo tomando como referencia el hombro. Para esto, se realiza una translación del origen del sistema de referencia del mundo a la posición del hombro, esto es:

## 5.2. Extracción de características de movimiento

$$D = C-H \quad (5.3)$$

donde el vector  $D$  representa la nueva posición del codo después de aplicar la translación. De esta forma, para convertir las coordenadas cartesianas a esféricas se aplican las siguientes ecuaciones [5]:

$$\begin{aligned} \rho &= \sqrt{X_D^2 + Y_D^2 + Z_D^2} \\ \theta &= \tan^{-1} \left( \frac{Y_D}{X_D} \right) \\ \phi &= \cos^{-1} \left( \frac{Z_D}{\rho} \right) \end{aligned} \quad (5.4)$$

donde  $\theta$  es el ángulo que se forma al proyectarse el segmento de línea en el plano  $(X, Y)$ .  $\Phi$  es el ángulo que indica la apertura del segmento de línea con respecto al eje  $Z$  (ver Figura 5.2). Finalmente, los tres ángulos son normalizados entre  $[0,0 — 1,0]$ , al dividirse entre  $2\pi$  radianes:

$$\begin{aligned} \alpha^* &= \frac{\alpha}{2\pi} \\ \theta^* &= \frac{\theta}{2\pi} \\ \phi^* &= \frac{\phi}{2\pi} \end{aligned} \quad (5.5)$$

De esta forma, se puede representar la trayectoria descrita por el movimiento del brazo al realizar un gesto, bajo el sistema de coordenadas  $\alpha — \theta — \Phi$ , como:

$$F = \{(\alpha_1^*, \theta_1^*, \phi_1^*), \dots, (\alpha_n^*, \theta_n^*, \phi_n^*)\} \quad (5.6)$$

donde  $n$  es la longitud de la trayectoria.

La Figura 5.3 muestra la trayectoria del movimiento del brazo al realizar el gesto *circular*, en el espacio de características  $[\alpha — \theta — \Phi]$ , así como la variación de los ángulos con respecto al tiempo.

### 5.2.1. Discretización

Después de extraer el conjunto de características que describen la trayectoria del brazo, el siguiente paso es discretizarlas con el fin de obtener símbolos discretos que serán utilizadas por los *Modelos Ocultos de Markov* [43], en la fase del análisis del movimiento. Para llevar esto a cabo, se utiliza el algoritmo *K-medias* [21], el cual permite clasificar los datos en  $K$  clases o *clusters*, dentro de nuestro espacio de características  $(\alpha — \theta — \Phi)$ . En otras palabras, el algoritmo permite pasar del conjunto de datos continuos (que representan la trayectoria del brazo) a un conjunto de datos discretos más compacto representado por  $K$  clases.

## 5.2. Extracción de características de movimiento

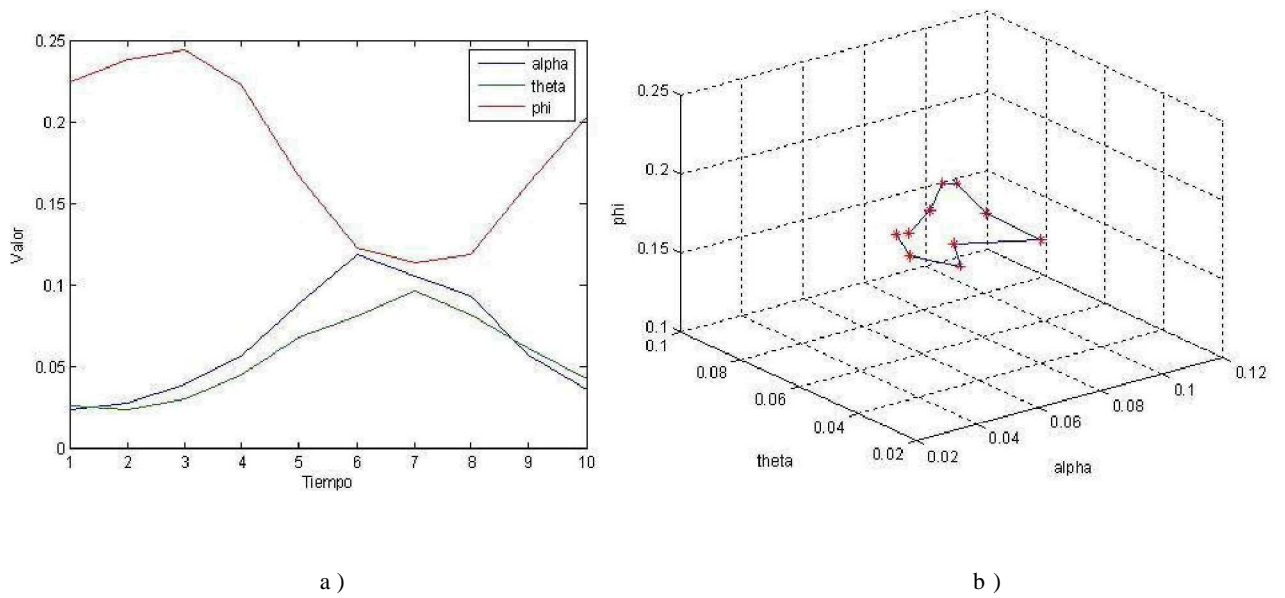


Figura 5.3: Imagen que ilustra la trayectoria del gesto *circular*, a) Variación de los ángulos con respecto al tiempo y, b) Trayectoria del gesto en el espacio de características  $[\alpha - \theta - \Phi]$ .

El algoritmo *K-medias*, es un método sencillo, pero muy eficiente para agrupar  $K$  clases en un espacio de características. Este algoritmo se basa en la distancia mínima entre el punto medio de cada clase ( $K$ ) y los puntos de las características.

Partiendo de un conjunto de datos a clasificar  $X_1, X_2, \dots, X_n$  (en nuestro caso  $X = [\alpha - \theta - \Phi]^T$ ), el algoritmo *K-medias* realiza las siguientes operaciones:

2. Clasificación. Se distribuyen todas las muestras  $\{X\}_{1 \leq i \leq n}$  entre las  $K$  clases  $C_i$  usando el clasificador de mínima distancia ( $X \in C_i$  *sii*  $\|X - Z_i\| \leq \|X - Z_j\|, \forall i = 1, 2, \dots, K, i \neq j$ )
3. Actualización. Una vez redistribuidos los elementos a agrupar entre las diferentes clases, se actualizan los centroides de las clases por medio de calcular el punto medio de los vectores de entrenamiento en cada grupo.
4. Finalizar. Se termina el agrupamiento cuando la diferencia entre la distancia de cada punto a los centroides de las clases es mínima, de lo contrario se continúa en el paso 2.

A este proceso de agrupamiento también se le conoce como *Cuantización Vectorial (VQ)*, el cual nos da como resultado un *Libro de códigos (Codebook)*, que contiene un número de símbolos y las coordenadas del centroide de cada clase. El proceso del reconocimiento del VQ es el siguiente. Una vez obtenido el libro de códigos por medio del algoritmo *K-medias*, se introduce un vector  $(\alpha - \theta - \Phi)$  representando los datos de un gesto, se calcula la distancia *Euclidea* con cada una de las clases, y se determina su pertenencia a alguna de las clases utilizando la distancia mínima. Así, obtenemos los símbolos de observación utilizados para entrenar los Modelos Ocultos de Markov [43]

$$(F = \{(\alpha_1, \theta_1, \phi_1), \dots, (\alpha_n, \theta_n, \phi_n)\} \xrightarrow{VQ} O = \{o_1, \dots, o_n\}).$$

### 5.3. Análisis del movimiento

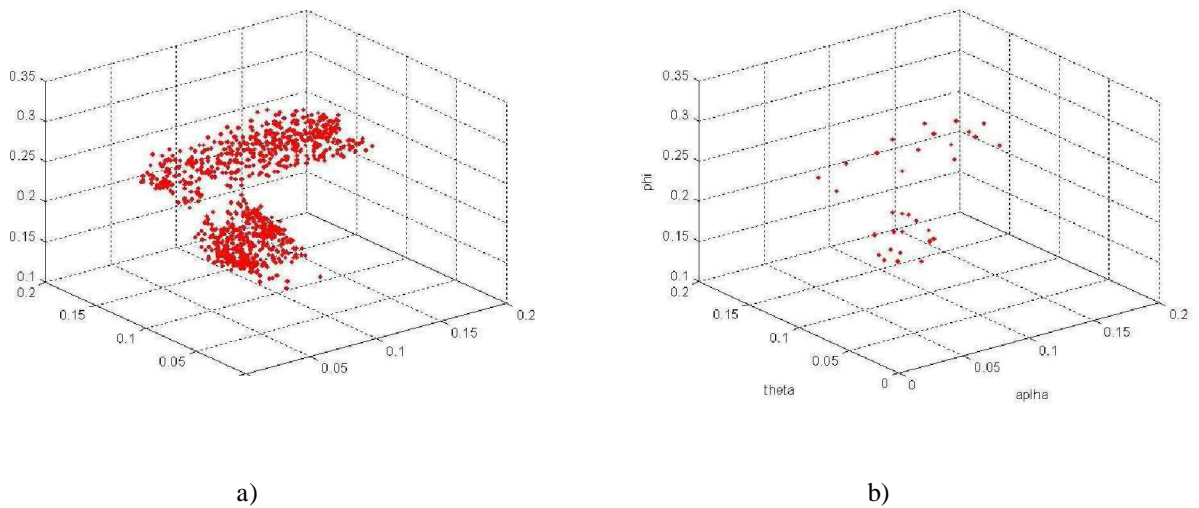


Figura 5.4: La imagen muestra el proceso de agrupamiento al aplicar el algoritmo K-medias sobre el conjunto de datos de las trayectorias de los gestos. a) Conjunto total de datos de trayectorias y, b) Agrupamiento en 32 clases.

La Figura 5.4 muestra el procedimiento de agrupamiento aplicando el algoritmo *K-medias*. La imagen a) muestra el conjunto total de datos de entrenamiento pertenecientes a los gestos *flexión-extensión* y *circular*, en el espacio de características  $[\alpha - \theta - \Phi]$ . La imagen b) muestra las clases en que fue agrupado el conjunto de datos, en este caso es de 32 clases.

## 5.3. Análisis del movimiento

Una vez establecido el procedimiento para el seguimiento del brazo de una persona y obtener su estructura 3D, así como su extracción de características de movimiento, el siguiente paso es modelar la trayectoria del brazo que describe un gesto. En esta tesis, el objetivo del análisis del movimiento es comparar un mismo gesto ejecutado *correctamente* por una persona sana con otro ejecutado por un paciente que haya sufrido un derrame cerebral. Para esto, se propone una métrica basada en modelos probabilistas, por medio del uso de los Modelos Ocultos de Markov (MOM) [43], que como se expuso en el capítulo 3, es una técnica que permite una representación apropiada a secuencias de datos *espacio-temporales*.

### 5.3.1. Topología de los MOM empleada

El primer paso, es escoger una topología para los MOM que sea adecuada para el análisis de gestos dinámicos. La topología escogida es la estructura *Izquierda-Derecha Continua* [43], la cual puede modelar fácilmente secuencias de observaciones cuya estructura cambia en el tiempo (ver Figura 5.5). En estos modelos, la transición entre estados ocurre de izquierda a derecha, regresando al estado inicial, conforme pasa el tiempo.

## 5.4. Resumen

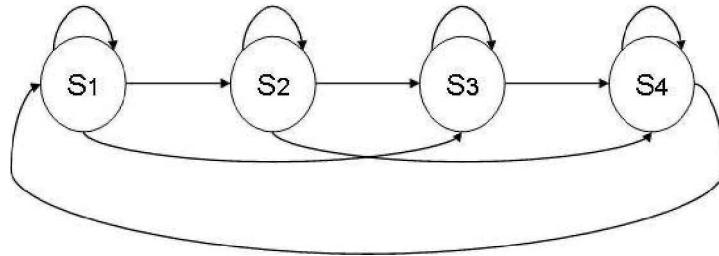


Figura 5.5: Topología empleada en los MOM.

### 5.3.2. Entrenamiento de los MOM

Cada uno de los gestos a analizar (*flexión-extensión* y *circular*), está representado por un MOM. El método para entrenar a cada MOM es el siguiente. Se obtuvieron distintas secuencias de observaciones de movimiento correspondientes a cada gesto. Posteriormente, para la estimación de los parámetros correspondientes a cada uno de los tres MOM se empleó el algoritmo *Baum-Welch*, detallado en la sección 3.1.

El método de entrenamiento se aplica a dos grupos: *normal* y *paciente*. El grupo *normal* está conformado por los gestos ejecutados por personas sanas, guardándose en un base de datos los tres MOM para cada gesto y sus correspondientes secuencias de observaciones. El grupo *paciente*, está conformado por los gestos realizados por el paciente a analizar, comparándose posteriormente cada uno de los MOM con los del grupo *normal*.

### 5.3.3. Comparación de gestos empleando MOM

El método para comparar dos gestos iguales es el siguiente. Una vez entrenado los dos modelos MOM's para cada gesto del grupo *normal* y del grupo *paciente*, se emplea la métrica basada en la distancia de *Kullback-Leibler* [31] detallada en la sección 3.2.2. Para esto, se toman dos modelos MOM's, uno del grupo *normal* y otro del grupo *paciente* correspondientes al mismo gesto, y se obtiene su distancia. En este trabajo postulamos la siguiente hipótesis: Si la distancia es cercana a cero, se estima que la ejecución del gesto realizada por el paciente es similar al gesto ejecutado *correctamente*, de lo contrario, si la distancia es mayor, se estima que el paciente no ha realizado correctamente el gesto.

## 5.4. Resumen

En este capítulo se presentó el análisis de movimiento del brazo de un individuo. El sistema se divide en tres módulos principales: *a) Módulo de seguimiento visual*, *b) Módulo de extracción de características* y *c) Módulo de análisis del movimiento*. El Módulo de seguimiento visual (expuesto en el capítulo anterior) localiza y sigue las articulaciones del brazo de una persona empleando cintas de color como marcas distintivas, de este modo se reconstruye el brazo en una estructura de alambre 3D. El módulo de extracción de características obtiene las trayectorias del movimiento del brazo.

#### 5.4. Resumen

Dichas trayectorias se modelan por medio de ángulos entre el brazo y el antebrazo, y entre el brazo y el torso del usuario, extraídas de la estructura de alambre 3D del brazo. Posteriormente, se discretizan los datos de las trayectorias para ser usados por los MOM. Finalmente, en el Módulo de análisis del movimiento, se entrenan los MOM correspondientes a los dos gestos, empleando las secuencias de observaciones de movimiento obtenidas. Se crean dos grupos de datos de entrenamiento, uno para los gestos *correctos* realizados por personas sanas, y otro para los gestos realizados por el paciente. Posteriormente, se obtiene una métrica entre dos MOM empleando la distancia de *Kullback-Leiblar*. Los resultados experimentales y el análisis sobre ellos, se detallan en el siguiente capítulo.



# Capítulo 6

## Resultados experimentales

En este capítulo se describen los experimentos y resultados obtenidos en los dos sistemas de seguimiento visual descritos en este documento: a) Sistema de seguimiento de la mano y b) Sistema de seguimiento de brazo para el análisis de gestos. En la sección 6.1 se presentan los resultados experimentales del sistema de seguimiento de la mano. En la sección 6.2 se describen los resultados experimentales del sistema de seguimiento de brazo para el análisis de gestos. Finalmente, en la sección 6.3 se presenta el resumen del capítulo.

### 6.1. Seguimiento visual de la mano: T-WREX

El sistema de seguimiento visual de la mano se divide en cuatro etapas: a) Calibración de las cámaras, b) Segmentación, b) Seguimiento 2D y c) Reconstrucción 3D. A continuación se detallan los cuatro procedimientos.

#### 6.1.1. Ambiente de trabajo

Antes de exponer los experimentos realizados, se presenta el ambiente de trabajo y la plataforma de desarrollo en las que se implementó y probó el sistema. El ambiente de trabajo se muestra en la Figura 6.1. El paciente se posiciona frente a la cámara frontal a una distancia de aproximadamente 1.5 metros. En la mesa, sobre la cual están colocadas las cámaras de video, se cubre con una tela de color uniforme (negro o azul) para reducir el ruido en el proceso de segmentación. De igual forma, el paciente deberá usar una camisa de manga larga para simplificar el seguimiento de la mano. Para las condiciones de iluminación se empleó luz artificial. Bajo estas condiciones el sistema puede localizar y seguir la mano del usuario eficientemente en tiempo real.

El sistema se implementó sobre una computadora Compaq Presario V2000 con un procesador Intel Celeron a 1.5 GHz y 256 MB de memoria RAM, ejecutándose sobre una plataforma Windows XP. Para la captura de video se emplearon dos cámaras web convencionales de bajo costo. En particular se ha empleado una cámara Logitech QuickCam Chat y una cámara Genius VideoCAM Messenger. Ambas cámaras tienen una velocidad de captura de 20 cuadros por segundo. Además se empleó un formato de salida de imagen de 320 x 240 píxeles.

## 6.1. Seguimiento visual de la mano: T-WREX

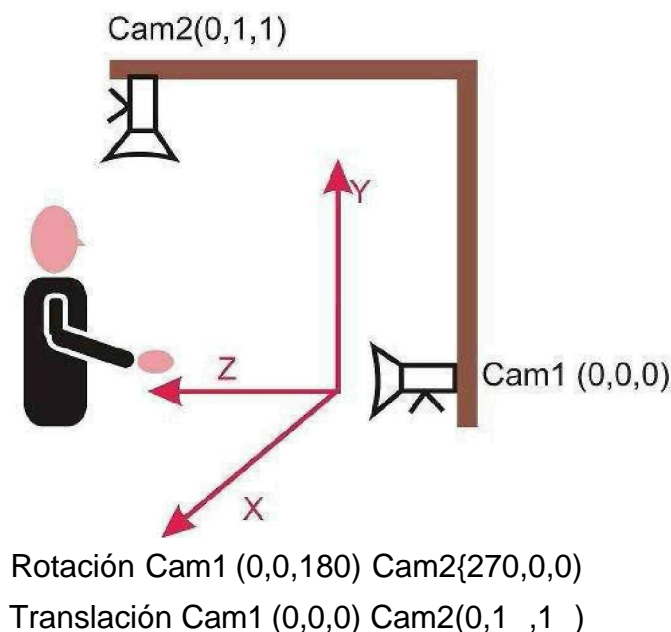


Figura 6.1: Imagen que muestra la configuración de las cámaras 1 y 2 en el ambiente de trabajo para el seguimiento de la mano: T-WREX.

### 6.1.2. Calibración de las cámaras

La calibración consiste en obtener los parámetros *intrínsecos* (longitud focal, centro óptico, tamaño del pixel) y los parámetros *extrínsecos* (posición y orientación) de las cámaras. Los parámetros intrínsecos son obtenidos por medio de un patrón de calibrado que es colocado frente a la cámara. El algoritmo de calibración detecta las esquinas de las cuadrículas del tablero, almacenando una secuencia de imágenes con diferentes vistas del tablero, como se muestra en la Figura 6.2. Dicha calibración se realizó empleando software basado en la biblioteca de visión computacional *OpenCV* [36].

Los parámetros extrínsecos se obtienen manualmente por medio de medir la posición y orientación del sistema de referencia de la cámara (centro del lente) con respecto al sistema de referencia del mundo. La Figura 6.1 muestra la configuración de las cámaras en el ambiente de trabajo para el sistema *T-WREX* [44]. La cámara 1 se encuentra en el centro de referencia del mundo en la posición (X: 0, Y: 0, Z: 0) (en metros) y rotada 180 grados en el eje Z, orientada hacia el plano XY. La cámara 2 se encuentra en la posición (X: 0, Y: 1, Z: 1) y rotada 270 grados en el eje X, orientada hacia el plano XZ.

El proceso de calibrado para obtener los parámetros intrínsecos se aplica sólo una vez y se guarda en la configuración inicial del sistema, por lo tanto en las sesiones subsecuentes este procedimiento no se necesita repetir. Por otro lado, si las cámaras son cambiadas de posición, los parámetros extrínsecos necesitan ser medidos nuevamente.

Las imágenes tomadas por dos cámaras cercanas, como en el caso de un par estéreo clásico, son muy parecidas, tienen pequeñas desviaciones una de otra. Por ello requieren de mayor precisión en sus parámetros de calibración para calcular la profundidad de los objetos con una exactitud aceptable. Por otro lado, es más fácil encontrar similitudes entre las dos imágenes y así localizar objetos por su forma, precisamente por tener pocas diferencias. En contraste, la colocación perpendicular de las

## 6.1. Seguimiento visual de la mano: T-WREX

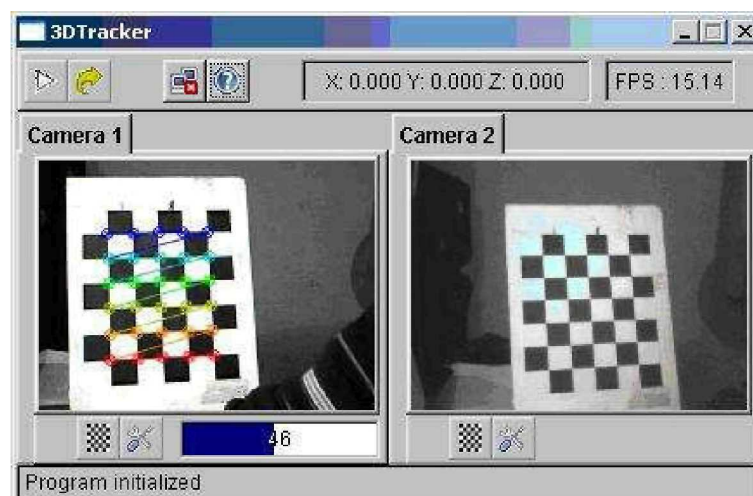


Figura 6.2: Patrón de calibrado para obtener los parámetros intrínsecos de cada cámara (cámara 1 y 2).

cámaras facilita la compensación de errores al tomar mayor información del entorno. Las cámaras lejanas perpendiculares dan mayor información al completar en una cámara lo que la otra no ve.

### 6.1.3. Segmentación

El proceso de localización de la mano consiste en dos algoritmos de segmentación: color y movimiento, expuestos en la sección 4.2.2. El algoritmo que combina color y movimiento sirve para localizar la posición inicial de la mano en la imagen. Posteriormente, sólo se aplica el algoritmo de segmentación basado color. Para el proceso de segmentación de color, se aplica el clasificador bayesiano que clasifica dos tipos de clases: pixeles con color de piel y pixeles que no son piel. Para establecer la distribución de probabilidad del color de la piel se tomaron muestras de imágenes (adquiridas de la *web*) de personas de diferentes etnias: Asiáticos, Africanos y Caucásicos, además de una secuencia de imágenes de video de personas para el grupo de la clase *piel*; así como de imágenes arbitrarias en ausencia de pixeles de piel, para el grupo de la clase *no-piel*. De las imágenes de la clase *piel*, y utilizando un editor gráfico, se extrajeron zonas que solo alberguen pixeles de color de la piel

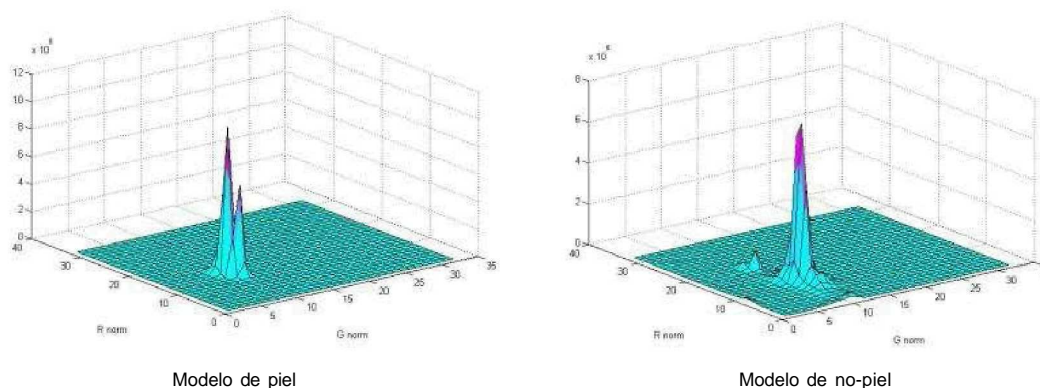


Figura 6.3: Distribución de los pixeles en los modelos de histograma de piel y no-piel, en el espacio de color H-S.

## 6.1. Seguimiento visual de la mano: T-WREX



Figura 6.4: Imagen que muestra el clasificador bayesiano para la segmentación de regiones de piel. La imagen de la izquierda muestra la imagen original y, la imagen de la derecha muestra los píxeles en escala de grises, donde el valor de gris (0 - 255) indica la pertenencia a la clase piel.

en su interior y se generaron nuevas imágenes. Para las imágenes de la clase *no-piel*, no se modificó su contenido, solo se seleccionaron las que no contenían píxeles de piel. El número total de píxeles de piel muestreado es igual a 3,354,974. El número total de píxeles de *no-piel* es igual a 3,459,358.

Con los dos grupos de imágenes *de piel* y *no-piel*, se construyeron 2 histogramas de color (en el espacio de color *H-S*) que representan la distribución de probabilidad discreta de ambas clases. La Figura 6.3 expone la distribución del color de las clases *piel* y *no-piel* en el espacio del color *H-S*. Se puede observar en ambas gráficas que los píxeles muestreados como piel ocurren también en el modelo de *no-piel*.

Para mejorar el proceso de segmentación, se crea un Mapa de Probabilidad de Piel (MPP), el cual consiste en una matriz bidimensional de tamaño 32x32, igual al tamaño de los modelos de histogramas de piel y *no-piel*. El MPP se obtiene a partir de la base de datos con valores de imágenes de piel y *no-piel*, calculando el clasificador bayesiano sobre todas las posibles combinaciones de valores que pueda tomar un píxel en el espacio de color *H-S*. De esta forma, para segmentar una imagen basta con tomar el valor de un píxel en la dupla *H-S* y acceder en la matriz bidimensional MPP para obtener la probabilidad de pertenencia a la clase piel. En la Figura 6.4 se muestra un ejemplo del clasificador bayesiano de piel. El clasificador se aplicó sobre cada píxel de la imagen original, dando como resultado una imagen en escala de grises. El nivel de gris de un píxel muestra el grado de pertenencia a la clase piel.

Existe un problema en el proceso de segmentación de píxeles de piel: Objetos o regiones en la imagen con valores de píxeles semejantes al color de la piel. Esto provoca ruido en la fase de seguimiento, causando la pérdida del seguimiento de la mano. Sin embargo, bajo ciertas condiciones de iluminación controlada (luz artificial), la segmentación se realiza de forma correcta.

### 6.1.4. Seguimiento 2D

El método para el seguimiento de la mano se basa en el algoritmo *Camshift* descrito en la sección 4.2.3. *Camshift* trabaja sobre una ventana de seguimiento y una imagen en escala de grises. La posición inicial de la ventana de seguimiento se obtiene por medio del algoritmo de segmentación de *color* y *movimiento*, descrito en la sección 4.2.2. En este trabajo se realiza una modificación al algoritmo *Camshift* original [13], el cual consiste en redimensionar la ventana a un tamaño máximo, esto

## 6.1. Seguimiento visual de la mano: T-WREX

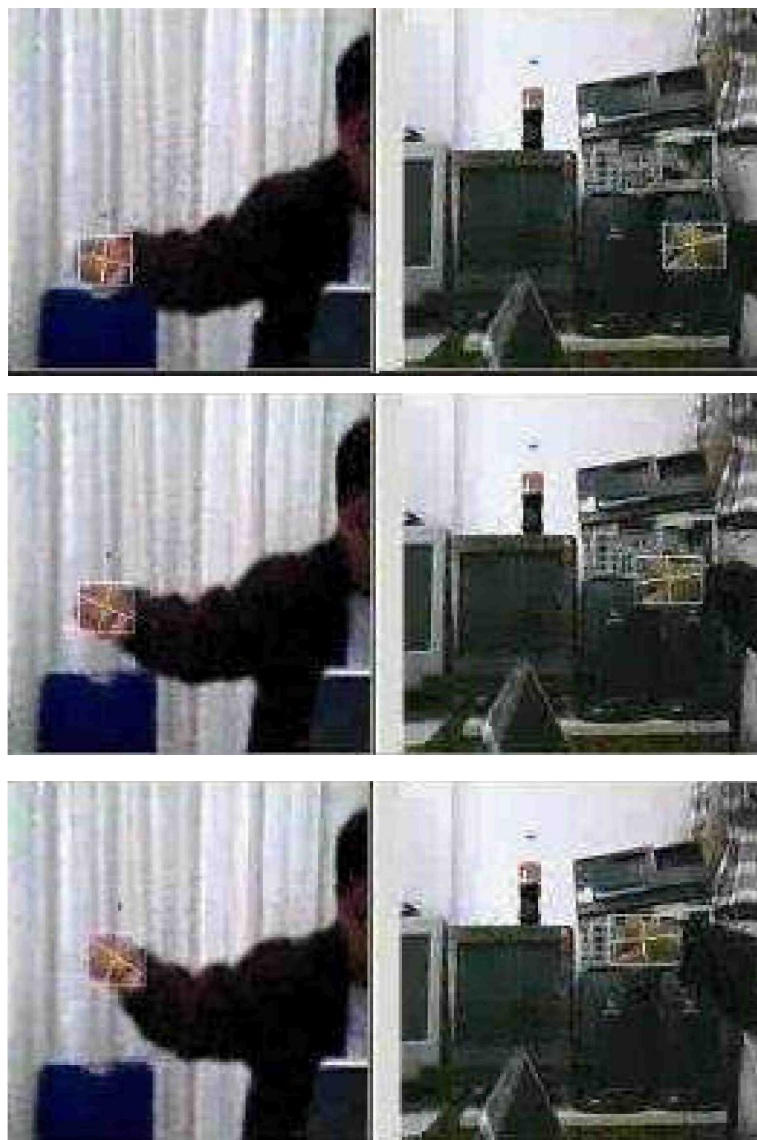


Figura 6.5: La imagen muestra el seguimiento de la mano en dos secuencias de imágenes capturadas por dos cámaras de video empleando el algoritmo Camshift.

es, Camshift modifica las dimensiones de la ventana de seguimiento hasta un tamaño permitido. Esto reduce el ruido en el seguimiento debido a la presencia de algunas regiones en la imagen semejantes al color de piel. El tamaño máximo de la ventana de seguimiento es de 100x100 píxeles.

Durante la etapa de seguimiento Camshift utiliza el MPP para obtener la imagen en escala de grises que representa la probabilidad de cada píxel de pertenecer a la clase piel. La segmentación de píxeles de color de piel se aplica en la ventana de seguimiento del Camshift. Posteriormente, Camshift calcula el centro de masa del objeto y obtiene sus coordenadas en el espacio 2D de la imagen y, al mismo tiempo, modifica las dimensiones de la ventana de seguimiento. El seguimiento de la mano se aplica en las secuencias de imágenes capturadas por las dos cámaras de video. En la Figura 6.5 se presenta una secuencia de imágenes que muestran el resultado del seguimiento de la mano. En los experimentos se empleó una camisa de manga larga para favorecer el seguimiento de la mano. Además, la distancia de las cámaras a la mano del usuario fue de aproximadamente 1.5 metros.

El empleo de una ventana de seguimiento reduce el cálculo en el proceso de segmentación

## 6.1. Seguimiento visual de la mano: T-WREX

$X_e$	$Y_e$	$Z_e$	$X_r$	$Y_r$	$Z_r$	Error
125	17	143	122	20	147	5.8
94	12	117	90	10	120	5.3
48	36	92	50	35	90	3
77	17	65	75	17	67	2.8
107	24	86	105	22	85	3
45	54	80	47	55	74	6.4

Tabla 6.1: Resultados de la etapa de reconstrucción 3D (en centímetros). Las columnas  $X_e, Y_e, Z_e$  muestran las posiciones de la mano estimadas por el sistema, las columnas  $X_r, Y_r, Z_r$  muestran las posiciones reales medidas manualmente, y la columna *Error* muestra la diferencia entre la posición estimada y la posición real.

para localizar la mano del usuario. Sin embargo, un problema en el seguimiento de la mano al emplear solamente el color de piel es la posible pérdida del seguimiento al detectar otra región de piel dentro de la ventana de seguimiento. Este problema se presenta en las secuencias de imágenes capturadas por la cámara frontal, ya que el seguimiento de la mano se pierde cuando pasa cerca del rostro del usuario.

### 6.1.5. Reconstrucción 3D

La etapa de Reconstrucción 3D tiene como objetivo estimar la posición del centroide de la mano del usuario en el espacio tridimensional. La Reconstrucción 3D se basa en el algoritmo de *Intersección de líneas de visión* descrito en la sección 4.2.4. Este algoritmo determina el punto de cruce más cercano entre dos líneas de visión en el espacio 3D. Cada línea de visión está formada por dos puntos que van del centro óptico de la cámara al centroide de la mano. Para establecer las líneas de visión se realizó la calibración de las cámaras, obteniéndose sus parámetros intrínsecos y extrínsecos, por medio del proceso de calibrado explicado anteriormente. En los experimentos realizados, el volumen cubierto por el campo de visión de ambas cámaras fue de 1.5 metros cúbicos aproximadamente.

Para caracterizar la precisión de la etapa de reconstrucción 3D se realizó una serie de medidas colocando la mano del usuario en distintas posiciones conocidas y anotando la estimación de la posición dada por el sistema de seguimiento. Estas medidas aparecen resumidas en la Tabla 6.1, en centímetros, siendo  $X_e, Y_e, Z_e$  la posición estimada y  $X_r, Y_r, Z_r$ , la posición real. Para obtener el *Error*, se calcula la distancia euclidia entre la posición estimada y las posición real.

La diferencia entre la posición real y la estimada se encuentra aproximadamente entre tres y cinco centímetros. Más allá de los números exactos, la principal conclusión de estas pruebas es que en el prototipo experimental el sistema entrega un error absoluto del orden de centímetros. Sobre este margen de error habría que considerar la posible existencia de errores del orden de centímetros en las medidas reales al realizarse con un metro convencional. También hay que matizar esos resultados pensando en que el volumen de solape entre los campos visuales de ambas cámaras es de 1.5 metros cúbicos. Habría que estudiar qué ocurre cuando el volumen cubierto es mayor y las cámaras están más alejadas.

Un factor crítico encontrado a la hora de acotar los errores son las variaciones de orientación de ambas cámaras. Incluso usando sistemas de calibración poco precisos, los errores de orientación fueron mucho más influyentes en el error final del sistema que los errores en los parámetros intrínse-



## 6.1. Seguimiento visual de la mano: T-WREX



Figura 6.6: Ejemplos del sistema de seguimiento visual de la mano al interactuar con la herramienta T-WREX. La ventana de seguimiento visual se despliega en la parte superior izquierda de la pantalla.

cos. Variaciones de menos de 5 grados en una de las cámaras respecto a la matriz de rotación usada, conducen en errores en la medida estimada alrededor de 30 cm, a una distancia de medio metro de la cámara.

### 6.1.6. Implementación

El sistema de seguimiento visual ha sido implementado en el lenguaje C++ usando la Programación Orientada a Objetos (POO), y se ha empleado la librería *OpenCV* [36], como herramienta auxiliar para el procesamiento y captura de las imágenes de las cámaras de video. El sistema de seguimiento se encuentra programado en una librería de enlace dinámico (DLL) que es utilizada por la herramienta *T-WREX* para medir la posición de la mano de un individuo en el espacio tridimensional.

## 6.2. Seguimiento visual de las marcas: Análisis de gestos

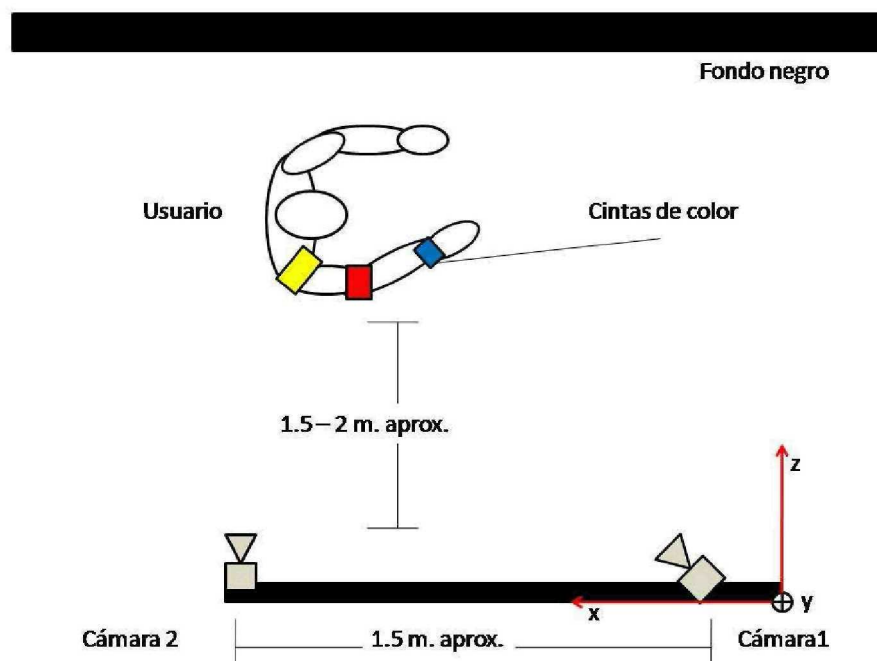


Figura 6.7: Ambiente de trabajo para el sistema de seguimiento de las marcas: análisis de gestos.

La interfaz gráfica del sistema de seguimiento visual se despliega en la pantalla cuando es solicitada por la herramienta T-WREX, mostrando el seguimiento de la mano en las secuencias de imágenes capturadas por las dos cámaras de video. La Figura 6.6 muestra el sistema de seguimiento visual de la mano interactuando con la herramienta *T-WREX*.

## 6.2. Seguimiento visual de las marcas: Análisis de gestos

El sistema de seguimiento visual de las marcas se divide en cinco etapas principales: a) Calibración de las cámaras, b) Segmentación, c) Seguimiento 2D, d) Reconstrucción 3D y e) Análisis de gestos. A continuación se describen los resultados de cada etapa.

### 6.2.1. Ambiente de trabajo

El ambiente de trabajo se muestra en la Figura 6.7. La distancia esperada entre el usuario y las cámaras es de 1.5 a 2 metros aproximadamente. En los experimentos se utilizó un soporte de madera para colocar las cámaras de video, las cuales se encuentran separadas entre ellas a una distancia de 1.5 metros aproximadamente. Se utilizaron cintas de color para identificar las articulaciones del brazo del individuo. Las cintas son fijadas a las articulaciones del brazo y cada una de ellas tiene un color distintivo. La cinta de color amarillo corresponde al hombro, la cinta de color rojo corresponde al codo y, la cinta de color azul corresponde a la muñeca. Se empleó una camisa de manga larga para favorecer la segmentación de las marcas de color. Por otro lado, se empleó una cortina de color negro como fondo de imagen y se utilizó luz artificial.



## 6.2. Seguimiento visual de las marcas: Análisis de gestos

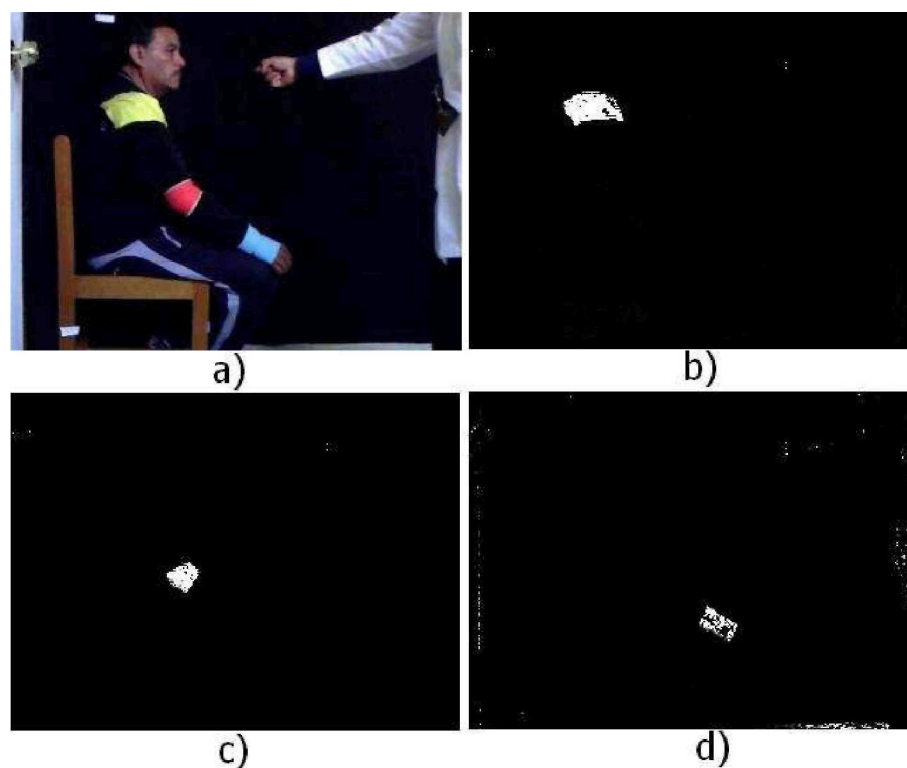


Figura 6.8: Ejemplo del algoritmo de segmentación *Retroproyección de Histogramas*. a) Imagen original, y b)-d), segmentación de las marcas de color amarillo, rojo y azul, respectivamente.

### 6.2.2. Calibración de las cámaras

El procedimiento para obtener los parámetros intrínsecos de las cámaras es básicamente el mismo explicado en la sección 6.1.2. Los parámetros extrínsecos se obtienen manualmente al medir la posición y orientación de la cámara con respecto a un punto de referencia del mundo, como se muestra en la Figura 6.7. En los experimentos se tomó la esquina del soporte de madera que sujeta a las cámaras de video, como el punto del sistema de referencia del mundo. La cámara 1 se encuentra en la posición (X: 0.04, Y: 0.03, Z: 0.13) (en metros) y rotada 315 grados en el eje Y y 180 grados en el eje Z. La cámara 2 se encuentra en la posición (X: 1.53, Y: 0.05, Z: 0.16) y rotada 180 grados en el eje Z.

### 6.2.3. Segmentación

El proceso de segmentación de las marcas de color se basa en el algoritmo *Retroproyección de Histograma*, descrito en la sección 4.3.3. Para obtener los histogramas de color se tomaron una secuencia de imágenes de ambas cámaras para cada una de las marcas de color. Empleando un editor gráfico se generaron imágenes conteniendo solo valores de píxeles pertenecientes al color específico de cada marca. El histograma de color para cada marca se forma contando el número de veces que una combinación de color *H-S* ocurre en la imagen. Los histogramas se dividen en 32 intervalos de clase por cada canal de color con el fin de mejorar su rendimiento. El número total de píxeles muestreados para la marca de color amarillo es igual a 15,081. El número total de píxeles muestreados para la marca de color rojo es igual a 30575. El número total de píxeles muestreados para la marca de color azul es igual a 40315. En la Figura 6.8 se muestra un ejemplo de la segmentación de cada una de las

## 6.2. Seguimiento visual de las marcas: Análisis de gestos

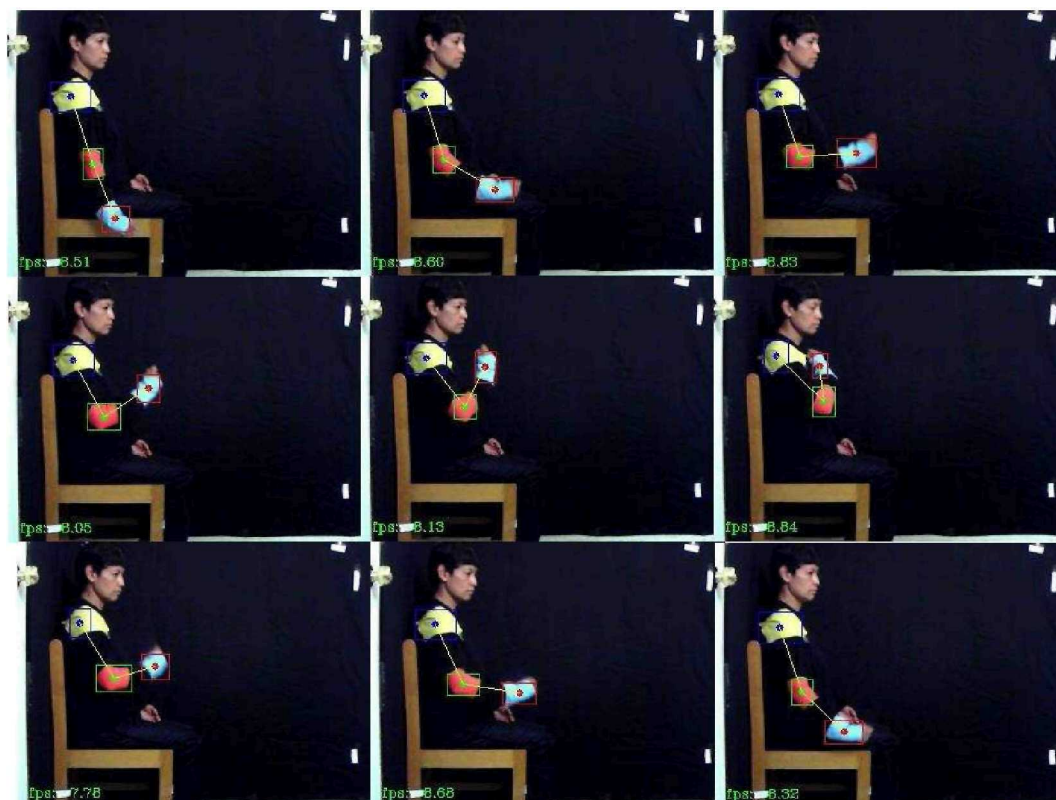


Figura 6.9: Secuencia de imágenes que muestran la ejecución del algoritmo Camshift en cada una de las marcas de color.

marcas de color.

### 6.2.4. Seguimiento 2D

Para el seguimiento de las marcas de color se emplea el algoritmo *Camshift*, descrito en la sección 4.2.3. El algoritmo Camshift se aplica sobre cada una de las marcas de color individualmente. Se emplean tres ventanas de seguimiento de 100 x 100 píxeles cuya posición se modifica de acuerdo al movimiento de las marcas. Un factor importante en el seguimiento de las marcas es la velocidad de desplazamiento. En los experimentos se observó que el algoritmo Camshit pierde el seguimiento de las marcas cuando el movimiento del brazo es rápido. Sin embargo, se debe considerar que el movimiento del brazo de un paciente con derrame cerebral no es demasiado rápido, por lo que el seguimiento se lleva a cabo de forma satisfactoria en desplazamientos moderados. En la Figura 6.9 se presenta una secuencia de imágenes mostrando el funcionamiento del algoritmo Camshift en cada una de las marcas de color. En las pruebas se emplearon camisas de manga larga para favorecer la segmentación de las marcas de color.

### 6.2.5. Reconstrucción 3D

El objetivo de esta etapa es la reconstrucción del brazo de un individuo por medio de una estructura de alambre en el espacio tridimensional, empleando el algoritmo de *Intersección de líneas de visión*, descrito en la sección 4.3.5. La información de seguimiento en el espacio bidimensional de

## 6.2. Seguimiento visual de las marcas: Análisis de gestos

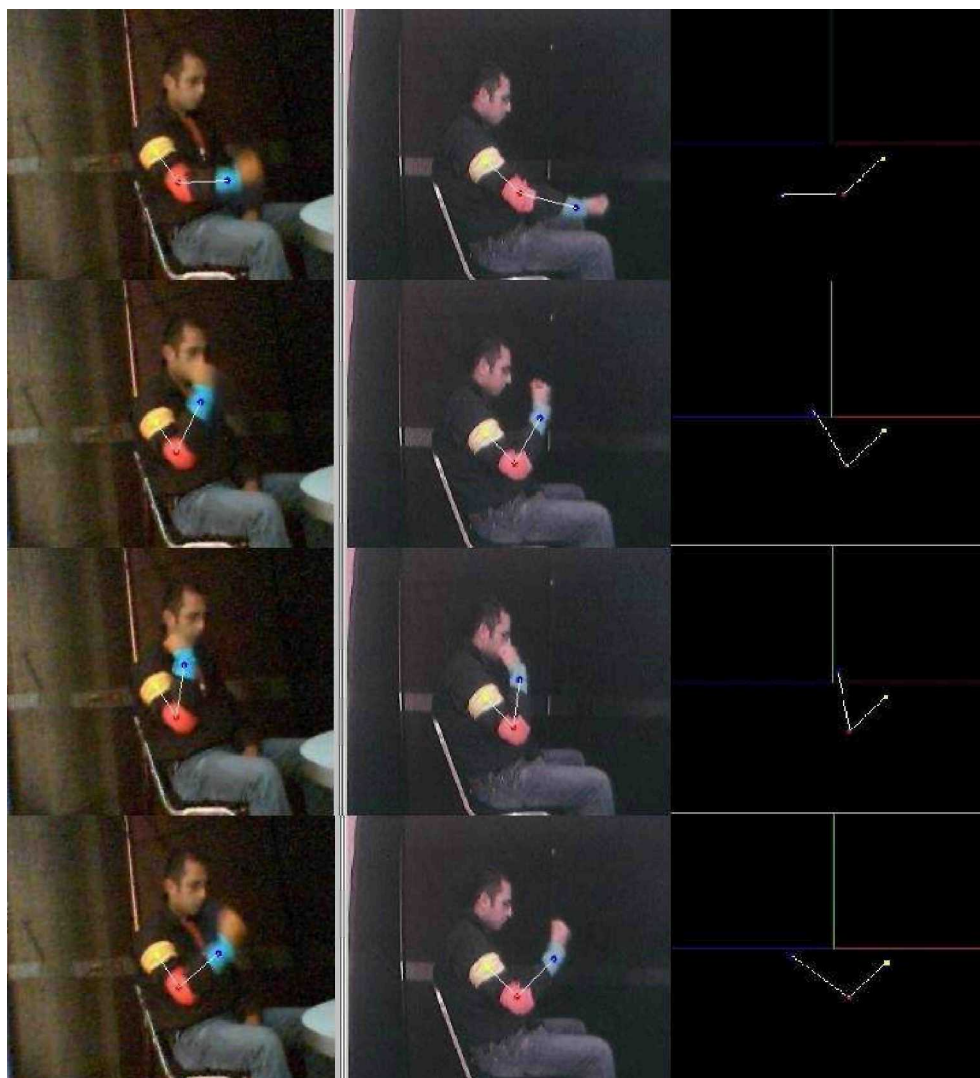


Figura 6.10: Secuencia de imágenes que muestran el seguimiento del brazo en el espacio tridimensional. Las imágenes de las dos primeras columnas muestran el seguimiento de las marcas de color en el espacio 2D, capturadas por las dos cámaras de video. Las imágenes de la tercera columna muestran la reconstrucción del brazo en el espacio tridimensional.

las imágenes de ambas cámaras es utilizada para estimar la posición 3D de cada una de las marcas de color. En los experimentos se observó que la colocación de las cámaras en forma perpendicular producía una pérdida del seguimiento de las marcas debido a oclusiones totales provocadas por el brazo del usuario, por lo que se optó por colocar las cámaras tal y como se ilustra en la Figura 6.7. Esta configuración permite un mayor campo de visión de la escena, permitiendo al usuario ejecutar movimientos del brazo libremente. El usuario se posiciona frente a las cámaras de forma lateral. La cámara 2 se encuentra rotada 45 grados, en sentido inverso a las manecillas del reloj, con respecto al plano XZ en el sistema de referencia del mundo. La distancia entre el usuario y las cámaras es de 1.5 a 2 metros aproximadamente.

La Figura 6.10 muestra un ejemplo de la reconstrucción tridimensional del brazo de un individuo. Las imágenes de las dos primeras columnas muestran el seguimiento en 2D de cada una de las marcas de color capturadas por las dos cámaras de video. Las imágenes de la tercera columna muestran la reconstrucción 3D del brazo por medio de una estructura de alambre. Las marcas de color en el espacio tridimensional se enlazan por medio de líneas de unión.

## 6.2. Seguimiento visual de las marcas: Análisis de gestos

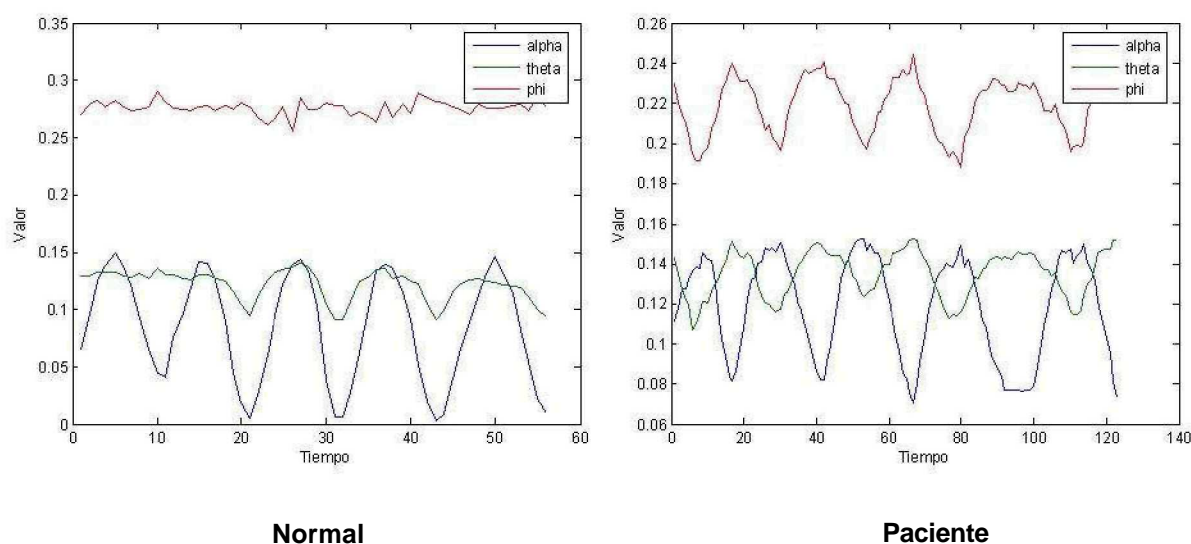


Figura 6.11: La imagen muestra las trayectorias del movimiento del brazo al ejecutar el gesto circular. La imagen de la izquierda muestra la trayectoria del brazo de una persona sana. La imagen de la derecha muestra la trayectoria del brazo del paciente.

### 6.2.6. Análisis de gestos

La etapa de *Análisis de gestos* tiene como objetivo analizar el movimiento de brazo de un paciente que ha sufrido un derrame cerebral. Los experimentos realizados se pueden dividir en dos pasos: el primero es el de entrenamiento de los Modelos Ocultos de Markov y el segundo es el de análisis del movimiento de los gestos considerados. Ambos procedimientos se describen a continuación.

### 6.2.7. Entrenamiento de los Modelos Ocultos de Markov

Para el entrenamiento de los Modelos Ocultos de Markov se empleó un promedio de 10 ejecuciones de secuencias de observaciones por gesto, los cuales se tomaron de dos personas sanas para el grupo *normal*. El entrenamiento del grupo *paciente* fue realizado por un paciente en proceso de rehabilitación, el cual consistió en 5 ejecuciones por cada gesto considerado. Para el entrenamiento de los gestos se pidió a los ejecutantes que realizaran los gestos empleando las formas definidas en la sección 4.3.1. Se utilizó el algoritmo Baum-Welch para entrenar los parámetros de los MOM correspondientes de cada gesto.

### 6.2.8. Análisis del movimiento

El análisis de movimiento consiste en comparar dos gestos iguales ejecutados por un paciente y por una persona sana, empleando una métrica basada en los Modelos Ocultos de Markov. En particular, se empleó la métrica de Kullback-Leibler descrita en la sección 3.2.2. Como se mencionó en la sección 3.2, existen diversos problemas al comparar dos trayectorias correspondientes a un mismo



## 6.2. Seguimiento visual de las marcas: Análisis de gestos

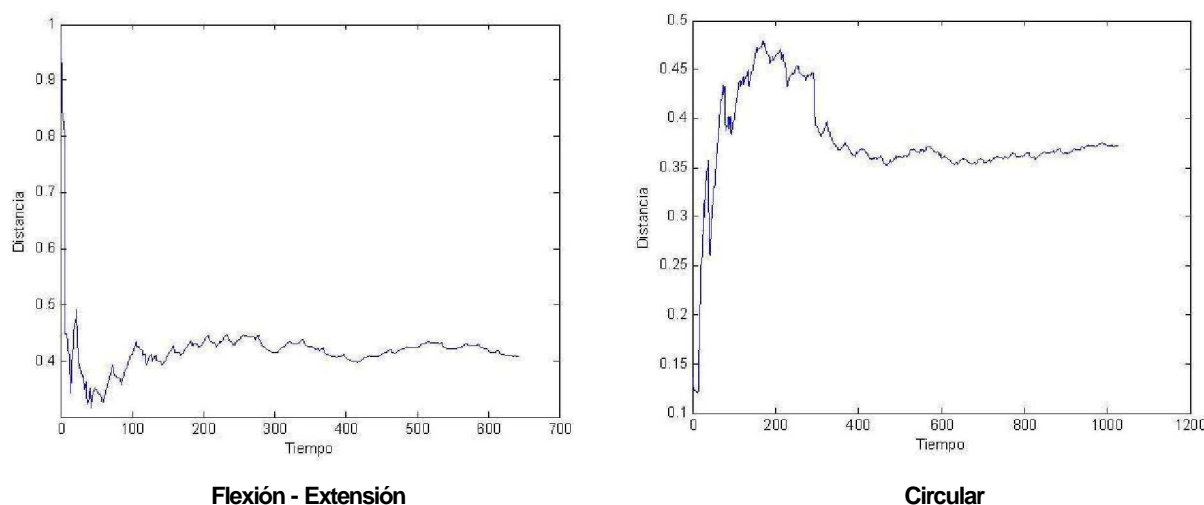


Figura 6.12: La imagen ilustra la distancia de Kullback-Leibler para los gestos de *flexión - extensión* y *circular*, respectivamente, ejecutados por un paciente. La distancia converge en medida que la longitud de la secuencia de observaciones incrementa.

gesto, esto debido a que las trayectorias presentan diferente duración de muestreo y los datos son tomados en distintos instantes de tiempo. En la Figura 6.11 se muestran las trayectorias del movimiento del brazo al realizar el gesto *flexión-extensión*, ejecutadas por una persona sana y por un paciente, respectivamente. Se puede observar claramente que el tiempo de captura de los datos y la variación de los ángulos en ambas trayectorias son diferentes.

En los experimentos realizados, el proceso para comparar dos gestos fué el siguiente. Se pidió al paciente ejecutar nuevamente un gesto, obteniéndose las secuencias de observaciones. Con dichas secuencias, y tomando dos modelos MOM's, uno del grupo *normal* y otro del grupo *paciente* correspondientes al mismo gesto, se aplicó el algoritmo de Kullback-Leibler y se obtuvo su distancia. Si la distancia es cercana a cero, se estima que el gesto realizado por el paciente es similar al gesto del grupo *normal*, por otro lado, si la distancia incrementa se estima que el paciente no ha realizado el gesto adecuadamente. La Figura 6.12 muestra el comportamiento de la distancia de Kullback-Leibler para los gestos *flexión-extensión* y *circular*, ejecutados por un paciente. Se puede observar que la distancia converge en medida que la longitud de la secuencia de observaciones incrementa.

Debido a que el proceso de rehabilitación en pacientes con derrame cerebral es prolongado, no se pudo mostrar el comportamiento de la métrica para evaluar el grado de avance en el proceso de rehabilitación del movimiento del brazo del paciente. Sin embargo, con fines comparativos, se realizaron pruebas con una persona sana aplicando el mismo procedimiento descrito anteriormente para un paciente. En la Figura 6.13 se muestra la distancia al ejecutar los gestos de *flexión-extensión* y *circular* en una persona sana, comparado con la distancia de un paciente. Se puede observar en la gráfica que las distancias de la persona sana son menores a las distancias del paciente. Esto se debe a que la ejecución realizada por la persona sana fué más similar al gesto del grupo *normal*.

El empleo de una métrica basada en los Modelos Ocultos de Markov tiene distintas ventajas. En primer lugar, no se requiere de una normalización en las trayectorias a comparar, ya que si existiera, se perderían las propiedades espacio-temporales de las mismas. En segundo lugar, se pueden modelar clases de trayectorias y parametrizarlas en un modelo, por lo que se pueden tener diferentes clases de trayectorias para ser evaluadas posteriormente. Además, los modelos se pueden reentre-

### 6.3. Resumen

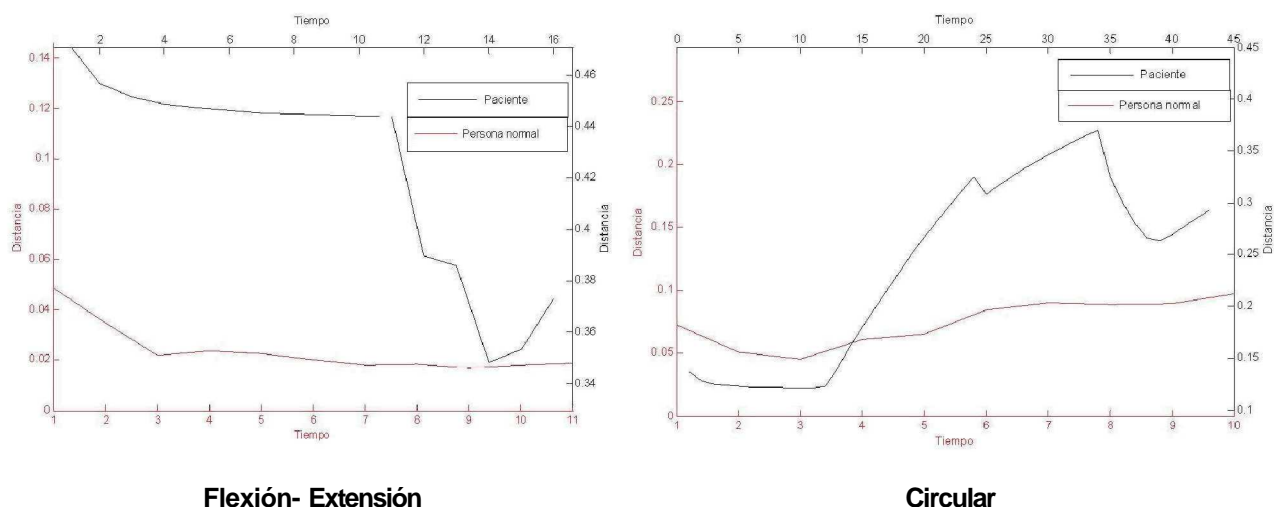


Figura 6.13: La imagen muestra las gráficas comparativas de la distancia de los gestos ejecutados por una persona sana y por un paciente. La gráfica de la izquierda muestra las distancias obtenidas al ejecutar el gesto *flexión-extensión* por una persona sana (gráfica de color rojo) y por un paciente (gráfica de color negro). La gráfica de la derecha muestra las distancias obtenidas al ejecutar el gesto *circular* por una persona sana (gráfica de color rojo) y por un paciente (gráfica de color negro).

nar con nueva información de trayectorias. Por otro lado, una desventaja es la sensibilidad de los parámetros del modelo MOM, ya que es sensible al número de estados y al número de observaciones por estado, provocando que la convergencia de la distancia varíe en la longitud de las secuencias de observaciones. Sin embargo, la distancia de Kullback-Leibler es una buena aproximación a los problemas que se presentan en la comparación de trayectorias, en nuestro caso, al análisis de gestos.

### 6.3. Resumen

En este capítulo se presentaron los experimentos realizados y los resultados obtenidos en la implementación del sistema de seguimiento visual en los dos escenarios descritos en este documento: a) Sistema de seguimiento de la mano: T-WREX, y b) Sistema de seguimiento de las marcas: Análisis de gestos. En el sistema de seguimiento de la mano, se muestran las etapas requeridas para el seguimiento de la mano en el espacio tridimensional. En la fase de segmentación basada en el clasificador bayesiano, se presentaron algunos problemas en la detección de la mano, debido a que algunas regiones presentes en la escena eran segmentadas erróneamente como piel. Sin embargo, bajo ciertas condiciones de iluminación, el clasificador trabaja adecuadamente. Así mismo, se demostró que utilizando ventanas de seguimiento podemos emplear el algoritmo Camshift para el seguimiento de la mano de un individuo a través de una secuencia de imágenes de video. Por otro lado, utilizar información de dos cámaras para reconstruir la posición de la mano en el espacio tridimensional es una técnica sencilla que puede ser utilizada para poder interactuar en ambientes virtuales, como se empleó con la herramienta T-WREX, diseñada para imitar situaciones de la vida real orientadas a la rehabilitación.

En el sistema de seguimiento de las marcas, se realizaron experimentos para analizar el movimiento del brazo de un paciente que ha sufrido un derrame cerebral. Se demostró que, bajo ciertas condiciones de iluminación, los histogramas de color son una técnica útil para la localización de las

### 6.3. Resumen

marcas de color. Además, empleando el algoritmo Camshift se puede seguir las marcas de color en forma individual en las secuencias de imágenes capturadas por dos cámaras de video. Por otro lado, el uso de una estructura de alambre para representar el brazo en el espacio tridimensional es una técnica sencilla para modelar objetos articulados. Finalmente, emplear los Modelos Ocultos de Markov es una buena aproximación para comparar trayectorias de movimiento, ya que estos modelos permiten una representación apropiada a secuencias de datos espacio-temporales.

# Capítulo 7

## Conclusiones

### 7.1. Conclusiones

En esta tesis se propone un sistema de seguimiento tridimensional basado en visión computacional para apoyar en el proceso de rehabilitación en pacientes que han sufrido un derrame cerebral. El sistema de seguimiento visual se probó en dos escenarios diferentes. El primer escenario tiene como objetivo el seguimiento de la mano de un individuo en el espacio tridimensional para interactuar con la herramienta T-WREX. La segmentación de la mano consiste en un clasificador de piel similar al propuesto por Jones [30] para la segmentación de regiones de piel. El seguimiento de la mano se realiza aplicando una variante al algoritmo Camshift propuesto por Bradski [13]. Para estimar la posición de la mano en el espacio tridimensional, se emplea un método basado en intersección de líneas de visión, similar al propuesto por González [23]. Finalmente, la información de la posición de la mano en el espacio 3D es empleada para poder interactuar con los ambientes virtuales de la herramienta T-WREX [44,45], la cual fue diseñada para realizar diferentes tareas imitando situaciones de la vida real orientadas a la rehabilitación.

En el seguimiento de las marcas para el análisis de gestos, se propone un método para representar el brazo de un individuo en una estructura de alambre en el espacio tridimensional, a diferencia del sistema de seguimiento propuesto por Tao [60], el cual sigue las marcas de color en el espacio 2D empleando una sola cámara de video. Además, se propone una métrica basada en modelos probabilistas para la comparación de trayectorias del movimiento del brazo de un individuo. El sistema de seguimiento visual del brazo toma ventaja de los dos sistemas de seguimiento visual presentes en la literatura: Sistemas de visión basados en marcas y Sistemas de visión libres de marcas. Empleando cintas de color se reduce la complejidad de detectar el brazo del usuario, en lugar de utilizar técnicas basadas en modelos cuyo costo computacional es alto. Para la detección de las marcas de color se emplea el algoritmo de Retroproyección de histograma, propuesto por Swain [59], el cual ha demostrado ser una técnica sencilla para la detección de objetos en una imagen. Para el seguimiento de las marcas, se utiliza el algoritmo Camshift, el cual sigue cada marca presente en la escena de forma individual. Finalmente, para la reconstrucción del brazo en una estructura de alambre, se emplea el método de intersección de líneas de visión.

En el análisis de gestos, se propone el uso de ángulos entre las articulaciones extraídas de la estructura de alambre del brazo en el espacio 3D. Los ángulos son invariantes a translaciones y escalamiento, y son independientes de las características físicas de la persona. Además, se propone el uso



## 7.2. Trabajo futuro

de los Modelos Ocultos de Markov para la comparación de trayectorias de movimiento, empleando la distancia de Kullback-Leibler. Emplear los MOM para parametrizar trayectorias de movimiento facilita la comparación de las mismas, ya que no se requiere un procesamiento previo en los datos.

El factor más crítico a la hora de obtener un buen funcionamiento en el sistema de seguimiento visual reside en dar correctamente las posiciones y orientaciones de las cámaras (parámetros extrínsecos). Incluso, un error en la calibración de los parámetros intrínsecos no afecta tanto en la obtención de las líneas de visión de un objeto desde la cámara como lo es la orientación de esta. Sin embargo, la calibración que se ha empleado, a pesar de ser poco precisa, ha demostrado ser satisfactoria.

Los resultados experimentales obtenidos en esta tesis fueron satisfactorios en los dos escenarios considerados. El sistema de seguimiento se implementó correctamente en el software T-WREX, sustituyendo satisfactoriamente al dispositivo robótico. Esto demuestra la factibilidad de desarrollo de este tipo de sistemas para apoyar en el proceso de rehabilitación. Por otro lado, el análisis de gestos demostró ser una buena forma de analizar el movimiento del brazo de un paciente en su progreso de rehabilitación. La métrica para comparar los movimientos del brazo, basada en modelos probabilistas, demostró ser una buena aproximación a los problemas que se presentan en la comparación de trayectorias.

El sistema de seguimiento visual propuesto (considerando ambos escenarios) demostró ser confiable para el seguimiento de la mano y brazo de un individuo en un espacio tridimensional, obteniendo una velocidad de procesamiento de 16 cuadros por segundo. Además, la implementación del sistema es de bajo costo ya que sólo se requieren cámaras de video convencionales, comparado con los sistemas de captura de movimiento humano existentes en el mercado, como son Coda, Qualisys y Vicon [51,56,57], los cuales requieren de dispositivos costosos y de una operación compleja. Por otro lado, la detección y seguimiento correcto del brazo y la mano de un individuo en las condiciones experimentales adoptadas en esta tesis muestran la efectividad del sistema de seguimiento, por lo cual puede ser implementado en otras aplicaciones, como pueden ser en sistemas de Interacción Humano - Computadora, tales como videojuegos, sistemas de navegación, simulación y reconocimiento de gestos.

## 7.2. Trabajo futuro

Dentro de las modificaciones que se pueden hacer al sistema de seguimiento visual propuesto, consideramos las siguientes:

- Mejorar los algoritmos de segmentación en el sistema de seguimiento de la mano. Para esto, consideramos agregar nueva información como textura o bordes, para hacer una representación más robusta de la mano en la escena. Esto reduciría los problemas existentes al detectar regiones similares al color de la piel, así como la pérdida del seguimiento cuando la mano pasa cerca del rostro del usuario.
- Incorporar al sistema un módulo de autocalibración de los parámetros extrínsecos. De esta forma se podrá aislar en mayor medida los resultados de los problemas encontrados en la calibración manual.
- Estudiar nuevas características de movimiento que puedan representar adecuadamente la postura del brazo en el espacio tridimensional.

## 7.2. Trabajo futuro

Implementar otras métricas basadas en modelos probabilistas para mejorar el análisis en el movimiento del brazo de una persona que ha sufrido un derrame cerebral. Además, es necesario realizar un estudio más prolongado en los pacientes para ver su grado de avance en el análisis de gestos orientado a la rehabilitación.

# Referencias

- [1] J. K. Aggarwal and Q. Cai. Human motion analysis: A review. Technical report, Department of Electrical and Computer Engineering, The University of Texas at Austin, 1999.
- [2] American Stroke Association. Disponible en: <http://www.strokeassociation.org/>, 2006.
- [3] H. H. Aviles-Arriaga. Reconocimiento visual de ademanes aplicados a robots móviles. Master's thesis, Tec de Monterrey, Campus Cuernavaca, 2000. (In Spanish).
- [4] H. H. Avilés-Arriaga, L. E. Sucar, and C. E. Mendoza. Visual recognition of similar gestures. *International Conference on Pattern Recognition*, 2006.
- [5] D. H. Ballard and C. M. Brown. *Computer Vision*. Prentice-Hall, 1982.
- [6] A. Baumberg and D. Hogg. An efficient method for contour tracking using active shape models. *Proc. IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, pages 194-199, 1994.
- [7] A. Baumberg and D. Hogg. Generating spatiotemporal models from examples. *IVC*, pages 525-532, 1996.
- [8] A. Bharatkumar, K. Daigle, M. Pandey, Q. Cai, and J. Aggarwal. Lower limb kinematics of human walking with the medial axis transformation. *Proc. of IEEE Workshop on Non-Rigid Motion*, pages 70-76, 1994.
- [9] M. Black, Y. Yacoob, A. Jepson, and D. Fleet. Learning parameterized models of image motion. *Proc. of CVPR*, pages 561-567, 1997.
- [10] A. Blake, R. Curwen, and A. Zisserman. A framework for spatio-temporal control in the tracking of visual contour. *Int. J. Computer Vision*, pages 127-145, 1993.
- [11] R. Bolan, A. Sharma, C. Han, A. Merians, G. Burdea, S. Adamovich, M. Recce, M. Tremaine, and H. Poizner. Virtual reality-based post-stroke hand rehabilitation. *Proc. of Medicine Meets VirtualReality 2002 Conference*, pages 64-70, 2002.
- [12] P. Bourke. The shortest line between two Unes in 3d. disponible en: <http://astronomy.swin.edu.au/pbourke/geometry/lineline3d/>, 2006.
- [13] G.R. Bradski. Computer vision face tracking as a component of a perceptual user interface. *Workshop on Applications of Computer Vision*, 1:214-219, 1998.
- [14] T. Brosnihan, A. Pisano, and R. Howe. Surface micromachined angular accelerometer with force feedback. *Digest ASME inter. Conf. andExp.*, 1995.

## REFERENCIAS

- [15] G. K. M. Cheung, S. Baker, and T. Kanade. Shape-from-silhouette of articulated objects and its use for human body kinematics estimation and motion capture. *ACM SIGGRAPH*, pages 77-84, 2003.
- [16] J. L. Crowley and J. Coutax. Vision for man machine interaction. *Robotics and Autonomous Systems*, 19:347-358, 1997.
- [17] Rehabilitacion despues de un derrame cerebral: Sumario. Disponible en: <http://www.xplain.com/>, 2006.
- [18] M. Falkhausen, H. Reininger, and D. Wolf. Calculation of distance measures between hidden markov models. Technical report, Institut für Angewandte Physik, Johann Wolfgang Goethe Universität, FRG, 1995.
- [19] D. Gavrilu. The visual analysis of human movement: A survey. *Computer Vision and Image Understanding: CVIU*, pages 82-98, 1999.
- [20] Z. Ghahramani. An introduction to hidden markov models and bayesian networks. *International Journal of Pattern Recognition and Artificial Intelligence*, pages 9-42, 2001.
- [21] D. M. Gómez-Allende. *Reconocimiento de Formas y Visión Artificial*. Editorial Ra-Ma, 1993.
- [22] R. C. González and R. E. Woods. *Digital Image Processing*. Adisson-Wesley, 1992.
- [23] P. B. González and J. M. Cañas. Seguimiento tridimensional usando dos cámaras. Technical report, Universidad Rey Juan Carlos, España, Octubre 2004. (In Spanish).
- [24] S.A.V.I Group. Disponible en: <http://ww.cs.toronto.edu/herpers/projects.html>, 28 de mayo de 1999.
- [25] R. Hartley and A. Zisserman. Multiple view geometry in computer vision. *Press Syndicate of the University of Cambridge*, 2000.
- [26] N. R. Howe, M. E. Leventon, and W. T. Freeman. Bayesian reconstruction of 3d human motion from single-camera video. Technical report, MERL - A Mitsubishi Electric Research Laboratory, 1999.
- [27] INNN. Instituto nacional de neurologia y neurocirugia. disponible en: <http://www.innn.edu.mx>. 2006.
- [28] C. Jaynes, S. Webb, R. Steele, and Q. Xiong. An open development environment for evaluation of video surveillance systems. *Proc. of PETS*, 2002.
- [29] G. Johansson. Visual motion perception. *Scientific American*, pages 76-88, 1975.
- [30] M. J. Jones and J. M. Rehg. Statistical color models with application to skin detection. *Proc. Computer Vision and Pattern Recognition, Ft. Collins, CO*, 1:274-280, January 1999.
- [31] B. H. Juang and L. R. Rabiner. A probabilistic distance measure for hidden markov models. *AT&T Technical Journal*, 64:391-408, 1985.
- [32] A. Kourepenis, A. Petrovich, and M. Meinberg. Development of a monotithic quartz resonator accelerometer. *Proc. of 14th Biennial Guidance Test Symp, Hollman AFB, NM*, pages 2-4, 1989.

## REFERENCIAS

- [33] V. Kravtchenko. Tracking color objects in real time. Master's thesis, The University of British Columbia, 1999.
- [34] S. Kullback. Information theory and statistic. *New York: Wiley*, 1958.
- [35] S. E. Levinson, L. R. Rabiner, and M. M. Sondhi. An introduction to the application of the theory of probabilistic function of a markov process to automatic speech recognition. *B. S. T. J.*, 62:1036-1074, 1983.
- [36] OpenCV Open Source Computer Vision Library. Disponible en: <http://www.sourceforge.net/projects/opencvlibrary>, 2006.
- [37] M. McNeill, L. Pokluda, S. McDonough, and J. Crosbie. Immersive virtual reality for upper limb rehabilitation following stroke. *IEEE International Conference on Systems, Man and Cybernetics*, pages 2783-2789, 2004.
- [38] J. A. Montero and L. E. Sucar. Feature selection for visual gesture recognition using hidden markov models. *Fifth Mexican International Conference on Computer Science*, 1:196-203, 2004.
- [39] J. O' Rourke and N. Badler. Model based image analysis of human motion using constraint propagation. *PAMI*, pages 522-536, 1980.
- [40] G. Pajares and J. M. De La Cruz. *Vision por computador: Imágenes y aplicaciones*. Editorial Alfa-Omega, 2002.
- [41] F. Porikli. Trajectory distance metric using hidden markov model based representation. Technical report, Mitsubishi Electronic Research Laboratories, 2005.
- [42] T. Pridmore, S. Cobb, D. Hilton, J. Green, and R. Eastgate. Mixed reality environments in stroke rehabilitation: Interfaces across the real/virtual divide. *International Journal on Disability and Human Development*, 2007.
- [43] L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77:257-286, February 1989.
- [44] D. Reinkensmeyer, L. Kahn, M. Averbuch, A. McKenna-Cole, B. Schmit, and W. Rymer. Understanding and treating arm movement impairment after chronic brain injury : Progress with the arm-guide. *Journal of Rehab. Res. and Develop*, pages 653-662, 2000.
- [45] D. Reinkensmeyer, C. Pang, J. Ñessler, and C. Painter. Web-based telerehabilitation for the upper-extremity after stroke. *IEEE Transactions on Neural Science and Rehabilitation Engineering*, pages 1-7, 2002.
- [46] R. J. Sanchez, J.Liu, S.Rao, P. Shah, R. Smith, T. Rahman, S. C. Cramer, J. E. Bobrow, and D. Reinkensmeyer. Automating arm movement training following severe stroke : Functional exercise with quantitative feedback in a gravity-reduced environment. *IEEE Trans. Neural Sci. Rehabil. Eng*, pages 378-389, 2006.
- [47] C. Schlegel, J. Illman, H. Jaberg, M. Schuster, and R. Worz. Vision based person tracking with a mobile robot. *British Machine Vision Conference*, 1998.
- [48] A. Senior, A. Hampapur, Y. Tian, L. Brown, S. Pankanti, and R. Bolle. Appearance models for occlusion handling. *Proc. of PETS*, December 2001.

## REFERENCIAS

- [49] H. Sidenbladh. *Probabilistic Tracking and Reconstruction of 3D Human Motion in Monocular Video Sequences*. PhD thesis, Dept. of Numerical Analysis and Comp. Sci., Sweden, 2001.
- [50] C. Sminchisescu. *Estimation Algorithms for Ambiguous Visual Models Three-Dimensional Human Modelling and Motion Reconstruction in Monocular Video Sequences*. PhD thesis, Institute National Politechnique de Grenoble (INRIA), July 2002.
- [51] CODA Software. Disponible en: <http://www.charndyn.com/>, 2006.
- [52] CyberGlove software. Disponible en: <http://www.vrealities.com/cyber.html>, 2006.
- [53] DataGlove software. Disponible en: <http://www.5dt.com/>, 2006.
- [54] Dexterous Hand Master software. Disponible en: <http://www.exos.com/>, 2006.
- [55] MT9 software. Disponible en: <http://www.xsens.com/>, 2006.
- [56] Qualisys Software. Disponible en: <http://www.qualisys.com/>, 2006.
- [57] VICON Software. Disponible en: <http://www.vicon.com/>, 2006.
- [58] T. E. Starner. Visual recognition of american sign language using hidden markov models. Master's thesis, MIT . Program in Media Arts and Science, 1995.
- [59] M. J. Swain and D. H. Ballard. Color indexing. *International Journal of Computer Vision*, pages 11-32, 1991.
- [60] Y. Tao and H. Hu. Building a visual tracking system for home-based rehabilitation. *Proc. of the 9 Chinese Automation and Computing Society Conference in the UK, Luton, Eng*, pages 343-448,2003.
- [61] Y. Tao, H. Hu, and H. Zhou. Integration of vision and inertial sensors for home-based rehabilitation. *Proc. of IEEE International Conference on Robotics and Automation, Barcelona, Spain*, pages 106-112,2005.
- [62] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfunder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 780-785, 1997.
- [63] H. Xie and G. Fedder. A cmos z-axis capacitive accelerometer with comb-finger sensing. Technical report, The Robotics Institute, Carnegie Mellon University, 2000.
- [64] H. Zhou y H. Hu. A survey - human movement tracking and stroke rehabilitation. Technical report, Department of Computer Science, University of Essex, 2004.
- [65] H. Zhou and H. Hu. Inertial motion tracking of human arm movements in stroke rehabilitation. *Proc. of IEEE International Conference on Mechatronics and Automation*, pages 1306-1311, 2005.