

INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE MONTERREY



**TECNOLÓGICO
DE MONTERREY®**

SEGUIMIENTO VISUAL ESTEREOSCÓPICO DE INSTRUMENTOS QUIRÚRGICOS EN AMBIENTES CONTROLADOS

Autor:

Juan Paulo Sánchez 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. Juan Frausto Solís

MC. Alejo Mosso Vázquez

Cuernavaca, Morelos. Diciembre de 2007

**SEGUIMIENTO VISUAL ESTEREOSCÓPICO DE
INSTRUMENTOS QUIRÚRGICOS EN AMBIENTES
CONTROLADOS**

Presentada por:

Juan Paulo Sánchez Hernández

Aprobada por:

Dr. Juan Frausto Solís

Profesor del Departamento de Computación,

ITESM Campus Cuernavaca

Asesor de Tesis

M.C. Alejo Mosso Vázquez.

Profesor del Departamento de Computación,

ITESM Campus Cuernavaca

Asesor de Tesis

Dr. Fernando Ramos Quintana

Director del programa de posgrado,

ITESM Campus Cuernavaca

Sinodal

DEDICATORIA

A mis padres por darme la vida.

A Deny Lizbeth, por ser mi mejor amiga.

A esas personas que han tocado mi vida en algún instante de tiempo y que me han ayudado a crecer cada día como ser humano.

AGRADECIMIENTOS.

Agradezco al Tecnológico de Monterrey campus Cuernavaca, por darme el apoyo financiero para realizar mis estudios de posgrado en esta institución.

Agradezco a mi asesor Dr. Juan Frausto, por ser estricto conmigo en la realización de este trabajo, pues eso me permitió mejorar la calidad de este trabajo.

Agradezco a mi asesor M.C. Alejo Mosso, por ser paciente conmigo y ser un guía para la realización de este trabajo y enseñarme cosas que en ocasiones me resultaban difíciles de visualizar.

Agradezco al Dr. José Luis Mosso Vázquez, por las atenciones dadas e invitaciones realizadas su lugar de trabajo, ya que, me permitieron comprender mejor la aplicación de esta tesis.

Agradezco al Dr. Fernando Ramos, por sus atenciones y la oportunidad brindada de seguir estudiando en esta gran institución.

A los amigos que hice durante la realización de este trabajo que me ayudaron a aligerar un poco la carga.

RESUMEN

Esta tesis presenta un sistema que permite reconocer un instrumento quirúrgico de entre un conjunto finito de los mismos. Una vez identificado el instrumento, se calcula la profundidad utilizando la estereoscopia y se obtiene la distancia con respecto a la cámara. Finalmente, se sigue el objeto en el plano de la imagen.

Se abordan tres tareas específicas de la visión computacional: Reconocimiento, localización y seguimiento visual. Cada una de las tareas, es abordada en forma individual. En el reconocimiento, se hace uso de los Modelos Ocultos de Markov, como técnica para el entrenamiento y reconocimiento de los instrumentos quirúrgicos. En la localización, dado que se hace uso de una cámara estereoscópica, se aprovecha dicha característica y se utiliza la estereoscopia, que por triangulación es posible conocer la distancia del instrumento quirúrgico. En el seguimiento visual, se utiliza una técnica ampliamente utilizada y fácil de implementar, la cual es el camshift.

Mediante experimentación se demostraron cada uno de los tres puntos descritos previamente y se comprueba que el sistema es viable.

ÍNDICE:

1. INTRODUCCIÓN	8
1.1 Antecedentes	9
1.2 Definición del problema	11
1.3 Motivación	12
1.4 Objetivos	12
1.4.1 Objetivo General	12
1.4.2 Objetivo específico	13
1.5 Alcances y Limitaciones	13
1.6 Hipótesis	14
1.7 Contribuciones	14
1.8 Organización de la tesis	15
2. TRABAJOS PREVIOS	16
2.1 Contexto del reconocimiento de patrones	17
2.1.1 Robot Asistente instrumentista	18
2.1.2 Comparativa de clasificadores en instrumentos quirúrgicos	19
2.2 Estimación de las coordenadas 3D de los objetos	22
2.2.1 Métodos de localización 3D	23
2.2.1 Visión estereoscópica	25
2.3 Contexto de seguimiento visual	26
2.3.1 Seguimiento 2D de instrumentos quirúrgicos	28
2.3.1.1 Seguimiento con marcas	
2.3.1.2 Seguimiento sin marcas	
2.4 Resumen	32

3. ARQUITECTURA DEL SISTEMA SRLQ	33
3.1 Procesos del sistema SRLQ	34
3.2 Preprocesamiento	36
3.3 Reconocimiento	40
3.4 Localización	40
3.5 Seguimiento	41
3.6 Integración dinámica de la escena	42
3.1 Resumen	42
4. RECONOCIMIENTO, LOCALIZACIÓN Y SEGUIMIENTO	43
4.1 Descripción de contornos con códigos de cadena	44
4.2 Definición y arquitectura de los HMM	47
4.2.1 Resolución al problema de evaluación	51
4.2.2 Aprendizaje de los HMM	53
4.3 Visión Estereoscópica	56
4.3.1 Geometría del sistema estereoscópico	57
4.3.2 Reconstrucción 3D	58
4.4 Seguimiento visual	60
4.4.1 Algoritmo Mean Shift	60
4.4.2 Algoritmo Camshift	61
4. 5 Resumen	63
5. ANÁLISIS DE RESULTADOS	64
5.1 Instrumental quirúrgico y cámara estéreo	65
5.2 Análisis de resultados en reconocimiento	67
5.3 Análisis de resultados en localización	74
5.4 Análisis de resultados en seguimiento	77
5.5 Resumen	80
6. CONCLUSIONES Y TRABAJOS FUTUROS	81
REFERENCIAS	83
A.SISTEMA SRLQ	89

LISTA DE FIGURAS.

Figura 2.1 Asistente instrumentista Penélope 3.0	18
Figura 2.2 Metodología de entrenamiento propuesta por Sossa	20
Figura 2.3 Metodología de evaluación propuesto por Sossa	20
Figura. 2.4 Imagen de instrumento quirúrgico reconocido y localizado.	22
Figura 2.5 Ejemplo de cálculo de distancia por ultrasonido	23
Figura 2.6 Imagen que ejemplifica la combinación de imágenes en el cerebro	25
Figura 2.7 Sistema de visión estéreo montado en humanoide	26
Figura 2.8 Distintas representaciones del seguimiento. a) un punto, b) un conjunto de puntos, c) rectángulo, d) elipse, e) conjunto de elipses, f) forma esquelética, g) h) i) silueta.	28
Figura. 2.9 Imagen de una operación laparoscópica	29
Figura. 3.1 Arquitectura del sistema SRLQ	35
Figura 3.2 Proceso de filtrado.	36
Figura 3.3 Aplicando Filtro Gaussiano	37
Figura 3.4 Ejemplo de discontinuidades. Arriba se muestra una imagen con una discontinuidad y abajo la grafica de su función.	38
Figura. 3.5 Instrumento segmentado con Canny	39
Figura. 3.6 Movimiento de un punto en el plano de la imagen	41
Figura 4.1 Código de 4 direcciones	44
Figura 4.2 Código de 8 direcciones	45
Figura 4.3 Ejemplo de código de cadena de 8 direcciones	46
Figura 4.4 Imagen binaria	46
Figura 4.5 Arquitectura ergódica de los HMM	49
Figura 4.6 Arquitectura Izquierda – Derecha de los HMM	50
Figura 4.7. Arquitectura Izquierda – Derecha paralelas de los HMM	50
Figura 4.8 Ilustración de operaciones del algoritmos Forward	52
Figura 4.9 Ilustración de operaciones por el algoritmo Backward	54
Figura 4.10 Geometría de un par de cámaras en estéreo con ejes ópticos paralelos	57

Figura 4.11 Geometría de dos cámaras en estéreo con ejes ópticos paralelos.	59
Figura 5.1 Instrumental quirúrgico.	65
Figura 5.2 Cámara estéreo BumbleBee	66
Figura 5.3 Instrumento codificado por su forma	67
Figura 5.4 Entrenamiento de separador Farefeuf variando estados.	68
Figura 5.5 Entrenamiento de pinza Metzemaum	69
Figura 5.6 Grupo de instrumentos a identificar	72
Figura 5.7 Instrumento identificado con modelo entrenado del separador Farefeuf	73
Figura 5.8 centro de masa del objeto	74
Figura 5.9 imágenes de Mango Bisturí	75
Figura 5.10 imágenes de pinzas metzenbaum	75
Figura 5.11 pinza Kelly curva.	76
Figura 5.12 Separador Farefeuf.	76
Figura 5.13 imagen binarizada	77
Figura 5.14 Seguimiento de pinza Kelly curva	78
Figura 5.15 Separador Farefeuf.	79
Figura 5.16 pinza Metzenbaum.	79
Figura 5.17 Mango Bisturí	80
Figura A.1 Pantalla que muestra el sistema SRLQ	90
Figura A.2 Menú principal del sistema SRLQ	91
Figura A.3 Barra de herramientas del sistema SRLQ	91
Figura A.4 Áreas de renderizado de las imágenes de captura	91
Figura A.5 Información de salida del sistema	92
Figura A.6 Pantalla del sistema funcionando en línea.	92

LISTA DE TABLAS.

Tabla 2.1 Porcentajes de reconocimiento de los clasificadores	22
Tabla 4.1 Código de cadena 8 direcciones de la Figura 4.4	47
Tabla 5.1 Secuencia de cadena decodigos de 8 direcciones.	68
Tabla 5.2 Valores del modelo entrenado para el separador Farefeuf	70
Tabla 5.3 Valores del modelo entrenado para la pinza Metzembraum	71
Tabla 5.4 Probabilidades dado el modelo entrenado del separador Farefeuf	73
Tabla 5.5 Tabla de distancias calculas por estereoscopia.	77

Capítulo 1

INTRODUCCIÓN

En este capítulo se presenta el contexto en el que se ha desarrollado este trabajo, cabe mencionar que el trabajo se enfoca en el seguimiento de instrumentos quirúrgicos apoyado inicialmente en su reconocimiento previo, así como su localización cartesiana 3D. Se continúa con la descripción general de los problemas a resolver así como la justificación del trabajo. Se plantea el objetivo general y los objetivos específicos, los alcances y limitaciones consideradas, la hipótesis del trabajo, así como una descripción general de las contribuciones de este trabajo de tesis.

1.1. Antecedentes

La disciplina llamada “visión computacional”, ha venido desarrollándose desde los años 1970, en un intento por emular las capacidades humanas. La visión computacional, es un área de la inteligencia artificial y su objetivo principal es la extracción de información relevante de imágenes. La visión computacional intenta emular la visión humana con el fin de extraer información necesaria para caracterizar escenas, reconocer objetos en el entorno, seguirlos y posteriormente tomar decisiones para realizar alguna acción determinada. Sin embargo enfrenta grandes dificultades para realizar dicha emulación.

Algunos de los campos de estudio de la visión computacional son: Segmentación, Reconocimiento de Imágenes, visión Estereoscópica, Seguimiento de Objetos, Mapeo de Escenas, búsqueda de imágenes en bases de datos y reconocimiento de rostros.

El reconocimiento de objetos a partir de una imagen o una secuencia de las mismas, es un problema importante en la visión computacional. La solución del reconocimiento es útil en sistemas que deseen conocer o identificar algún objeto deseado de entre un conjunto de los mismos. A lo largo de 50 años una gran variedad de trabajos han sido publicados relacionados con reconocimiento y sus distintos enfoques [Laveen, 1974][Anil et al, 2000].

La obtención de distancia ha sido una preocupación constante en algunos sistemas, particularmente en los diseñados para navegación en robótica, donde el robot necesita conocer en cada momento la estructura de la escena para navegar en el entorno. Por tal motivo se han desarrollado ciertas técnicas y dispositivos tales como:

1. **Sensores ultrasónicos**, que utilizan la energía acústica.
2. **Triangulación**, una fuente de luz emite un haz con cierto ángulo que al incidir sobre una superficie es reflejado, este haz reflejado es recogido por un detector, al estar la fuente y el detector separados por cierta distancia es posible conocer la distancia del objeto.

3. **Telémetro**, consiste en determinar la distancia de los objetos midiendo el tiempo que tarda la onda en retornar.
4. **Visión estereoscópica**, se refiere a la captura de dos imágenes que tiene la misma escena y por triangulación es posible determinar la distancia.

La visión estereoscópica artificial tomo como referencia el modelo estereoscópico biológico. La posición de los dos ojos en los humanos permite obtener la misma escena del mundo con un ángulo ligeramente diferente, a la diferencia entre estas dos imágenes se le denomina disparidad. La vista de cada ojo se envía por separado al cerebro de tal manera que el cerebro se encarga de combinarlas para producir finalmente una imagen estéreo y de esa forma podemos percibir la sensación de profundidad, lejanía, cercanía de los objetos.

En visión estereoscópica artificial se utilizan dos o mas cámaras separadas entre si a una cierta distancia, generalmente se utilizan dos como en el modelo biológico. El procedimiento consiste en capturar las dos imágenes de una misma escena. Cada imagen es capturada desde una posición de las cámaras ligeramente diferente a las anteriores, de tal manera que las imágenes están ligeramente desplazadas, siendo este el fundamento básico de la estereoscópica, este hecho es el que nos permite la obtención de la distancia de algún determinado objeto.

Sin embargo el seguimiento de objetos, ha merecido una especial atención dentro de la comunidad científica. La capacidad visual que posee por excelencia el ser humano nos permite realizar un sin fin de tareas cotidianas, tales como leer, caminar o conducir un automóvil.

Existe una gran variedad de aplicaciones en el seguimiento visual mediante las técnicas de visión artificial [Matsushita et al, 1998] [Lepetit y Fua, 2005]. Muchas de ellas están inspiradas en los seres humanos y su sistema visual biológico. La navegación robótica es un ejemplo en que la visión artificial es utilizada, al igual que las personas utilizan marcas para caminar y evitar los objetos, los robots pueden utilizar marcas contrastantes

que permitan la navegación. El seguimiento de rostros es otra de las aplicaciones en que ha incursionado el seguimiento visual [Douglas, 1998]. La vigilancia fue una de las primeras aplicaciones en las que se utilizaron algoritmos de seguimiento. Normalmente los sistemas desarrollados constan de cámaras estáticas con módulos de detección que activan la grabación de secuencia.

Es importante mencionar que este trabajo de tesis combina estos tres problemas mencionados anteriormente en un sistema computacional, el cual pretende ser una herramienta para un asistente robot que reconozca, siga y localice un instrumento quirúrgico.

La combinación de los Modelos Ocultos de Markov (HMM, por sus siglas en inglés), para el reconocimiento y algoritmo camshift [Bradsky, 1998], para el seguimiento visual, actualmente no han sido utilizados en los instrumentos quirúrgicos de operaciones laparoscópicas.

1.2. Definición del problema

Lo que se desea resolver en esta tesis es diseñar un sistema computacional para el seguimiento visual, que pueda apoyarse inicialmente en el reconocimiento de algún instrumento quirúrgico deseado, de entre un conjunto finito de instrumentos dispuestos en una mesa y la localización cartesiana en 3D del instrumento reconocido.

Las dificultades a resolver son el reconocimiento de objetos quirúrgicos y la localización 3D en el mundo real, además del seguimiento del mismo. Una vez encontradas las coordenadas 3D del instrumento, el sistema puede ser usado en trabajos posteriores como herramienta computacional de localización continua y reconocimiento.

1.3. Motivación

La visión computacional es un mundo apasionante que nos permite emular los ojos humanos, esos ojos humanos que realizan tareas tan sorprendentes que nos ayudan a comprender de una forma visual lo que nos rodea. Esta tesis intenta proporcionar una herramienta computacional que le permita a algún sistema mecánico distinguir objetos en el mundo real, es la principal motivación, además de contribuir en las aplicaciones de visión computacional que permitan al ser humano realizar sus tareas cotidianas o especializadas de una forma más fácil.

Buscar enfoques prácticos en la visión computacional que sirvan de apoyo en alguna tarea determinada, ya sea en el área médica como aquí se pretende, y aterrizar los métodos, que algunas ocasiones parecen muy abstractos, en aplicaciones tangibles y de gran utilidad, es otra de nuestras motivaciones.

1.4. Objetivos

1.4.1 Objetivo General

El objetivo general de este trabajo es desarrollar un sistema computacional que siga instrumentos quirúrgicos, reconocidos previamente, de entre un conjunto finito de instrumentos quirúrgicos usados en operaciones, así como su localización en un ambiente 3D mediante el uso de la estereoscopia, todo esto en un ambiente controlado.

1.4.2 Objetivos Específicos

Para alcanzar el objetivo general planteado se lograron realizar los siguientes objetivos específicos:

1. Analizar e implementar técnicas para el procesamiento de imágenes.
2. Implementar técnicas de segmentación visual de instrumentos quirúrgicos dentro de la imagen.
3. Implementar Modelos Ocultos de Markov (HMM) para reconocimiento de instrumentos quirúrgicos.
4. Implementar el cálculo de las coordenadas cartesianas 3D, para la localización del instrumento quirúrgico, en base a la estereoscopia.
5. Implementar el seguimiento de los instrumentos quirúrgicos utilizando el algoritmo camshift.

1.5 Alcances y limitaciones

Los alcances del sistema son:

1. Reconocer hasta 4 instrumentos quirúrgicos dispuestos en una mesa.
2. Encontrar las coordenadas cartesianas 3D del centro geométrico del instrumento quirúrgico.
3. Seguir un instrumento quirúrgico con un alcance visual dependiente de las características del hardware de la cámara utilizada.

Las limitaciones son las siguientes:

1. El entorno es controlado, ya sea bajo una iluminación adecuada y bajo un color de fondo contrastante, específicamente colores oscuros.
2. En la localización solo se encuentra la posición del centro geométrico del instrumento quirúrgico.
3. Se asume que la cámara está fija, mientras el objeto se encuentra en movimiento uniforme controlado.
4. La velocidad considerada del instrumento quirúrgico es por debajo de los 4 metros por minuto.

1.6 Hipótesis

La hipótesis planteada en este trabajo es la siguiente: Los HMM y el algoritmo camshift son eficientes y eficaces en el reconocimiento y seguimiento de instrumentos quirúrgicos, respectivamente.

1.7 Contribuciones

La contribución de este trabajo se centra en el diseño de un sistema computacional que pueda seguir un instrumento quirúrgico con una cámara estéreo, que inicialmente el objeto será reconocido para apoyar al seguimiento y proporcione información cartesiana 3D de la localización de un punto en una secuencia de imágenes.

La aplicación de los Modelos Ocultos de Markov enfocado en el reconocimiento de instrumentos quirúrgicos, como apoyo inicial al seguimiento de los mismos es una de las contribuciones de esta tesis, además de la localización continua de dicho instrumento quirúrgico utilizando la cámara estéreo

1.8 Organización del documento

El presente documento está estructurado en 8 capítulos, como sigue:

En el **Capítulo 2** veremos un breve estado del arte de los trabajos más relacionados con el tema de tesis, enfocados al reconocimiento, localización y seguimiento de instrumentos quirúrgicos.

En el **Capítulo 3** se presenta la arquitectura a seguir en el desarrollo de la presente tesis. Se aborda la arquitectura propuesta para el reconocimiento, localización y seguimiento de instrumentos quirúrgicos; se presenta una arquitectura que permite el desarrollo de esas funciones.

En el **capítulo 4** se describen, para el reconocimiento de instrumentos quirúrgicos, los Modelos Ocultos de Markov así como su implementación y validación. En este mismo capítulo se describe la técnica de cálculo de las coordenadas cartesianas en 3D, utilizando la estereoscopía. Una vez obtenido el reconocimiento de los instrumentos y sus coordenadas cartesianas 3D, la fase final es seguir dicho instrumento. Este capítulo describe el seguimiento de los instrumentos quirúrgicos en base al camshift.

En el **Capítulo 5** se discute los experimentos desarrollados de las tres técnicas propuestas, con lo cual comprobamos la hipótesis planteada inicialmente.

En el **Capítulo 6** se muestra las conclusiones de la tesis desarrollada, así como los trabajos futuros.

Capítulo 2

TRABAJOS PREVIOS

En este capítulo se realiza una revisión del estado actual de la investigación enfocado en el seguimiento visual 2D de los instrumentos quirúrgicos, esto quiere decir que solo es en el plano de la imagen. Este estudio tiene particular interés en el seguimiento de objetos quirúrgicos, el reconocimiento de los mismos y su localización cartesiana 3D, por tal motivo se analizan trabajos ya desarrollados en estas tres tareas específicas.

Este capítulo tiene tres secciones importantes en las cuales se exploran, en cada una de ellas, los trabajos relacionados al tema de tesis. Inicialmente exploramos los trabajos desarrollados en reconocimiento de los instrumentos quirúrgicos. Posteriormente abordamos las técnicas más relevantes en localización 3D de objetos. Para finalizar, abordamos el seguimiento visual que tiene particular atención en esta tesis, en ella, revisamos brevemente aquellos trabajos que se apeguen al seguimiento de instrumentos quirúrgicos. El sistema que realiza estas tres funciones Sistema de Reconocimiento, Localización y Seguimiento de instrumentos quirúrgicos (SRLQ).

2.1 Contexto del reconocimiento de patrones

El reconocimiento de patrones, se refiere a la clasificación de objetos en ciertas clases. Dependiendo de la aplicación, los objetos pueden ser imágenes, señales o cualquier otro tipo de información que pueda ser clasificada. El reconocimiento de patrones, se clasifica en tres tipos de enfoques: Enfoque Neuronal, enfoque sintáctico, enfoque estadístico.

El enfoque neuronal data de los años 40 a partir del trabajo de [McCulloch y Pitts, 1943] donde se propusieron modelos de neuronas en la forma de dispositivos binarios basados en un umbral y algoritmos estocásticos que implicaban cambios binarios 0-1 y 1-0 en estados de las neuronas como la base para el modelado de sistemas neuronales. La idea de este enfoque es muy sencilla. Consiste en cada una o mas muestras del conjunto de objetos a aprender, ajustar de forma iterada los pesos internos de un arreglo de pequeños procesadores interconectados entre ellos llamados neuronas. Cuando el llamado error de aprendizaje ha descendido a un valor inferior a un umbral dado se dice que la red neuronal ha sido entrenada. Durante la etapa de reconocimiento una muestra de alguno de los objetos aprendidos es presentada en la entrada de la red entrenada, produciendo como salida el índice de la clase correspondiente.

El enfoque sintáctico estructural, por otra parte, consiste en descomponer cada uno de los objetos en un conjunto de partes representativas que forman parte de un alfabeto. Un conjunto de reglas de producción permiten conectar las demás partes para dar forma a cada uno de los objetos. Durante la etapa del reconocimiento a partir de una imagen de uno de los objetos previamente aprendidos, se busca encontrar en dicha imagen partes de dicho objeto, el conjunto de reglas de producción junto con un autómata permiten reconstruir la estructura del objeto.

Por ultimo, el enfoque estadístico, consiste en una etapa de entrenamiento y a partir de un conjunto de muestras de un objeto a reconocer se calcula un vector de rasgos representativos que nos permitan describir el objeto.

El reconocimiento de patrones ha tenido una variedad de aplicaciones y enfoques distintos [Kanal, 1974] [Campbell y Flynn, 2001]. Algunos de ellos con propuestas muy interesantes, sin embargo, este trabajo tiene como principal enfoque el reconocer patrones en base a una imagen o una secuencia de ellas. Los objetos a reconocer son los instrumentos quirúrgicos, por tal motivo, este apartado enfocara los trabajos relacionados en reconocimiento de dichos instrumentos y las diferentes propuestas más interesantes.

2.1.1 Robot Asistente instrumentista

Actualmente se han desarrollado varios sistemas robots que sirven como asistentes instrumentistas, como [Penelope, 2005], este sistema reconoce los instrumentos quirúrgicos solicitados por el medico por medio de la voz, posteriormente Penélope identifica los instrumentos quirúrgicos, los toma y se los da al medico. La empresa que los desarrolla, Robotic Surgical Tech Inc, ya tiene varias versiones desarrolladas, como Penélope 3.0 (Figura 2.1) y actualmente cuentan con la versión Penélope 4.0.



Figura 2.1 Asistente instrumentista Penélope 3.0

2.1.2 Comparativa de clasificadores en instrumentos quirúrgicos

[Sossa et al, 2005], propone el desarrollo de una metodología sencilla para el reconocimiento y la localización (en el plano de la imagen) de un conjunto de cuatro instrumentos quirúrgicos. En dicho trabajo realizan una comparativa entre un conjunto de nueve clasificadores para reconocer dichos instrumentos quirúrgicos. El trabajo está clasificado en dos problemas principales:

Problema 1: Reconocimiento. Determinación de forma automática la identidad del objeto, de entre un conjunto de ellos.

Problema 2: Localización. Determinar automáticamente el centro geométrico del instrumento, la dirección de su eje mayor, así como las coordenadas de los centros de los agujeros de toma. Para resolver el primer problema ellos realizan una etapa de entrenamiento (figura 2.2).

Como paso inicial, binarizan cada instrumento quirúrgico, posteriormente extraen las componentes conectadas y eliminan el ruido de las componentes para obtener un vector de rasgos característicos de cada instrumento, para finalizar, dicho vector de rasgos es utilizado para entrenar cada uno de los clasificadores.

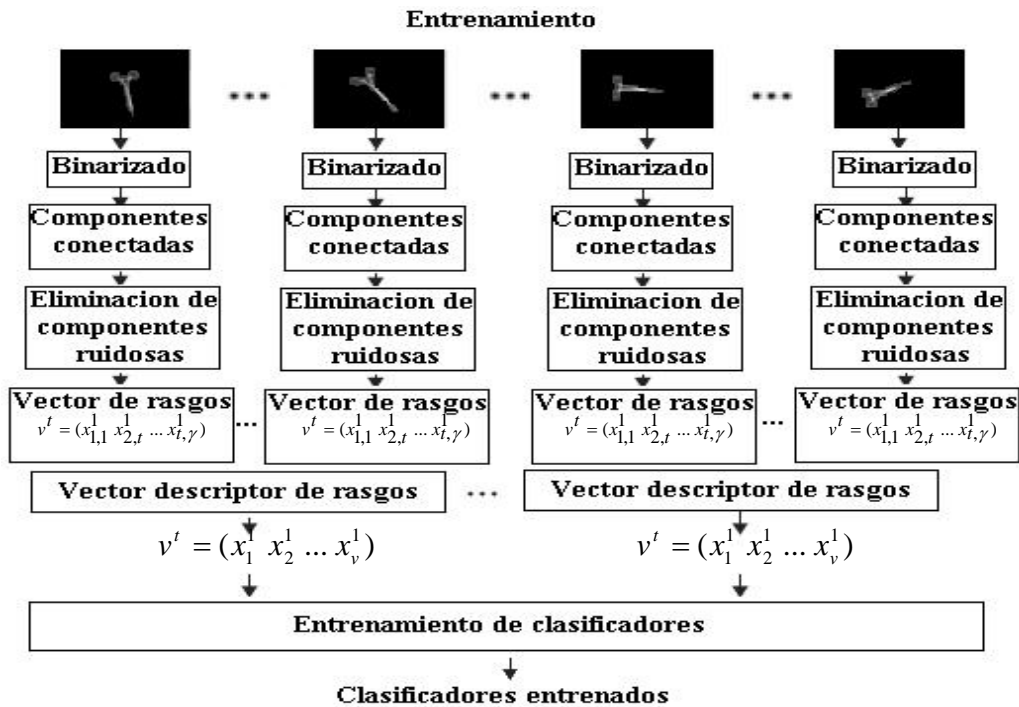


Figura 2.2 Metodología de entrenamiento propuesta por Sossa [Sossa et al, 2005]

Una vez entrenado con cada clasificador propuesto, En el segundo problema, evalúan el porcentaje de reconocimiento con cada uno de los clasificadores propuestos (Figura 2.3).

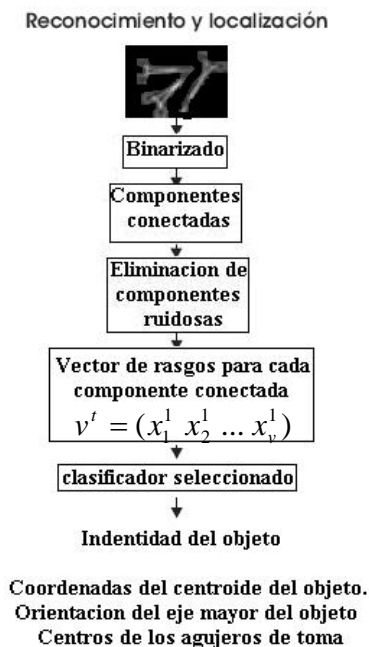


Figura 2.3 Metodología de evaluación propuesto por Sossa [Sossa et al, 2005]

Cada clasificador utilizado en la fase de entrenamiento y evaluación mostró distintos comportamientos, entre los que destaca el enfoque Bayesiano, pues es el que mejor resultados de clasificación les dio, con un 96% en promedio como muestra la Tabla 2.1.

Porcentaje de reconocimiento para cada uno de los clasificadores propuestos									
Pinza Tipo	Manhattan modificado	SSE modificado	Basado en el ángulo	Distancia Chi	Distancia de Canberra	Coficiente de correlación	Manhattan	Euclideo	Bayesiano
1	98%	91%	85%	89%	98%	83%	88%	88%	100%
2	75%	76%	76%	76%	76%	76%	76%	76%	99%
3	97%	97%	98%	99%	99%	97%	98%	97%	99%
4	98%	86%	75%	94%	95%	67%	79%	74%	89%
% total	92%	87.5%	83.5%	89.5%	92%	80.75%	85.25%	83.75%	96.75%

Tabla 2.1. Porcentajes de reconocimiento de los clasificadores [Sossa et al, 2005].

El segundo problema ellos lo resuelven utilizando los momentos geométricos, los momentos de orden cero y uno (Ecuación 2.1) para encontrar el centro geométrico de los instrumentos quirúrgicos y localizar los centros de los agujeros de toma.

Centroide del objeto: Las coordenadas (x,y) , del centroide de la región R conectada binaria de píxeles se puede obtener a partir de momentos geométricos estándar. (ver ecuación 2.1).

$$x = \frac{m_{10}}{m_{00}} = \frac{\sum_{x,y \in R} xf(x,y)}{\sum_{(x,y) \in R} f(x,y)} \quad y = \frac{m_{01}}{m_{00}} = \frac{\sum_{(x,y) \in R} yf(x,y)}{\sum_{(x,y) \in R} f(x,y)} \quad (2.1)$$

Los momentos de segundo orden (Ecuación 2.2), lo utilizan para encontrar su ángulo de orientación como muestra la figura 2.4.

Angulo de orientación y dirección del eje mayor: El ángulo de orientación con respecto al eje x de la región R se obtiene con los momentos geométricos de segundo orden.

$$\phi = \frac{1}{2} \tan^{-1} \left(\frac{2l_{11}}{l_{20} - l_{02}} \right) \quad (2.2)$$

Donde:

$$l_{11} = m_{11} - ym_{10}$$

$$l_{20} = m_{20} - xm_{10}$$

$$l_{02} = m_{02} - ym_{01}$$

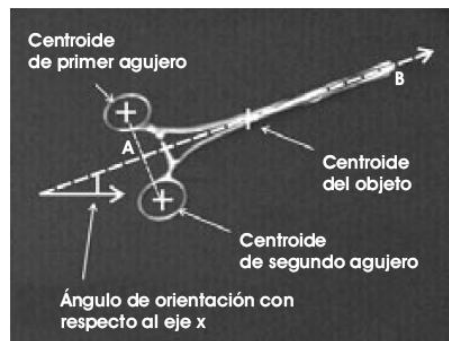


Figura. 2.4 Imagen de instrumento quirúrgico reconocido y localizado. [Sossa et al, 2005]

2.2 Estimación de las coordenadas 3D de los objetos

El cálculo de distancias de objetos, es de gran importancia puesto que nos permite conocer las distancias reales con respecto al dispositivo que se utilice. Dicho cálculo de la distancia ha sido una preocupación constante en algunos sistemas, tales como la navegación robótica, los cuales necesitan conocer en cada momento la estructura de la escena para navegar en el entorno.

En esta sección abordamos algunos de los enfoques más relevantes para la obtención de las coordenadas 3D de los objeto.

2.2.1 Métodos de localización 3D

Existen varios tipos de métodos para obtener la distancia y son clasificadas en dos grupos:

- Directas y Activas.
- Pasivas.

Las técnicas directas y pasivas se basan principalmente en el cálculo del tiempo que tarde la señal en viajar, que puede ser ultrasónica o electromagnética. Algunas de las técnicas que tienen especial importancia son las siguientes:

- a) **Ultrasonidos:** Los ultrasonidos son antes que nada sonidos, exactamente igual a los que oímos normalmente, solo que tienen una frecuencia mayor que la máxima audible por el oído humano. Estas frecuencias tienen un rango de entre los 16 KHz hasta 20 KHz.

El funcionamiento básico de los ultrasonidos como medidores de distancia se muestra en la figura 2.5, donde se tiene un transmisor que emite un pulso de ultrasonido que rebota sobre un determinado objeto y la reflexión de ese pulso es determinada por un receptor de ultrasonido.

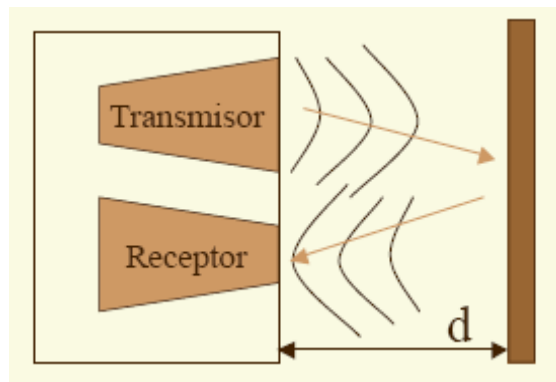


Figura 2.5 Ejemplo de cálculo de distancia por ultrasonido

La mayoría de los sensores de ultrasonido de bajo costo se basan en la emisión de un pulso de ultrasonido cuyo lóbulo, o campo de acción, es de forma cónica. Midiendo el tiempo que transcurre entre la emisión del sonido y la percepción del eco se puede establecer la distancia a la que se encuentra el obstáculo que ha producido la reflexión de la onda sonora, mediante la fórmula:

$$d = \frac{1}{2} V.t \quad (2.3)$$

Donde V es la velocidad del sonido en el aire y t es el tiempo que transcurre entre la emisión y recepción del pulso.

- b) **Luz estructurada.** Esta técnica consiste en la proyección de una configuración de luz (laser). La distorsión de la luz es captada por una cámara de visión y con ello se puede obtener el alcance.
- c) **Triangulación.** Una fuente de luz emite un haz con un cierto ángulo, que al incidir sobre una superficie es reflejado, este haz reflejado, es recogido por un detector, al estar la fuente y el detector separados por una cierta distancia se obtiene un triángulo con el haz incidente y reflejado con lo que se puede conocer la distancia del objeto.

Por otro lado las técnicas pasivas, que son basadas en imágenes y son de dos tipos:

- a. Análisis de una sola imagen. Las cuales realizan algún tipo de análisis sobre una sola imagen tales como: Análisis de gradiente de la textura, tamaño de objeto y análisis del enfoque.
- b. Análisis de más de una imagen, que son utilizadas para calcular la distancia de un objeto en la escena a través de técnicas basadas en la disparidad estéreo, el flujo óptico y mediante triangulación.

De las dos técnicas mencionadas la que tiene un mayor rango de aplicación son las pasivas puesto que no requiere de una fuente de energía artificial y puede ser transportado a exteriores. Las desventajas que presentan los métodos pasivos es que son computacionalmente complejos y son muy sensibles a cambios en las condiciones de la escena.

2.2.1 Visión Estereoscópica

Una de las técnicas pasivas, mencionadas en el cálculo de la distancia utilizando dos o más imágenes, es la visión estereoscópica, la cual se basa en la triangulación entre un punto de la escena y por lo menos dos proyecciones de este punto en dos imágenes tomadas desde distintas perspectivas, esta técnica es una de las más utilizadas para la reconstrucción tridimensional.

La visión estereoscópica se basa en el modelo visual propio del ser humano, esto es, los seres humanos tenemos dos ojos localizados en cada lado de la cabeza, cada ojo, obtiene una vista de la escena con un ángulo ligeramente diferente esta diferencia se le denomina disparidad. La imagen de cada ojo se envía al cerebro el cual realiza un proceso de combinación, emparejando las similitudes y añadiendo las diferencias y de esta forma producir una imagen en estéreo y de esa forma podemos percibir la profundidad, la lejanía o cercanía de los objetos, tal como se muestra en la figura 2.6.

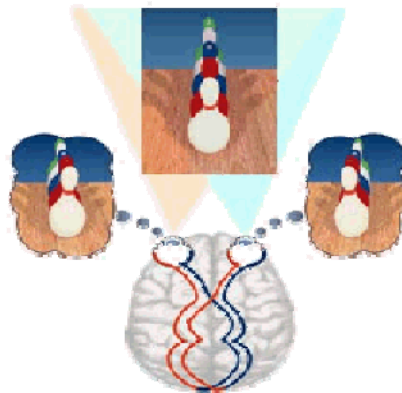


Figura 2.6 Imagen que ejemplifica la combinación de imágenes en el cerebro.

Algunos trabajos utilizan la visión estéreo para la navegación en robot, como el propuesto por [Thompson y Kagami, 2005] en el que montan un sistema estéreo en un robot humanoide el cual le permite navegar en un ambiente con ruido utilizando marcas colocadas en el piso, las cuales el robot humanoide puede identificar.

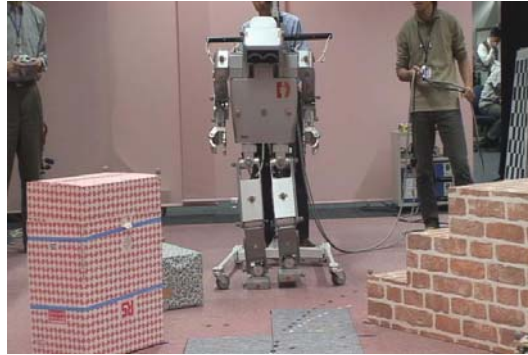


Figura 2.7 Sistema de visión estéreo montado en humanoide [Thompson y Kagami, 2005]

2.3 Contexto de seguimiento visual

El seguimiento visual de objetos es una tarea importante dentro de la visión computacional. Los avances tecnológicos en hardware de computadora, la disponibilidad de cámaras de video con mejor resolución y velocidad han generado una creciente necesidad por aumentar los algoritmos de seguimiento de objetos. El uso del seguimiento de objetos se aplica en diferentes tareas, tales como:

- El reconocimiento basado en el movimiento, que significa la identificación de humanos basado en el movimiento del caminar [Jia-Ching y Moura, 1998].
- La vigilancia automatizada, que es monitorear la escena para detectar actividades sospechosas [Sandoval, 2003].
- La interacción humano-computadora, en la que se reconocen gestos para comunicarse de forma natural con la computadora [Anjum y Aggarwal, 2001].
- Navegación de vehículos planeando rutas y evadir obstáculos [Deans et al, 2005].

El seguimiento de objetos, quiere decir, que en una secuencia de video, el objeto estará etiquetado en diferentes imágenes de video sin perderlo de vista. Sin embargo el seguimiento también tiene sus complejidades, como ejemplo:

- Ruido en la imagen,
- Movimiento bruscos del objeto,
- Oclusiones parciales o totales,
- Cambios de iluminación en la escena y
- Requerimientos de procesamiento en tiempo real.

En la actualidad numerosas aproximaciones han sido propuestas en el seguimiento de objetos , sin embargo muchas de ellas son aplicaciones a tareas específicas o para una variedad de escenarios particulares. Algunas de las aproximaciones [Aggarwal y Cai, 1999] son enfocadas al seguimiento de articulaciones de los humanos y al seguimiento de objetos rígido.

En el seguimiento visual, existen varias representaciones para poder seguir el objeto de interés, algunas de las representaciones comunes para seguimiento visual son las siguientes:

- Los objetos son representados por un punto, que es el centroide del objeto, o por un conjunto de puntos [Veeman et al, 2001].
- Por objetos geométricos primitivos, como rectángulos o elipses [Comaniciu et al, 2003], son otro de las representaciones comunes.
- La silueta también es utilizada, la cual está definida por el contorno del objeto a seguir [Yilmaz et al, 2004].
- Un modelo esquelético del objeto también es muy empleado [Ballard y Brown, 1982].

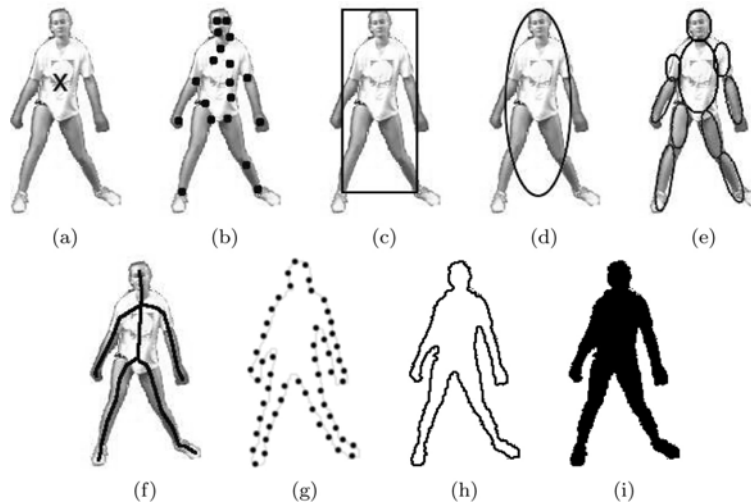


Figura 2.8. Distintas representaciones del seguimiento. a) un punto, b) un conjunto de puntos, c) rectángulo, d) elipse, e) conjunto de elipses, f) forma esquelética, g) h) i) silueta.

2.3.1 Seguimiento 2D de instrumental quirúrgico

En esta sección se analizan trabajos sobre seguimiento de instrumentos quirúrgicos en operaciones laparoscópicas y en cirugía abierta. De acuerdo a la bibliografía explorada en seguimiento de instrumentos quirúrgicos en operaciones laparoscópicas, dichos trabajos se han enfocado en el entorno físico del paciente, para seguimiento 2D, es decir, en el plano de la imagen. En esta sección exploramos dichos trabajos, sin embargo se debe aclarar que el enfoque de este trabajo es el seguimiento visual de los instrumentos quirúrgicos fuera del entorno físico del paciente.

2.3.1.1 Seguimiento con marcas de instrumentos laparoscópicos

La cirugía laparoscópica es una técnica quirúrgica mínimamente invasiva. El procedimiento se refiere a que el cirujano inserta en el paciente el instrumento requerido junto con el laparoscopio a través de múltiples incisiones y realiza la operación observando una pantalla que muestra las imágenes capturadas por el laparoscopio. El

problema fundamental del uso de esta técnica, radica en el entendimiento entre el cirujano y el asistente de la cámara, pues el asistente puede cansarse de sostener el laparoscopio, debido a que las operaciones normalmente son largas o no comprender una orden del cirujano.



Figura. 2.9 Imagen de una operación laparoscópica

Para solucionar dicho problema se han desarrollado varios sistemas de asistencia robotizados [Hurteau et al, 1994] [Taylor et al, 1995] que utilizan procesamiento de imágenes para mantener siempre centrado el instrumento en la pantalla. El proceso para mantener siempre centrado el objeto de interés es seguirlo en el plano de la imagen. El seguimiento de los instrumentos quirúrgicos en operaciones laparoscópicas inicialmente se había hecho con marcas [Wei et al, 1997], dichos sistemas segmentaban el color de las marcas adheridas al instrumentos, localizarlo visualmente y de esa manera seguirlo.

Uno de los trabajos enfocados en el seguimiento de los instrumentos en el entorno físico del paciente, utilizando marcas, es el propuesto por [West y Maurer, 2004], su enfoque es colocar varias marcas en el instrumento quirúrgico que emita luz infrarroja y de esa forma ellos puedan localizar (de forma visual) y seguir dicho instrumento en una operación.

Otra propuesta interesante, en el que utilizan un tipo de marca, es el propuesto por [Krupa et al, 2002], en el que desarrollan un sistema computacional automático que

recupera la posición del instrumento quirúrgico en una cirugía laparoscópica. Ellos colocan en el efector de un cirujano robot el instrumento quirúrgico y la localización visual la realizan utilizando los puntos de las incidencias del láser en la escena. De esa manera segmentar dichos puntos de color discriminante con la escena y centrar la imagen con respecto a las coordenadas de los puntos de incidencia del láser.

2.3.1.2 Seguimiento sin marcas

El seguimiento sin marcas es un enfoque que evita colocar algún tipo de marca visual a instrumentos quirúrgico para ser seguidos con el endoscopio, ya que las marcas pueden resultar una desventaja, pues pueden entorpecer el manejo de los instrumentos quirúrgicos en la operación.

Un trabajo desarrollado por [Climent y Mares, 2004], en donde ellos desarrollan un sistema de seguimiento de objetos quirúrgicos, en base a una técnica muy sencilla basada en encontrar las líneas estructuradas, mediante la transformada de Hough [Hough, 1962]. La transformada de Hough, es una técnica muy conocida para la detección de líneas rectas y otras figuras geométricas primitivas. El sistema que ellos implementan no necesita marcas ni alguno color específico para su localización. Este sistema localiza y sigue los instrumentos quirúrgicos en una operación de cirugía laparoscópica.

Básicamente la idea de este sistema se basa en que los instrumentos son los objetos más estructurados en la imagen laparoscópica. El uso de la transformada de Hough es para detectar las líneas estructuradas en la escena, ellos afinan el sistema con el movimiento del instrumento quirúrgico para distinguirlo de otra forma estructurada.

Otro de los trabajos relacionados al seguimiento de instrumentos quirúrgicos sin marcas, es el propuesto por [McKenna et al, 2004], en dicho trabajo ellos proponen seguir los instrumentos quirúrgicos utilizando el color de la punta de dicho instrumento. Para realizar la clasificación entre instrumento y no instrumento ellos proponen utilizar el

clasificador Bayesiano. A continuación explicamos brevemente dicha propuesta, esto debido a que el seguimiento que realizamos en esta tesis es similar:

Dado un píxel γ en valores RGB, la probabilidad a posteriori de un instrumento quirúrgico de este píxel se puede modelar con la regla de Bayes. (ver ecuación. 2.4):

$$P(\tau | \gamma) = \frac{P(\gamma | \tau)P(\tau)}{P(\gamma | \tau)P(\tau) + P(\gamma | \neg\tau)(1 - P(\tau))} \quad (2.4)$$

Donde τ y $\neg\tau$ son la clase instrumento y no instrumento, respectivamente.

De esta forma ellos logran segmentar el instrumento quirúrgico, en base al histograma de color que representan los píxeles de dicho instrumento quirúrgico. Una vez identificado el instrumento en la escena utilizan los filtros de partículas para estimar los estados posteriores en el tiempo y así poder realizar el seguimiento visual de dichos instrumentos quirúrgicos.

McKenna y sus colaboradores reportan buenos resultados además que presentan una nueva estrategia para el seguimiento de los instrumentos quirúrgicos sin necesidad de utilizar algún tipo de marcas, en cuestión de eficiencia su propuesta es muy buena para correr en tiempo real.

2.4 Resumen

El sistema SRLQ comprende las funciones de Reconocimiento, Localización y Seguimiento de instrumentos quirúrgicos. En este capítulo se han descrito los fundamentos de las tres tareas de SRLQ las cuales individualmente tienen una gran variedad de aplicaciones potenciales de enorme interés tanto científico como tecnológico. En el capítulo también se revisaron los trabajos con mayor relación al sistema propuesto en esta tesis.

Capítulo 3

ARQUITECTURA DEL SISTEMA SRLQ

Una vez descrito los componentes del sistema SRLQ, en este capítulo se presenta en este capítulo su arquitectura para realizar cada una de las tareas que lo conforman: Reconocimiento, Localización y Seguimiento de los instrumentos quirúrgicos.

Abordaremos cada uno de los procesos de la arquitectura del sistema SRQL, algunos procesos son abordados detalladamente, sin embargo, otros son solo explicados brevemente debido a su importancia en esta tesis.

3.1 Procesos del sistema SRLQ

La arquitectura de SRLQ, está conformada por los elementos dentro del cuadro de la figura 3.1. La entrada al sistema SRLQ es una imagen o una secuencia de imágenes de instrumentos quirúrgicos, las cuales llegan al sistema como un conjunto de píxeles en escala de gris.

El sistema SRLQ contiene ocho procesos, mediante los cuales se realizan las tres tareas principales de reconocimiento, localización y seguimiento. La descripción de cada uno se describe en las subsecciones 3.2- 3.6.

En cada proceso se realizan diferentes tareas, los procesos uno y dos constituyen lo que se denomina preprocesamiento, el cual tiene como función la limpieza de la imagen y segmentación de la misma; la tarea de preprocesamiento se conoce también como visión de bajo nivel . Los procesos tres, cuatro y cinco constituyen la tarea de reconocimiento de SRLQ, es decir realizan el reconocimiento de los instrumentos quirúrgicos representados en la entrada. Esta tarea de reconocimiento requiere de las fases de codificación de las observaciones (proceso tres), la de entrenamiento (proceso cuatro) y la de reconocimiento (proceso cinco). En el proceso seis se realiza la tarea de la localización que consiste en el cálculo de las coordenadas 3D del instrumento quirúrgico. En el proceso siete, se aborda el seguimiento de los instrumentos quirúrgicos de la escena proporcionada en la entrada como una secuencia de imágenes. El proceso ocho, finalmente, se realiza la integración de los procesos anteriores.

El sistema SRLQ tiene como salida la Evaluación del Objetivo General (ver figura 3.1), es decir, desarrollar un sistema computacional que siga instrumentos quirúrgicos, reconocidos previamente, de entre un conjunto finito de instrumentos quirúrgicos, así como su localización en un ambiente 3D.

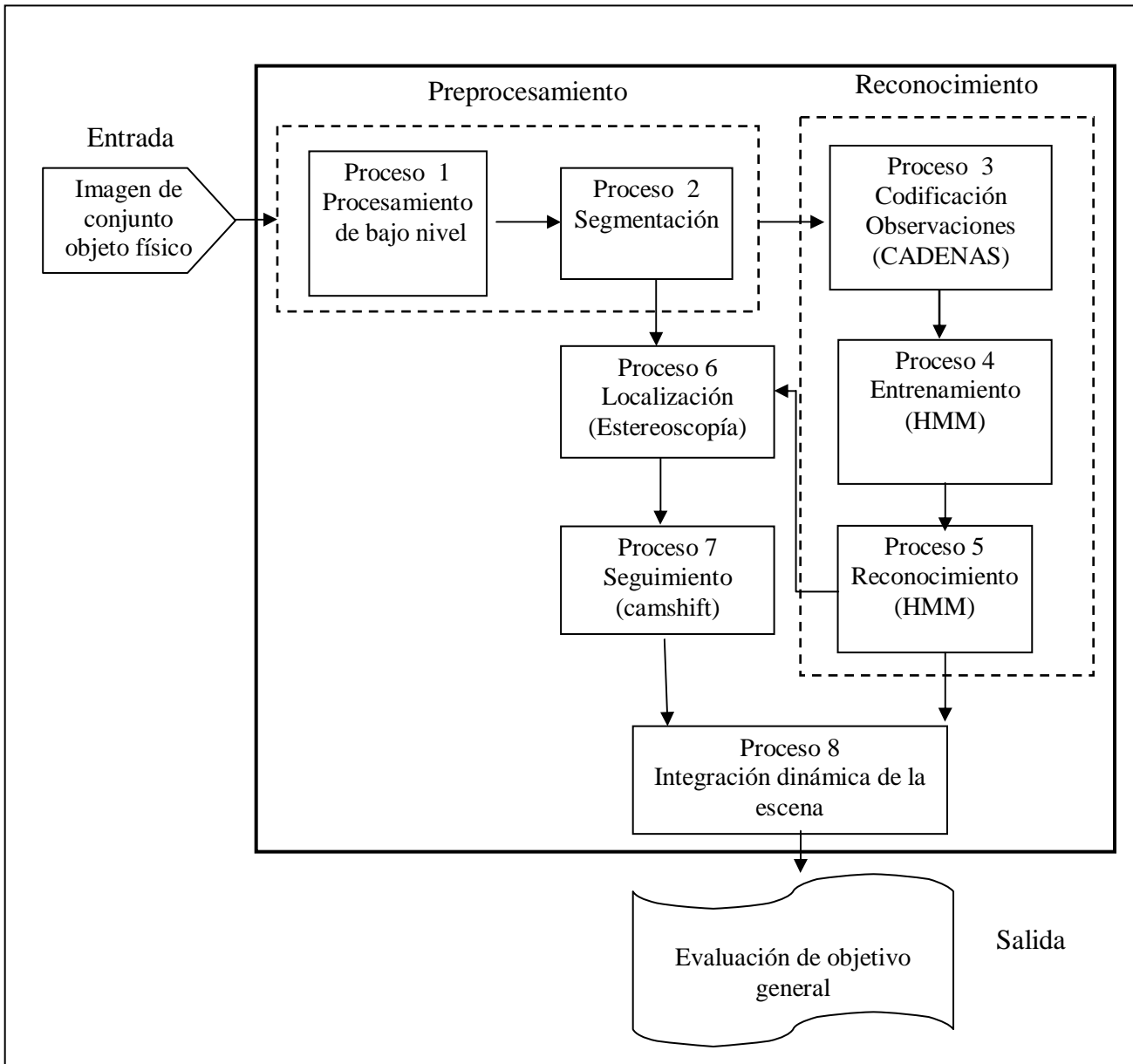


Figura. 3.1 Arquitectura del sistema SRLQ

3.2 Preprocesamiento

Los procesos uno y dos que constituyen la etapa de preprocesamiento o visión de bajo nivel, en los cuales se limpia la imagen para eliminar aspectos no deseados aplicando un filtro a la imagen o secuencia de imágenes recibidas en la entrada y se segmenta el instrumento quirúrgico usando los bordes del objeto, respectivamente.

El proceso uno, que corresponde en la etapa de preprocesamiento al procesamiento previo, se filtra una imagen (f), se aplica una transformación (T) para obtener una nueva imagen (g) de forma que ciertas características son acentuadas o disminuidas. (ver ecuación 3.1).

$$g(x,y)=T[f(x,y)] \quad (3.1)$$

Se puede ver el proceso de filtrado de la imagen pasar a través de una caja o sistema (filtro) cuya salida es la imagen filtrada (Figura 3.2).

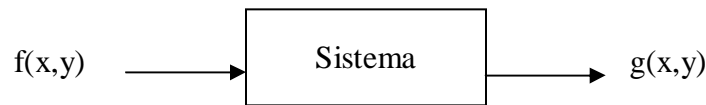


Figura 3.2 Proceso de filtrado.

Un ejemplo de filtro aplicado a una imagen se muestra en la figura 3.3, en la que se tiene un instrumento quirúrgico laparoscópico y es aplicado un filtro Gaussiano con una máscara de 5 X 5.



Figura 3.3 Aplicando Filtro Gaussiano

El proceso dos en la etapa de preprocesamiento, es la segmentación de la imagen y es aplicada después de haber filtrado la imagen, para discriminarlo entre los objetos en la escena, este paso es llamado por tal motivo “segmentación”.

Existen una variedad de trabajos en donde se ha aplicado la segmentación por color en piel, para el reconocimiento de rostros unos ejemplos de ellos se encuentran en [Vezhnevets et al], algunos otros relacionados con la segmentación de objetos usando de igual manera el color [Gil et al, 2004].

Sin embargo en este trabajo de tesis se pretende usar la segmentación utilizando la forma de nuestro instrumento, los bordes pueden proporcionarnos dicha información además de presentar ciertas propiedades, ya que son invariantes a cambios de luminosidad y color de la fuente de luz, que son algunos de los problemas típicos en la segmentación por color.

La extracción de bordes es una técnica para resaltar las orillas de una imagen y consisten en aplicar operadores de gradiente. Los operadores de gradientes son mascarar con la propiedad de resaltar los cambios de intensidad de la imagen es decir donde la intensidad

aumenta o disminuye bruscamente. Estos cambios de intensidad (bordes) son las fronteras entre regiones homogéneas de la imagen.

La manera más común de detectar orillas o bordes es utilizar algún tipo de derivada o diferencial. Las derivadas nos permiten calcular las variaciones entre un punto y su vecindario. Viendo la imagen como una función, un contorno implica una discontinuidad (figura 3.4) en dicha función es decir donde la función tiene un valor de gradiente o derivada “alta”.

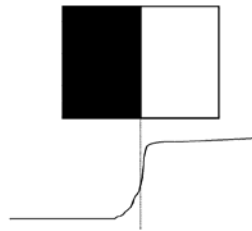


Figura 3.4. Ejemplo de discontinuidades. Arriba se muestra una imagen con una discontinuidad y abajo la grafica de su función.

El gradiente de una función $f(x,y)$ se define como: (ver ecuación 3.2):

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) \quad (3.2)$$

Y su magnitud la podemos calcular de la siguiente manera:

$$|\nabla f| = \sqrt{\left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2} \quad (3.3)$$

Sin embargo, en el caso discreto se puede hacer una aproximación de dicha derivada, tomando simplemente la diferencia entre sus valores contiguos (ver ecuación 3.4).

$I_{1,1}$	$I_{1,2}$	$I_{1,3}$
$I_{2,1}$	$I_{2,2}$	$I_{2,3}$
$I_{3,1}$	$I_{3,2}$	$I_{3,3}$

(3.4)

$$\frac{\partial f}{\partial x} = (I_{3,1} + I_{3,2} + I_{3,3}) - (I_{1,1} + I_{1,2} + I_{1,3})$$

$$\frac{\partial f}{\partial y} = (I_{1,3} + I_{2,3} + I_{3,3}) - (I_{1,1} + I_{2,1} + I_{3,1})$$

Estas operaciones pueden ser implementadas mediante mascarar o también llamados operadores tales como: Sobel, Prewitt y Robinson.

Existe otra técnica de extracción de bordes que hemos estudiado, es el llamado Canny [Canny, 1986]. No sólo usa convoluciones (operadores de gradiente), sino que busca el máximo gradiente a lo largo de un borde, lo que le da mayor robustez en la búsqueda del borde en la imagen. Este algoritmo de extracción de bordes es el que aplicamos en este trabajo de tesis, debido a que nos dio mejores resultados en la fase de segmentación por bordes, como muestra la figura 3.5.

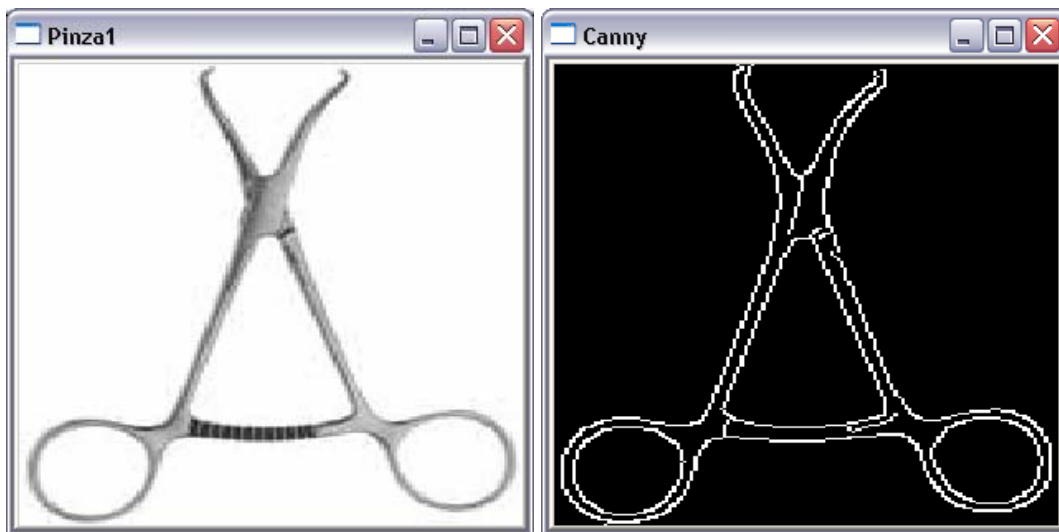


Figura. 3.5 Instrumento segmentado con Canny

3.3 Reconocimiento

Esta etapa consta de tres procesos agrupados: Codificación de observaciones, Entrenamiento y Reconocimiento, los cuales se desarrollan en los procesos tres, cuatro y cinco, respectivamente.

En el proceso de codificación de observaciones nos enfocamos en la extracción de los patrones que describen la imagen, con el fin de encontrar un modelo individual para cada instrumento quirúrgico. Estos datos, que describen al instrumento quirúrgico, nos sirven como observaciones y dichas observaciones son los datos de entrada en los HMM para la fase de entrenamiento. El proceso detallado de la extracción de las observaciones se verá en la sección 4.1.1.

Una vez extraída las observaciones de cada instrumento quirúrgico, el paso a seguir es el entrenamiento. El entrenamiento es realizado utilizando el algoritmo Baum-Welch [Rabiner y Juang, 1986], muy frecuentemente usado en los HMM para ajustar los parámetros del modelo. El proceso detallado de este paso, se verá en la sección 4.1.2.

Entrenado el modelo con las observaciones extraídas, el paso a seguir es reconocer el instrumento quirúrgico modelado anteriormente, esta fase es la de reconocimiento. Puesto que se utilizan los Modelos Ocultos de Markov, el algoritmo a utilizar para el reconocimiento de los instrumentos quirúrgicos es el Forward [Rabiner y Juang, 1986]. El proceso de reconocimiento de los instrumentos quirúrgicos es explicado con detalle en la sección 4.1.3.

3.4 Localización

En el proceso seis, una vez identificado el instrumento quirúrgico, de entre un conjunto finito de instrumentos, podemos recuperar sus coordenadas cartesianas 3D. Este paso, se refiere a la localización 3D del instrumento, específicamente solo nos enfocaremos en el

centro de masa de dicho instrumento. Para la realización del cálculo de las coordenadas 3D aprovecharemos las características de la cámara estéreo y utilizaremos la estereoscopia, debido a que la cámara tiene especial relación con el los ojos humanos. Este paso es explicado detalladamente en la sección 4.2.

3.5 Seguimiento

El proceso siete, es la fase del seguimiento de nuestro instrumento, consideramos que el instrumento tiene movimiento uniforme y el seguimiento solo se hará en el rango 2D a partir de una secuencia en el tiempo de la imagen. Por seguimiento se entiende como la habilidad para identificar el punto dado en una imagen como el mismo de la imagen previa tal como muestra la Figura. 3.7.

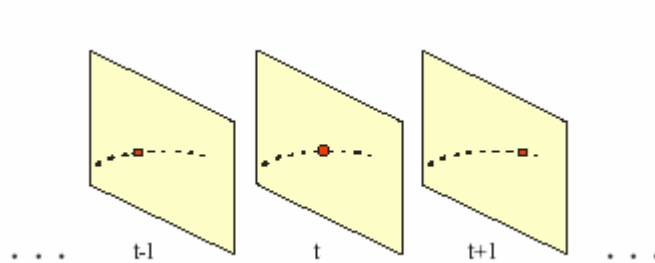


Figura. 3.6 Movimiento de un punto en el plano de la imagen

La técnica que usaremos, para dicho seguimiento, es el algoritmo camshift. Este algoritmo se explica en la sección 4.3.

3.6 Integración dinámica de la escena

En el proceso ocho, se refiere a la integración de todas las tareas descritas anteriormente, es decir, adquirida la imagen, se limpia la escena, se segmenta el instrumento quirúrgico para posteriormente extraer las observaciones necesarias para el entrenamiento. Una vez realizado dicho entrenamiento se reconoce un instrumento cualquiera. A dicho instrumento quirúrgico se le localiza o calcula las coordenadas 3D, para finalmente seguirlo en una secuencia de imágenes.

3.9 Resumen

En este capítulo se describió brevemente la arquitectura del sistema SRLQ para el desarrollo de esta tesis, algunos de los procesos se dejaron pendientes debido a la importancia en esta tesis de esas secciones. Se le dedica un capítulo al entrenamiento y reconocimiento, debido a su importancia en la tesis, al cálculo de las coordenadas 3D de un punto del instrumento quirúrgico utilizando la estereoscopía y al seguimiento siendo otra de las secciones que tiene especial atención en esta tesis.

Capítulo 4

RECONOCIMIENTO, LOCALIZACIÓN Y SEGUIMIENTO

Una vez expuesta la arquitectura del sistema SRLQ, este capítulo presenta como sus elementos realizan sus tres etapas principales: el reconocimiento, la localización y el seguimiento. Además en este capítulo, abordaremos con más detalle la extracción de características, las cuales como mencionamos anteriormente son las observaciones que nos permiten realizar el entrenamiento. Se describe el reconocimiento con los HMM y el entrenamiento. Se calculan las coordenadas 3D de los instrumentos quirúrgicos usando estereoscopía una vez reconocido el instrumento quirúrgico. Finalmente, se aborda la técnica para seguir dicho instrumento quirúrgico en la escena.

4.1 Descripción de contornos con códigos de cadena

El sistema SRLQ, tiene como estrategia de discriminación del objeto de interés en la imagen, la utilización de bordes para la segmentación de los instrumentos quirúrgicos. Una vez segmentada la imagen es importante representar la forma de los instrumentos quirúrgicos esto se logra con la descripción de contornos la cual es el paso previo a la clasificación.

Una de las técnicas que sirven para extraer el contorno de la imagen, son los llamados códigos de cadena [Freeman y Saghri,1978]. Los cuales se utilizan para representar un conjunto de puntos, que constituyen el contorno del instrumento quirúrgico. La forma de realizar dicha representación con estos códigos se explica a continuación:

En los códigos de cadenas la localización de un píxel (i,j) se señala con las coordenadas del píxel y la orientación de sus vecinos. Existen en realidad varios códigos de cadena, los mas utilizados son los de 4 y 8 direcciones. Los códigos de cadena de 4 direcciones solo consideran 4 vecinos en las direcciones oeste (O), sur (S), este (E) y norte (N); estas direcciones se les asocia los números 2, 3, 0, y 1, respectivamente, como se muestra en la figura 4.1.

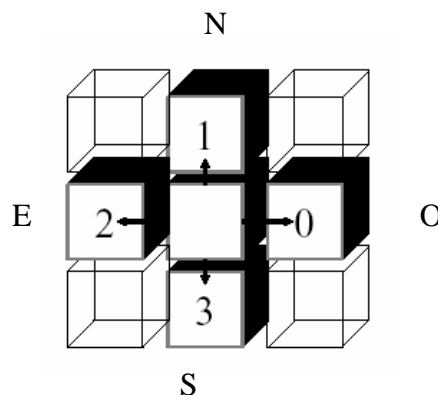


Figura 4.1 Código de 4 direcciones

En los códigos de cadena de 8 direcciones se consideran 8 vecinos oeste (O) , noreste (NO), norte (N), noreste (NE), este (E), sureste (SE), sur (S) y suroeste (SO); a estas direcciones les corresponden los números 0, 1, 2, 3, 4, 5, 6, y 7, respectivamente, como se muestra en la Figura 4.2.

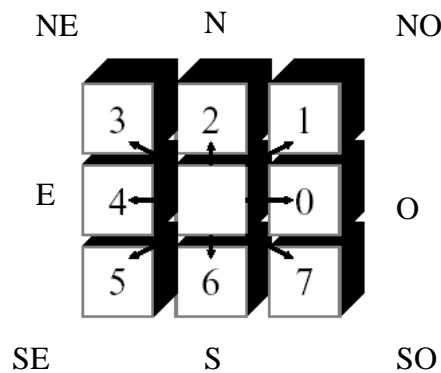


Figura 4.2 Código de 8 direcciones

Si se tiene el contorno de un objeto, en una imagen binaria, representado por una cadena de píxeles con valor 1 en la imagen binaria. El código de cadena es un tipo de estructura de dato que representa el contorno de un objeto en una imagen binaria mediante una secuencia de segmentos, conectados consecutivamente de longitud y orientación específica, que conectan píxeles adyacentes.

El código de cadena genera una secuencia de números que representan las orientaciones de los segmentos conectados consecutivamente, partiendo de un punto del contorno y siguiendo el sentido de la aguja del reloj. Un ejemplo es la cadena 05577443221 que en el código de cadena de 8 direcciones representa el contorno de la Figura 4.3.

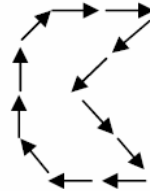
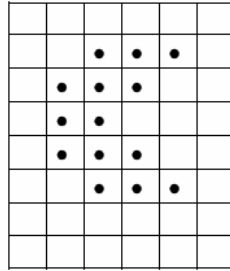


Figura 4.3 Ejemplo de código de cadena de 8 direcciones

Las ventajas que presentan los códigos de cadenas son:

- Los códigos de cadena son invariantes a traslaciones, rotaciones y escalado.
- El código de cadena es una representación compacta de un objeto binario. Suministra una buena comprensión de la descripción de contornos.

Como ejemplo tomemos la imagen binaria de la figura 4.4 (parecida a un triángulo) cuyo código de cadena de 8 direcciones de dicha imagen es extraída en la tabla 4.1.

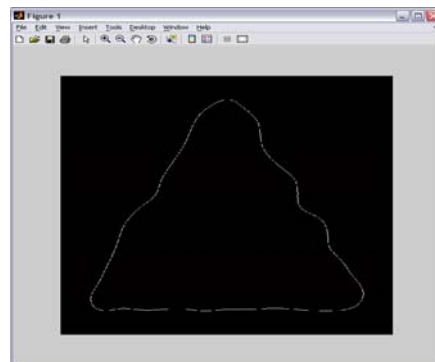


Figura 4.4 Imagen binaria

Secuencia de código de cadena de 8 direcciones															
0	0	6	0	6	6	0	6	0	6	6	6	0	6	6	0
6	6	0	6	6	4	4	4	4	4	4	4	4	4	4	4
4	4	2	4	1	2	0	2	2	2	0	2	0	2	2	0
2	2	2	0	2											

Tabla 4.1 Código de cadena 8 direcciones de la Figura 4.4:

4.2 Definición y arquitectura de los HMM

Un HMM, es un proceso doblemente estocástico, que consta de un proceso de Markov no observable (oculto), lo cual son las secuencias de estados no observables y un proceso observado, los cuales son la secuencia de observaciones.

Un HMM es un modelo matemático capaz de predecir probabilísticamente el comportamiento de un proceso de Markov. Un proceso estocástico se puede llamar de Markov solo si se cumple que en todo instante de tiempo t , el estado en el proceso en el instante t , depende únicamente del estado del sistema en el instante $t-1$ y no de la historia del proceso.

La base teórica de los HMM fue desarrollada por [Baum et al, 1970] [Baum2,1970], a finales de 1960, sin embargo en los últimos años ha sido extensamente aplicada en un gran número de problemas. Algunas de las aplicaciones en que se utilizan los HMM son en el reconocimiento de voz [Rabiner, 1989], reconocimiento de caracteres escritos a mano [Hu et al, 1996], modelado de proteínas y DNA[Hughey y Krogh, 1996], y reconocimiento de gestos[Donal y Tanguay, 1993].

En esta sección describiremos formalmente los HMM, puesto que los utilizamos para el reconocimiento y aprendizaje de los instrumentos quirúrgicos. Formalmente los HMM están definidos por los siguientes elementos:

Las variables N y M son el tamaño máximo de estados y el tamaño máximo de observaciones en los HMM, respectivamente.

Un conjunto $Q = \{q_1, q_2, \dots, q_N\}$ de estados ocultos.

Una distribución de probabilidad de los estados llamada matriz $A = \{a_{ij}\}$, que definen las transición del estado Q_i al estado Q_j .

$$a_{ij} = P[q_{t+1} = Q_j \mid q_t = Q_i] \quad 1 \leq i, j \leq N \quad (4.1)$$

Donde $a_{ij} \geq 0$ y $\sum_{j=1}^N a_{ij} = 1$

Un conjunto $V = \{v_1, v_2, \dots, v_M\}$ de símbolos de observación.

Una distribución de probabilidad de los símbolos de observación $B = \{b_j(k)\}$, donde en cada tiempo t la matriz de observación nos dice la probabilidad del símbolo v_k estando en el estado Q_j .

$$b_j(k) = P[v_k \text{ en tiempo } t \mid q_t = Q_j] \quad 1 \leq j \leq N, 1 \leq k \leq M \quad (4.2)$$

Donde $b_j(k) \geq 0$ y $\sum_{k=1}^M b_j(k) = 1$

Una distribución de probabilidad de estados iniciales $\pi = \{\pi_i\}$

$$\pi_i = P[q_1 = S_i] \quad 1 \leq i \leq N \quad (4.3)$$

Donde $\pi_i \geq 0$ y $\sum_{i=1}^N \pi_i = 1$

En los HMM existen tres problemas a resolver, los cuales se describen a continuación:

Problema 1: Dada una secuencia de observaciones $O = O_1O_2\dots O_T$ y un modelo λ , ¿Cómo calcular eficientemente la $P(O|\lambda)$, la probabilidad de la observación dado nuestro modelo?

Problema 2: Dada una secuencia de observaciones $O = O_1O_2\dots O_T$ y un modelo λ , ¿Cómo elegir su correspondiente secuencia de estados $Q = q_1q_2\dots q_T$?

Problema 3: ¿Cómo ajustar el modelo de parámetros $\lambda = (A, B, \pi)$ que maximice $P(O|\lambda)$?

Dentro de los HMM existen diferentes arquitecturas que los representan gráficamente. Un HMM puede ser representado como un grafo dirigido. La arquitectura específica que permita modelar de la mejor forma posible las propiedades observadas depende en gran medida de las características del problema. Las arquitecturas mas usadas son:

1. **Ergódicas.** En este tipo de arquitectura cada estado del modelo puede ser alcanzado desde cualquier otro estado, como muestra la figura 4.5.

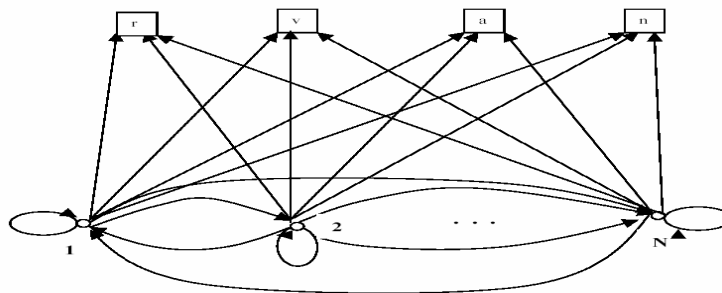


Figura 4.5 Arquitectura ergódica de los HMM

2. **Izquierda-Derecha.** También llamada Bakis, 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. Como muestra la Figura 4.6.

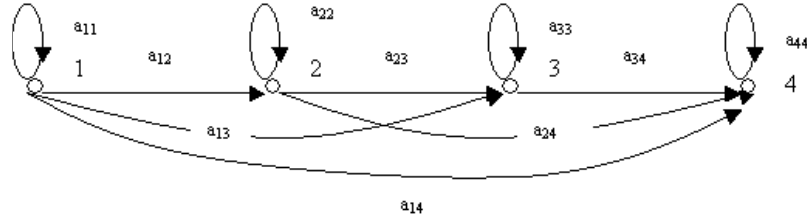


Figura 4.6 Arquitectura Izquierda – Derecha de los HMM

3. Izquierda – Derecha paralelas. Son similares a la arquitectura izquierda - derecha, la diferencia es que esta arquitectura tiene dos arquitecturas izquierda - derecha en paralelo. Como muestra la Figura 4.7.

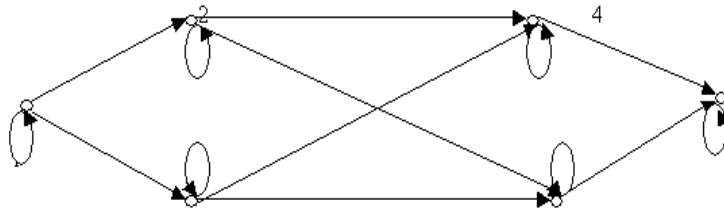


Figura 4.7. Arquitectura Izquierda – Derecha paralelas de los HMM

Una vez definidos los HMM y expuestas sus distintas arquitecturas, cabe resaltar que en este trabajo se resuelven solo dos problemas de los anteriormente expuestos. Inicialmente resolveremos el problema 1, el cual nos permite encontrar la probabilidad de la observación dado el modelo $P(O|\lambda)$, previamente entrenado. Posteriormente resolveremos el problema 3, el cual nos permitirá ajustar los parámetros de un modelo propuestos $\lambda = (A, B, \pi)$ con un conjunto de observaciones dadas $O = O_1O_2...O_T$, este problema nos sirve para entrenar nuestro modelo. Estos dos problemas son descritos en las siguientes dos subsecciones.

4.2.1 Resolución al problema de reconocimiento

El problema de reconocimiento o evaluación, consiste en calcular eficientemente la probabilidad de que ocurra una determinada secuencia de observación con un modelo $P(O | \lambda)$. El cálculo puede ser realizado con la ecuación 4.4, sin embargo esto implicaría un costo computacional muy caro. Si se cuenta con N estados y T observaciones la complejidad es $2TN^T$ cálculos. Afortunadamente, se puede reducir este número de cálculos con el algoritmo Forward [Rabiner, 1989].

$$\begin{aligned}
 P(O | \lambda) &= \sum_{\text{todo } Q} P(O, Q | \lambda) \\
 P(O | \lambda) &= \sum_{\text{todo } Q} P(O | Q, \lambda) P(Q | \lambda) \\
 &= \sum_{q_1, q_2, \dots, q_T} \pi_{q_1} b_{q_1}(O_1) a_{q_1 q_2} b_{q_2}(O_2) \dots a_{q_{T-1} q_T} b_{q_T}(O_T)
 \end{aligned} \tag{4.4}$$

Considérese la variable $\alpha_t(i)$ definida por la ecuación 4.5 que representa la probabilidad condicional de la secuencia de observación, $O_1 O_2 \dots O_t$ y el estado S_i , en tiempo t dado un modelo λ .

$$\alpha_t(i) = P(O_1 O_2 \dots O_t, q_t = S_i | \lambda) \tag{4.5}$$

El algoritmo Forward está compuesto por tres pasos: 1) Inicialización, 2) Inducción y 3) Terminación. Estos pasos conjuntamente resuelven la variable $\alpha_t(i)$, descrita anteriormente, utilizando inducción. El proceso de solución se describe a continuación.

1. Inicialización:

$$\alpha_1(i) = \pi_i b_i(O_1) \quad 1 \leq i \leq N \tag{4.6}$$

2. Inducción:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(O_{t+1}) \quad 1 \leq t \leq T-1, \quad 1 \leq j \leq N \quad (4.7)$$

3. Terminación

$$P(O | \lambda) = \sum_{i=1}^N \alpha_T(i) \quad (4.8)$$

En el paso 1, denominado inicialización, se establece un valor inicial de $\alpha_t(i)$ como la probabilidad de que ocurra π_i en un tiempo t y se observe O_1 . El paso 2 de inducción se ilustra en la Figura 4.8 que muestra como el estado S_j puede ser alcanzado en el tiempo $t + 1$ de entre N posibles estados S_i .

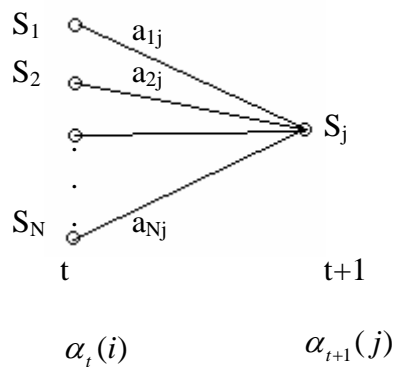


Figura 4.8 Ilustración de operaciones del algoritmos Forward

Finalmente, en el paso 3, se termina el algoritmo calculando $P(O | \lambda)$ como la suma de la variable $\alpha_T(i)$ final, de esta manera se resuelve el problema 1, el cual nos permite reconocer el instrumento deseado de entre un conjunto de instrumentos, teniendo solamente las observaciones.

De este modo los cálculos de operaciones se reducen a $N^2 T$, lo cual resulta factible utilizar este algoritmo.

4.2.2 Entrenamiento de los HMM

El entrenamiento o aprendizaje de los HMM, es una de las partes importantes cuando se desea hacer reconocimiento de patrones. En los HMM la fase de entrenamiento consiste en alimentar de observaciones obtenidas con los códigos de cadenas y ajustar los parámetros del modelo $\lambda = (A, B, \pi)$.

El problema mas difícil de los HMMs es determinar un método para ajustar los parámetros (A, B, π) del modelo para satisfacer los criterios de optimización. No se conoce una forma analítica para fijar los parámetros que maximice la probabilidad de la secuencia de observación. En realidad no existe un modo analítico para resolver el problema completo de entrenamiento del modelo. A lo más que se puede aspirar es encontrar un máximo local que proporcione unos parámetros razonables para el modelo. Para ellos se puede utilizar el algoritmo forward-backward, de Baum-Welch [Rabiner, 1989], como se describe a continuación.

Para desarrollar el algoritmo, es necesario definir la variable de retroceso $\beta_t(i)$ como:

$$\beta_t(j) = P(O_{t+1}O_{t+2}\dots O_T | q_t = S_j, \lambda) \quad (4.9)$$

La ecuación anteriormente descrita, nos dice cual es la probabilidad condicional de una secuencia de observaciones en tiempo $t+1$ hasta el tiempo T , dado el estado S_i y el modelo λ . Nuevamente podemos definir la variable de retroceso por inducción, a partir del instante T .

1. Inicialización:

$$\beta_T(i) = 1 \quad 1 \leq i \leq M \quad (4.10)$$

2. Inducción:

$$B_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j) \quad t = T-1, T-2, \dots, 1 \quad 1 \leq i \leq N \quad (4.11)$$

En el paso 1, arbitrariamente el valor $\beta_T(i)$ se define arbitrariamente como la unidad, el paso 2 se ilustra en la Figura 4.9.

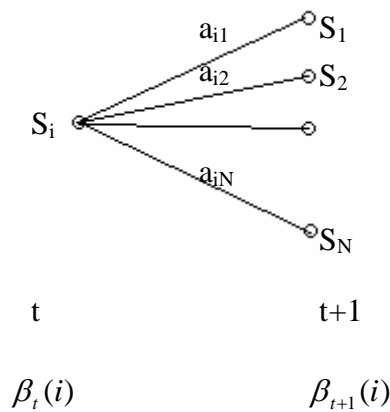


Figura 4.9 Ilustración de operaciones por el algoritmo Backward

El procedimiento de entrenamiento para reestimar los parámetros A, B y π , utiliza la variable $\xi_t(i, j)$, que es la probabilidad de encontrarse en el estado i en el tiempo t y en el estado j en el tiempo $t+1$, dado el modelo λ y la secuencia de observación O , esto es:

$$\xi_t(i, j) = \frac{P(q_t = S_i, q_{t+1} = S_j, O | \lambda)}{P(O | \lambda)} \quad (4.12)$$

La variable $\xi_t(i, j)$ también puede reescribirse en términos de $\alpha_t(j)$ y $B_t(i)$ de la siguiente manera:

$$\xi_t(i, j) = \frac{\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)} \quad (4.13)$$

Asimismo, se puede definir la variable $\gamma_t(i)$ como la probabilidad de que el proceso pase por el estado Q_i en el instante t , dados la secuencia de observaciones y el modelo. Esta variable simplemente es la suma de las variables $\xi_t(i, j)$ para cualquier estado oculto en $t+1$, es decir:

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j) \quad (4.14)$$

En donde las dos variables nos dan:

$$\sum_{t=1}^T \gamma_t(i) = \text{número de transiciones esperadas desde el estado } S_i. \quad (4.15)$$

$$\sum_{t=1}^T \xi_t(i, j) = \text{número de transiciones esperadas del estado } S_i \text{ a } S_j. \quad (4.16)$$

Usando estas dos variables se pueden calcular los parámetros del modelo:

$$\bar{\pi} = \text{frecuencia esperada de ocurrencia del estado } S_i \text{ en } t+1 = \gamma_1(i) \quad (4.17)$$

$$\begin{aligned} a_{ij} &= \frac{\text{numero de transiciones esperadas del estado } S_i \text{ al estado } S_j}{\text{frecuencia esperada de ocurrencia del estado } S_i} \\ &= \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \end{aligned} \quad (4.18)$$

$$\bar{b}_j(k) = \frac{\text{frecuencia de ocurrencia del estado } S_j \text{ y observado símbolo } v_k}{\text{frecuencia esperada del estado } S_j}$$

$$= \frac{\sum_{t=1}^T \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)} \quad (4.19)$$

Con las ecuaciones 4.17-19, si tenemos un modelo para la secuencia de observaciones de nuestro objeto a reconocer, la evaluación de un modelo propuesto inicial $\bar{\lambda} = (\bar{A}, \bar{B}, \bar{\pi})$, debe dar como resultado los parámetros del modelo original $\lambda = (A, B, \pi)$ iterando los cálculos hasta que la variación de los parámetros entre dos pasos iterativos sea suficientemente pequeña.

Es importante notar que, dado el procedimiento de estimación de los parámetros del modelo, no se asegura la obtención del máximo global, es conveniente seguir un proceso de inicialización que sitúe los parámetros dentro de un rango razonable.

4.3 Visión Estereoscópica

Como ya se mencionó en la sección 2.3, del capítulo 2, la visión estereoscópica toma como referencia el modelo estereoscópico biológico. En estos sistemas el desplazamiento de los ojos permite obtener la profundidad de los objetos mediante un simple proceso de triangulación a partir de dos imágenes generadas por el mismo objeto en la escena 3D en cada ojo. Esto es posible por que el hecho de que los ojos estén desplazados entre sí, hace que las imágenes de los objetos se muestren desplazadas según la distancia de los objetos a los ojos.

La visión estereoscópica artificial hace uso de dos o más cámaras separadas entre sí a cierta distancia con las que se puede obtener las imágenes correspondientes a cada cámara. El procedimiento consiste en captar dos o más imágenes de una misma escena

con una posición ligeramente diferente a las anteriores, pues esto permite la obtención de la distancia a la que se encuentra determinado objeto.

4.3.1 Geometría del sistema estereoscópico

Para obtener las coordenadas cartesianas 3D del centro del instrumento quirúrgico utilizando la estereoscopía es necesario conocer la geometría del un sistema estereoscópico.

Un sistema convencional está caracterizado por un par de cámaras con sus ejes ópticos paralelos y separados por una distancia horizontal la cual se denomina distancia base (ver figura 4.10) el cual es el parámetro b .

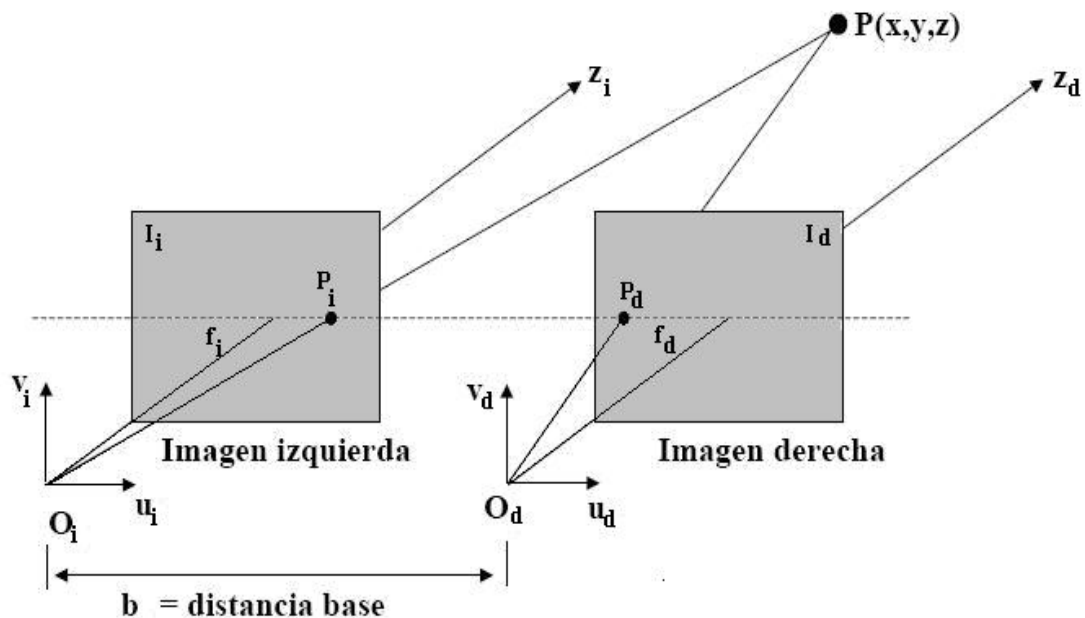


Figura 4.10. Geometría de un par de cámaras en estéreo con ejes ópticos paralelos

En este sistema de ejes ópticos paralelos el desplazamiento entre cada centro óptico de las dos cámaras es horizontal, esto quiere decir que un punto determinado de la escena captado por las dos cámaras difiere solamente en la componente horizontal. Por ejemplo

la figura 4.10 muestra una escena captada por dos cámaras que proporcionan una imagen izquierda y una imagen derecha en las que los puntos correspondiente de la escena P_i y P_d para las imágenes izquierda y derecha, respectivamente, puede apreciarse que el punto solo está desplazado sobre el eje u de las imágenes tomadas y el eje v es el mismo en las dos imágenes.

En la estereoscopia existen varios problemas a resolver, como el problema de la correspondencia estereoscópica, la cual se refiere al proceso mediante el cual dado un punto en la escena 3D, determinar que pareja de puntos de cada imagen se corresponden con un mismo punto en la escena. Es decir, resolver la correspondencia supone que dada una característica en la imagen (borde o punto) la proyección de una primitiva en un espacio 3D, obtener cual es la característica proyectada de la misma primitiva, en otra imagen desde un punto de vista diferente. Geométricamente, pueden existir infinitas soluciones o quizá no exista ninguna, esto, por oclusiones, lo que puede dar lugar a una falsa correspondencia.

En esta tesis no es necesario resolver el problema de correspondencia, puesto que se calculan los centros geométricos de los instrumentos quirúrgicos previamente, sin embargo, el problema de la reconstrucción, otro de los problemas de la visión estereoscópica si es resuelto pues su solución nos permite obtener las coordenadas cartesianas 3D del centro geométrico de los instrumentos quirúrgicos.

4.3.2 Reconstrucción 3D

El procedimiento de reconstrucción 3D consiste en calcular la ubicación (x,y,z) de cada punto $P(x,y,z)$ correspondiente en el par de imágenes aplicando la técnica de visión estereoscópica.

La figura 4.11 muestra la geometría de un par de cámaras en estéreo, representadas por sus modelos puntuales con sus planos imagen, I_l e I_r , reflejados sobre sus centros O_l y O_r , respectivamente.

El origen del sistema de coordenadas de la figura 4.11 está en O , la longitud focal efectiva de cada cámara es f y b es la línea base, explicada previamente. Para un punto dado P_l en la imagen izquierda, su punto correspondiente P_r en la imagen derecha. Estos puntos deben caer sobre la línea epipolar. La línea epipolar en la imagen derecha correspondiente a un punto P_l en la imagen izquierda define el espacio de búsqueda dentro del cual se situará el correspondiente punto de emparejamiento P_r . Las coordenadas en los puntos $P_l(x,y)$ y $P_r(x,y)$ puede obtenerse un valor de disparidad dado por $d = x_l - x_r$.

Considerando la relación geométrica de triángulos que existe en el sistema estereoscópico, las coordenadas del punto $P(x,y,z)$ pueden deducirse fácilmente observando la figura 4.11. Estas coordenadas se deducen en la ecuación 4.20, utilizando este tipo de geometría de cámaras, la profundidad Z es inversamente proporcional a la disparidad d de la imagen.

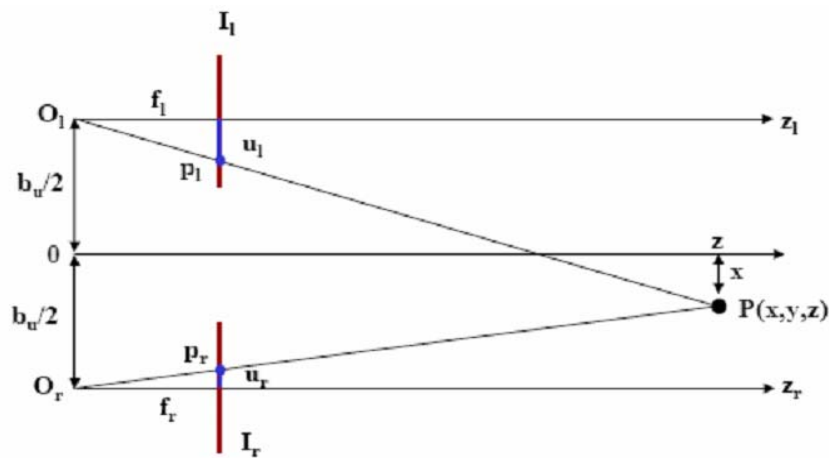


Figura 4.11. Geometría de dos cámaras en estéreo con ejes ópticos paralelos.

$$\left. \begin{array}{l} o_l : \frac{\frac{b}{2} + x}{z} = \frac{u_l}{f} \\ o_r : \frac{\frac{b}{2} - x}{z} = \frac{u_r}{f} \end{array} \right\} \Rightarrow \left. \begin{array}{l} u_l = \frac{f}{z} \left(x + \frac{b}{2} \right) \\ u_r = \frac{f}{z} \left(x - \frac{b}{2} \right) \end{array} \right\} \Rightarrow d = u_l - u_r = \frac{fb}{z} \Rightarrow z = \frac{fb}{d} \quad (4.20)$$

4.4 Seguimiento visual

En esta sección abordaremos una de las secciones previamente mencionadas en el Capítulo 3, la cual es el seguimiento. La técnica a desarrollar para el seguimiento de los instrumentos es el algoritmo camshift [Bradsky, 1998]. El algoritmo camshift se basa en una técnica robusta no paramétrica (el algoritmo de desplazamiento de la media -Mean Shift-) que ascendiendo en gradientes de densidad permite encontrar la moda de distribuciones de probabilidad. De esta forma podemos encontrar la zona con mayor densidad de puntos pertenecientes a los datos, que son los píxeles que pertenecen a los instrumentos quirúrgicos.

Camshift se basa en ideas de estadística robusta y distribuciones de probabilidad. Los estadísticos robustos son aquellos que tienden a ignorar los 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).

4.4.1 Algoritmo Mean Shift

El algoritmo en que se basa Camshift es Mean Shift el cual no fue diseñado para ser usado como algoritmo de seguimiento, sino como algoritmo de clasificación de datos, pero funciona bastante bien en ese papel.

Mean Shift es un algoritmo orientado a la clasificación de vectores de características. Es un procedimiento simple que desplaza cada punto de los datos hacia la media de los datos que se encuentran en su vecindad. Ahora bien, para establecer los datos que se encuentren en la vecindad de otro, es necesaria una medida cuantitativa de la vecindad de dos datos.

Básicamente el algoritmo Mean shift, permite encontrar agrupamientos de datos en la muestra. De esta forma, si nosotros tenemos una imagen con datos clasificados como el instrumento quirúrgico, éstos se encontrarán agrupados en una parte de la misma y utilizando este algoritmo podremos encontrar este agrupamiento.

4.4.2 Algoritmo Camshift

A diferencia del algoritmo de desplazamiento de la media, que se utiliza en distribuciones estáticas, Camshift se diseñó para distribuciones que cambian dinámicamente. Esto ocurre cuando los objetos se están siguiendo en una secuencia de vídeo, se mueven de tal forma que el tamaño y la posición de la distribución de probabilidad varía con el tiempo.

El algoritmo Mean shift, por sí solo fallaría como seguidor. Un tamaño de ventana que funciona a una escala de la distribución no es bueno para otra escala cuando el objeto de color se aleja y acerca de la cámara. Ventanas de tamaño fijo pequeño podrían perder el blanco por completo en grandes movimientos del objeto en la escena, mientras que ventanas de tamaño fijo grandes podrían incluir distracciones y demasiado ruido.

El algoritmo camshift ajusta el tamaño de la ventana de búsqueda en el transcurso de sus operaciones. El tamaño inicial de la ventana de búsqueda se puede establecer en cualquier valor razonable. Para distribuciones discretas (datos digitales), el tamaño mínimo de la ventana es 3. En lugar de un conjunto de tamaños o una adaptación externa del tamaño de la ventana, camshift confía en la información del momento de orden 0 (4.21) para adaptar continuamente el tamaño de la ventana en cada cuadro de vídeo.

El momento de orden cero representa el "área" de la distribución bajo la ventana de búsqueda. El tamaño de la ventana se establece como una función del momento de orden cero. El algoritmo camshift se calcula entonces utilizando cualquier tamaño de ventana que no sea nulo (mayor o igual que 3 si la distribución es discreta).

$$M_{00} = \sum_x \sum_y I(x, y) \quad (4.21)$$

El algoritmo camshift realiza los siguientes pasos:

1. Elegir el tamaño de la ventana de búsqueda.
2. Elegir la posición inicial de la ventana de búsqueda.
3. Desplazar la media (una o varias iteraciones) y actualizar el momento de orden 0.
4. Establecer el tamaño de la ventana de búsqueda igual a una función del momento de orden 0.
5. Repetir los pasos 3 a 4 hasta la convergencia.

Cuando sigue un objeto camshift trabaja con una imagen de la distribución de probabilidad del color. camshift calcula el centroide de la distribución de probabilidad del color dentro de la ventana de búsqueda, vuelve a centrar la ventana, y calcula el área para obtener el siguiente tamaño de ventana. De esta forma no necesitamos calcular la distribución de probabilidad del color en la imagen completa, sino que basta con restringir el cálculo de la distribución a una pequeña región de la imagen que envuelve a la ventana de búsqueda actual de camshift. Esto evitará gran cantidad de cálculos cuando la imagen del instrumento segmentado no domina toda la imagen.

4.4 Resumen

En este capítulo se abordan tres de las fases de nuestra arquitectura previamente descrita. Inicialmente. La fase de reconocimiento, que cuenta la extracción de características utilizando los códigos de cadena, los cuales nos permiten extraer las observaciones de nuestro instrumento quirúrgico. Una de las ventajas de utilizar los códigos de cadenas es que los datos extraídos están discretizados, es decir, no son datos continuos, lo cual nos permite usar de forma natural como observaciones. El reconocimiento con los HMM, dando su definición formal de los HMM, así como las arquitecturas habitualmente utilizadas. El entrenamiento de los HMM, el cual permite encontrar los parámetros del modelo.

Otra de las fases de nuestra arquitectura, previamente explicada, es la localización en la que se calculan las coordenadas cartesianas 3D de un solo punto del instrumento quirúrgico, en esta sección, se describió la técnica de calculo de profundidad basada en estereoscopia. Finalmente, se abordo la técnica para seguimiento, llamada Camshift, en la que se explica el funcionamiento de este algoritmo.

En el siguiente capítulo, se exponen los resultados encontrados de las fases explicadas en este capítulo.

Capítulo 5

ANÁLISIS DE RESULTADOS

Una vez explicadas las técnicas utilizadas en esta tesis, el paso a seguir es la exposición de resultados encontrados en nuestra experimentación. En este capítulo abordamos el análisis de los resultados de las fases anteriormente descritas. Abordaremos en forma individual cada una de las técnicas, inicialmente expondremos los resultados de la fase de reconocimiento, posteriormente expondremos los resultados del cálculo de las coordenadas cartesianas 3D del centro geométrico de un instrumento quirúrgico. Finalmente, una vez reconocido y calculadas las coordenadas 3D, expondremos los resultados del seguimiento de instrumentos quirúrgicos.

5.1 Instrumental quirúrgico y cámara estéreo

Para la realización de los experimentos se escogieron cuatro tipos de instrumentos quirúrgicos (figura 5.1), con el fin de comprobar cada uno de las fases descritas previamente en el capítulo 4, dichos instrumentos presentan la característica de ser instrumentos de cirugía abierta. Además de ser distintos en cuanto a la forma que presentan. Esto nos facilita en gran manera la realización de los experimentos.

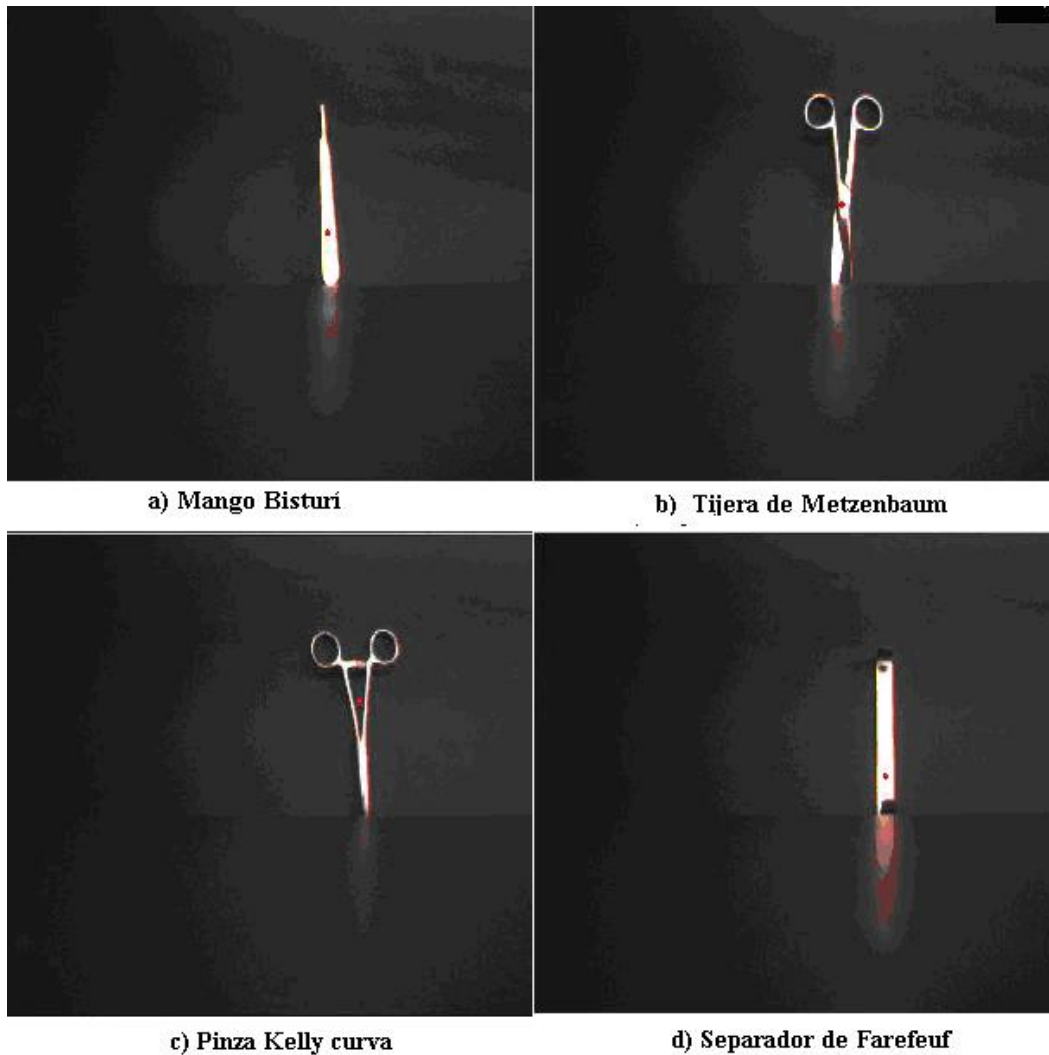


Figura 5.1 Instrumental quirúrgico.

La cámara estéreo que se utiliza en esta tesis, para medir la distancia del objeto con respecto a la cámara. Dicha cámara nos proporciona las dos imágenes necesarias para el cálculo de la distancia. El tamaño de las imágenes es de 320x240 y la velocidad de captura de ambas imágenes es de 5 frames / segundo.



Figura 5.2. Cámara estéreo BumbleBee

Los parámetros de la cámara estereos BumbleBee son:

- Longitud focal = 6mm.
- Separación entre las cámaras = 120 mm
- Dimensiones del sensor CCD de 1 cm² formado por 640 x 480 sensores ópticos con separación entre píxeles de 20 μm .

En las secciones siguientes se abordará cada una de las fases del sistema SRLQ. Se presentan los resultados inicialmente del reconocimiento de instrumentos quirúrgicos. Posteriormente se presentan resultados de localización o cálculo de la profundidad. Finalmente se muestran resultado obtenidos en seguimiento de cada uno de los instrumentos quirúrgicos.

5.2 Análisis de resultados en reconocimiento

En esta fase del reconocimiento se analizan los instrumentos quirúrgicos con los que se experimentó. Se extrajo la información del instrumento a analizar, la cual describe que describe la forma de cada objeto utilizando los códigos de cadena de 8 direcciones, obteniendo así la segmentación de la imagen como se ilustra la figura 5.3.



a) Imagen original

b) Imagen segmentada por bordes.

Figura 5.3. Instrumento codificado por su forma

Cuando se segmenta el instrumento quirúrgico por bordes, dependiendo de la forma del objeto se generan diversos bordes cerrados. Si se desean extraer los códigos de cadena de 8 direcciones, es necesario extraer de forma individual cada uno de los bordes cerrados. La tabla 5.1 de los dos bordes cerrados pequeños (agujeros de toma) de la figura 5.3b.

Secuencia de tamaño 111
100000777077767676667666665665665656565554554545
545445454444344343323232222222222222222122221212121
111111010010000
Secuencia de tamaño 113
1000000000000070707776777667667666666666566565556
55555454454444444444444434443433333232222222222221
22122112111211010

Tabla 5.1 Secuencia de cadena decodigos de 8 direcciones.

Una vez extraída las secuencias de cada uno de los instrumentos a reconocer, se inicia el proceso de entrenamiento, con los HMM. Para validar la fase de entrenamiento se prueban diferentes estados del modelo y se escoge el modelo que mejor resultados arroje. Iniciamos la fase de entrenamiento con 4 estados y haciendolos variar hasta 15 estados.

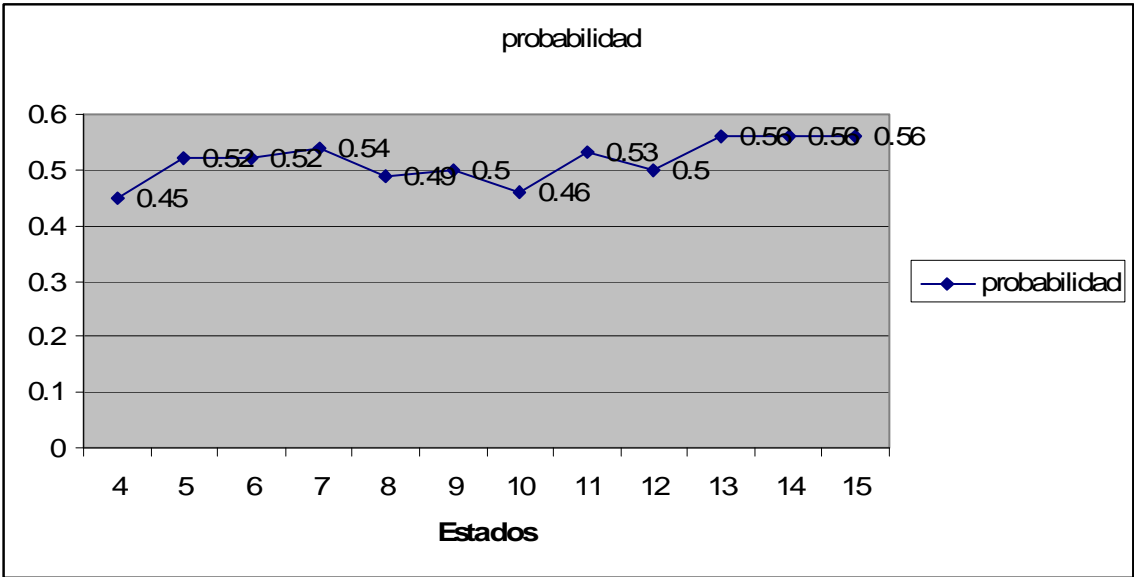


Figura 5.4. Entrenamiento del instrumento “Separador Farefeuf” variando estados.

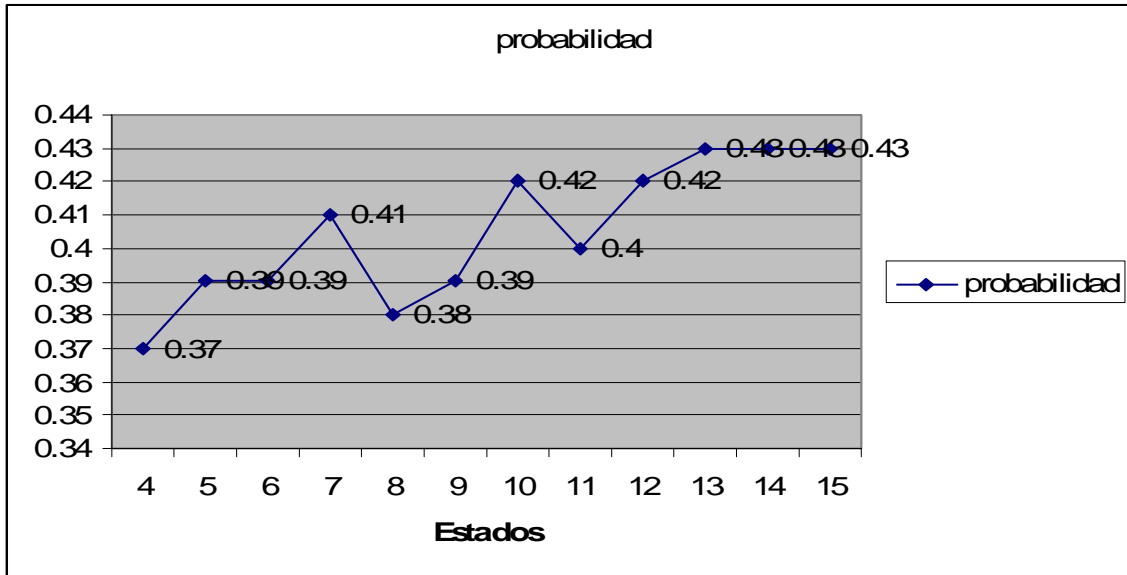


Figura 5.5. Entrenamiento de pinza Metzemaum

Como se muestra en la figura 5.4-5, conforme se agregan estados el modelo tiende al equilibrio, por tal motivo para el modelo de la “Pinza Metzemaum” se seleccionó el modelo de 13 estados, mientras que para el Separador Ferefeuf se escogio de 11 estados. La tabla 5.2 y 5.3 muestra los modelos entrenados para el Separador Forefeuf y la Pinza Metzemaum, respectivamente.

ESTADOS 11
 OBSERVACIONES 8
 TIPO DE ESTRUCTURA: BAKIS

MATRIZ A								
0.0000	0.9999	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.9999	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.9296	0.0703	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.6020	0.3979	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.9936	0.0063	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.8100	0.1899	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.9356	0.0643	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.9937	0.0062
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.7456	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MATRIZ B:								
0.0000	0.0000	0.0000	0.0000	0.0000	0.2152	0.7004	0.0082	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2134	0.0364	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.8133	0.0000
0.4771	0.0000	0.0219	0.0000	0.0000	0.0000	0.0000	0.1378	0.9999
0.0000	0.9999	0.9078	0.0000	0.0000	0.0000	0.0000	0.0041	0.0000
0.5228	0.0000	0.0702	0.5993	0.0150	0.0000	0.0321	0.0000	0.0000
0.0000	0.0000	0.0000	0.4006	0.8565	0.1148	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.1283	0.6699	0.0539	0.0000	0.0000
VECTOR Pi								
0.9999	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Tabla 5.2. Valores del modelo entrenado para el separador Farefeuf

ESTADOS =13
OBSERVACIONES 8
ESTRUCTURA: BAKIS
MATRIZ A
0.0000 0.9999 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
0.0000 0.0000 0.9999 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
0.0000 0.0000 0.9162 0.0837 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
0.0000 0.0000 0.0000 0.0017 0.9982 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
0.0000 0.0000 0.0000 0.0000 0.8360 0.1639 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
0.0000 0.0000 0.0000 0.0000 0.0000 0.9692 0.0307 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.9999 0.0000 0.0000 0.0000 0.0000 0.0000
0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.9792 0.0207 0.0000 0.0000 0.0000 0.0000
0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.7964 0.2035 0.0000 0.0000 0.0000
0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.8801 0.1198 0.0000 0.0000
0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.9809 0.0190 0.0000
0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.9071 0.0928
0.0357 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.9642
MATRIZ B
0.0732 0.3337 0.7203 0.0000 0.0000 0.0000 0.0000 0.0000 0.3774 0.0313 0.0000 0.0000 0.0000
0.5826 0.4038 0.1648 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0076 0.0462 0.0000 0.1971
0.0747 0.0668 0.0000 0.0835 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.8068 0.2014 0.7956
0.0000 0.0000 0.0000 0.0843 0.0000 0.0000 0.0000 0.0000 0.0000 0.2785 0.1451 0.4946 0.0000
0.0662 0.1955 0.0215 0.2495 0.3651 0.0000 0.0000 0.0057 0.0000 0.5435 0.0000 0.3038 0.0036
0.1326 0.0000 0.0000 0.0000 0.4856 0.0000 0.7753 0.1260 0.1937 0.1263 0.0017 0.0000 0.0000
0.0000 0.0000 0.0000 0.0000 0.0454 0.7716 0.0000 0.8250 0.0445 0.0000 0.0000 0.0000 0.0035
0.0702 0.0000 0.0932 0.5825 0.1037 0.2283 0.2246 0.0431 0.3843 0.0124 0.0000 0.0000 0.0000
VECTOR PI
0.9999 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000

Tabla 5.3 Valores del modelo entrenado para la pinza Metzemaum

Una vez realizado el entrenamiento con cada uno de los instrumentos, el paso siguiente es la evaluación de los modelos entrenados y verificar si dado el modelo entrenado es posible reconocer cada uno de los instrumentos. Para lo anterior se coloca en la escena el grupo de objetos previamente entrenados, en este caso son dos instrumentos, y se extraen los códigos de cadena de cada uno de los instrumentos. Posteriormente se prueban las observaciones extraídas, las cuales son los códigos de cadena, y se verifica la probabilidad que calcula dado el modelo entrenado.



a) Imagen izquierda

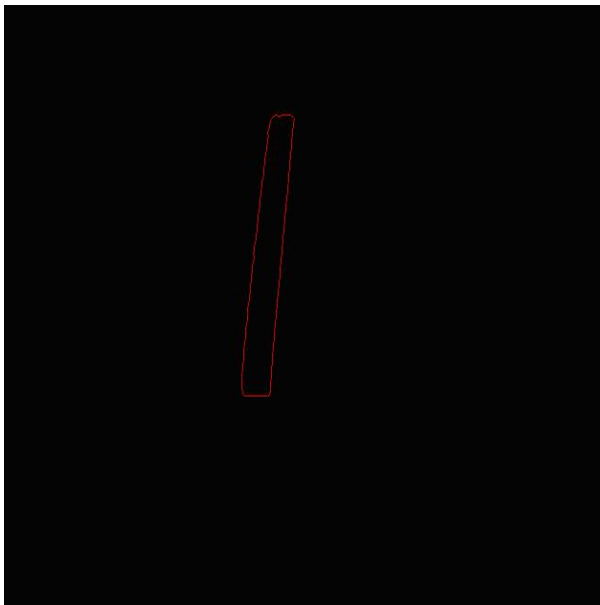
b) Imagen derecha

Figura 5.6 Grupo de instrumentos a identificar

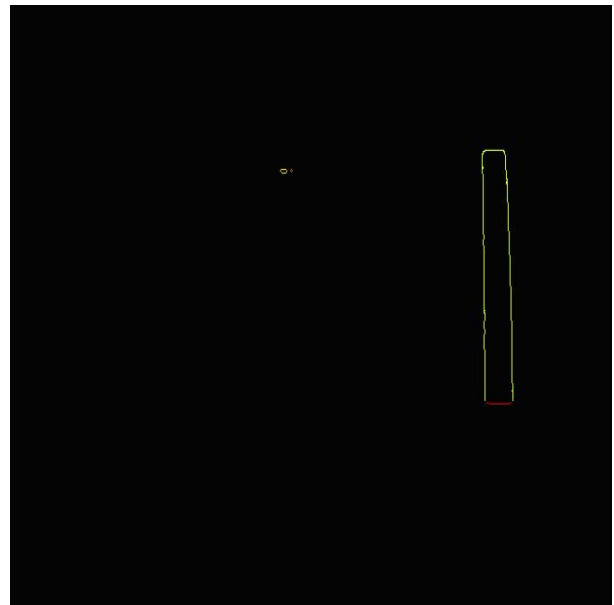
Si se quiere identificar el “Separador Farefeuf”, se tienen que evaluar las observaciones obtenidas de los dos objetos y verificar las probabilidades que se calculan, la probabilidad que sea mayor de las calculadas de las dos secuencias de observaciones corresponderá el objeto y por tanto es reconocido.

Modelo entrenado del Separador Farefeuf			
Imagen	Probabilidad del separador Farefeuf	Probabilidad de la pinza Metzemaum	Instrumento reconocido
Imagen derecha	0.61	0.42	Separador Farefeuf
Imagen izquierda	0.48	0.31	Separador Farefeuf

Tabla 5.4. Probabilidades dado el modelo entrenado del separador Farefeuf



a) Instrumento reconocido de imagen izquierda



a) Instrumento reconocido de imagen derecha

Figura 5.7 Instrumento identificado con modelo entrenado del Separador Farefeuf

5.3 Análisis de resultados en localización

En esta sección, se muestran los resultados obtenidos en localización de los instrumentos quirúrgicos utilizando estereoscopía. Para poder calcular la distancia “z”, con respecto a la cámara inicialmente se escoge un punto del objeto para el que se desea calcular su distancia. El punto escogido del objeto es el centro de masa (figura 5.2). Como se aprecia en la figura 5.8, el objeto se encuentra solamente desplazado en el eje x, el eje “y” queda fijo para las dos imágenes, por tal motivo no es necesario calcular la correspondencia. Si embargo para el cálculo de la distancia, si es necesario calculas su disparidad con respecto a cada una de las imágenes.



Figura 5.8 centro de masa del objeto

Para realizar la comprobación de la localización de los instrumentos quirúrgicos, se hicieron pruebas de cálculo con cada uno de los instrumentos quirúrgicos a distintas distancias.

En la figura 5.9 se muestra el Mango Bisturí, con las imágenes de los dos focos de la cámara (derecho e izquierdo). Estas dos imágenes calculando su disparidad el cual nos permite cálculo su distancia “z”, por estereoscopía



b) Imagen izquierda

a) Imagen derecha

Figura 5.9 imágenes de Mango Bisturí

De la misma manera, en la figura 5.10, se muestran las dos imágenes de la Tijera de Metzemberbaum, donde el punto rojo, al igual que la imagen anterior representa el centro de masa de dicho instrumento, el cual sus coordenadas en “x”, de los puntos de cada imagen nos permiten calcular la disparidad que hay entre ellos.



a) Imagen izquierda

b) Imagen derecha

Figura 5.10 imágenes de pinzas Metzenbaum

En la figura 5.11, muestra otro de los instrumentos llamado Pinza Kelly curva, al cual se le calcula la disparidad de los puntos centrales de cada imagen y se obtiene la distancia por estereoscopía.

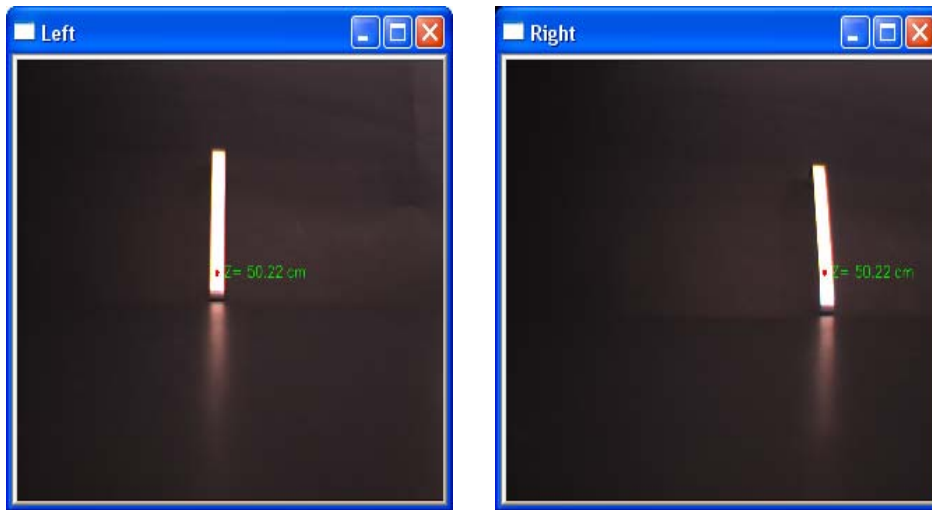


a) Imagen izquierda

b) Imagen derecha

Figura 5.11 pinza Kelly curva.

Finalmente, en la figura 5.12, calculamos la distancia con respecto a la cámara del último instrumento quirúrgico.



a) Imagen izquierda

b) Imagen derecha

Figura 5.12 Separador Farefeuf.

En la tabla 5.5, se muestra los cálculos de distancia real de cada uno de los instrumentos con respecto a la cámara, la distancia calculada y el error de la distancia calculada.

Instrumento	Distancia real (cm)	Distancia calculada (cm)	Error (cm)
Mango Bisturí	57.1	58.48	1.38
Tijera de Metzenbaum	45.3	46.54	1.24
Pinza Nelly curva	30	32.92	2.92
Separador Farefeuf	50	50.22	0.22

Tabla 5.5 Tabla de distancias calculas por estereoscopía.

5.4 Análisis de resultados en seguimiento

Para realizar la comprobación del seguimiento se diseño un experimento en donde se tiene un fondo e iluminación controlada. Se coloca cada uno de los instrumentos de forma individual en la escena controlada y se aplica el algoritmo camshift a la imagen binarizada (figura 5.13).

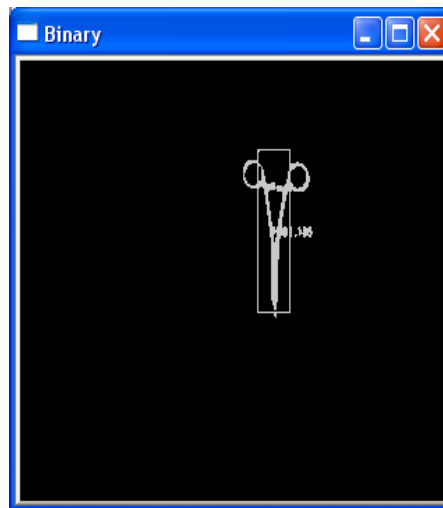
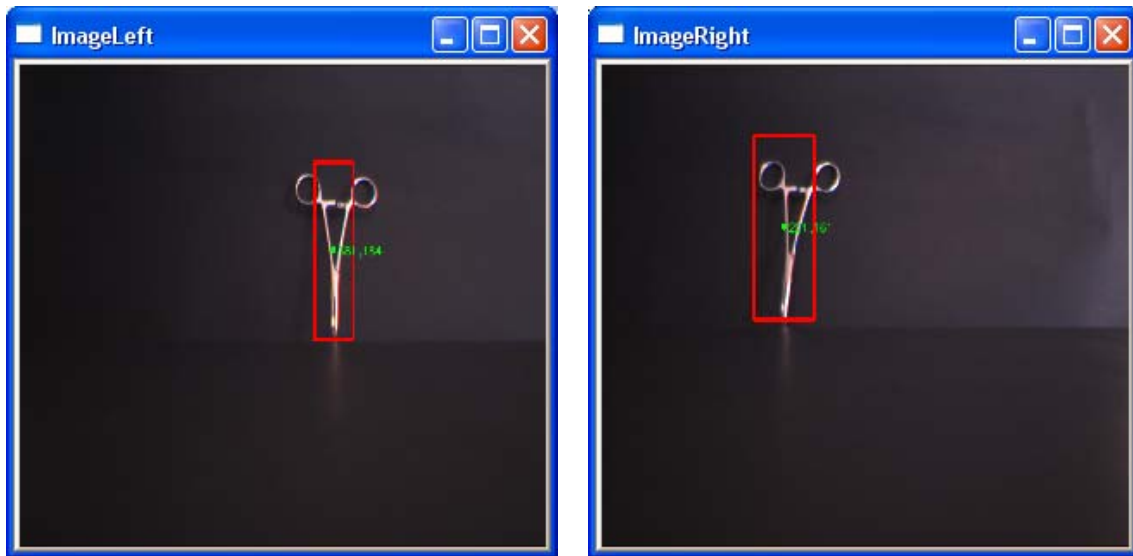


Figura 5.13 imagen binarizada

Una vez obtenida la imagen binaria, se pasa la imagen al algoritmo camshift, el cual se encarga de seguir el área que tenga la mayor cantidad de píxeles válidos que en este caso son del instrumento quirúrgico binarizado.

La figura 5.14 muestra el instrumento encerrado en un rectángulo de color rojo, lo cual significa que el algoritmo lo ha localizado en las dos imágenes capturadas de la cámara.

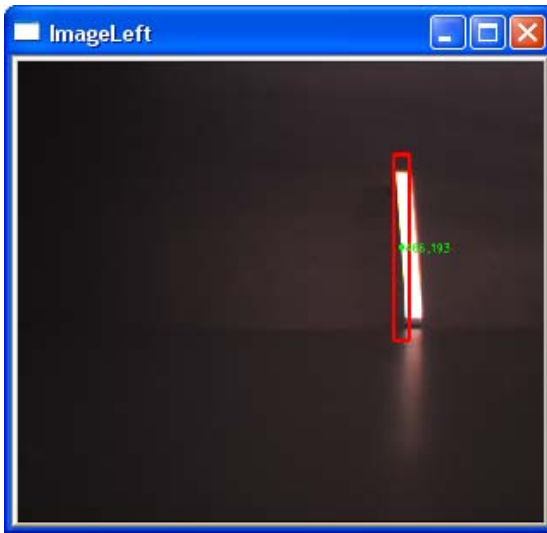


a) Imagen izquierda

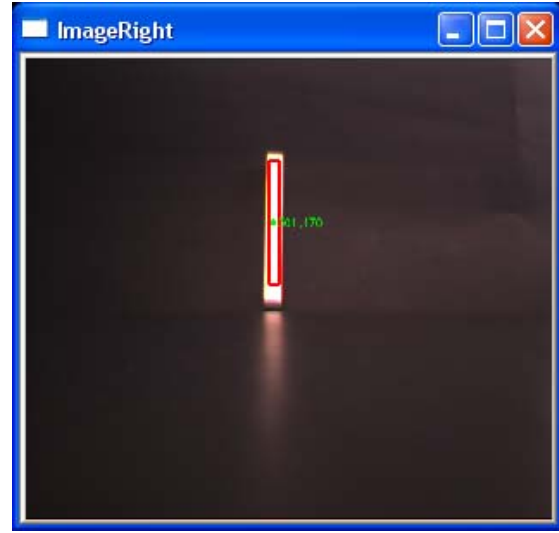
b) Imagen derecha

Figura 5.14 Seguimiento de Pinza Kelly curva

La figura 5.15, muestra el Separador Farefeuf, que de la misma forma que el instrumento anterior es encerrado en un rectángulo de color rojo durante una secuencia de imágenes, esto quiere decir que entre cada cuadro de la imagen capturada, el algoritmo identifica el objeto en la escena.



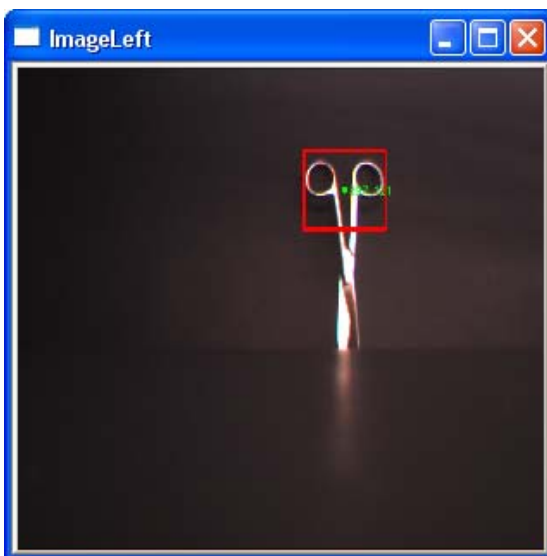
a) Imagen izquierda



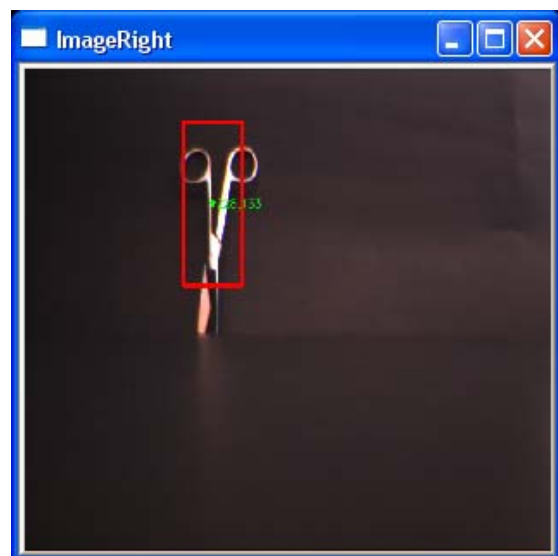
b) Imagen derecha

Figura 5.15 Separador Farefeuf.

De la misma forma que el instrumento anterior, en la figura 5.16 muestra otro de los instrumentos planteados, se binariza el instrumento y se aplica el algoritmo camshift, el cual nos permite identificar el instrumento en una secuencia de imágenes.



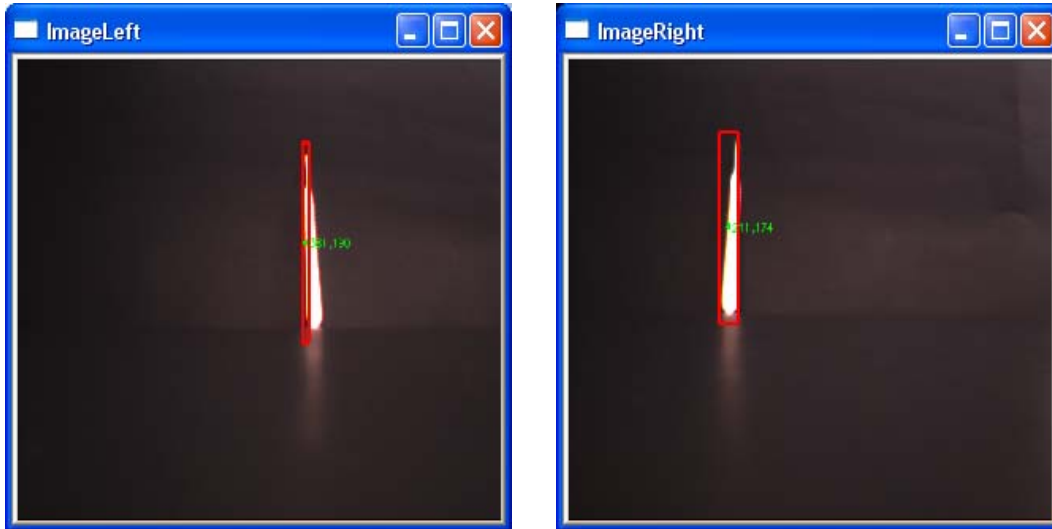
a) Imagen izquierda



b) Imagen derecha

Figura 5.16 Pinza Metzenbaum.

Finalmente, en la figura 5.17, colocamos el último instrumento quirúrgico en la escena y realizamos la misma metodología que los anteriores instrumentos quirúrgicos.



a) Imagen izquierda

b) Imagen derecha

Figura 5.17 Mango Bisturí

5.4 Resumen

En este capítulo se abordaron los experimentos realizados de esta tesis. Se presentan los experimentos de forma individual para mayor explicación. Sin embargo es importante aclarar que las condiciones bajo las que se realizaron los experimentos fueron controladas tanto en iluminación como en el fondo, que para este caso se utiliza uno que discrimina de la mejor forma el objeto de la escena.

Los resultados fueron buenos en localización y seguimiento. En la sección de reconocimiento, solo se muestra el entrenamiento y el reconocimiento con dos instrumentos quirúrgicos, los cuales nos arrojan cualitativamente buenos resultados de reconocimiento. Sin embargo se requiere experimentación adicional para cuantificar dichos resultados.

Capítulo 6

CONCLUSIONES Y TRABAJOS FUTUROS

Este estudio tuvo particular interés en tres tareas específicas de la visión computacional: El reconocimiento, la estereoscopia, y el seguimiento de objeto. Sin embargo no se pretende con esta tesis el descubrir nuevas técnicas en cada tarea específica, sino, el utilizar las técnicas ya desarrolladas y ampliamente probadas en una tarea específica, como la de reconocer instrumental quirúrgico, conocer su profundidad con respecto a la cámara estéreo y seguir los instrumentos en una secuencia de imágenes.

Mediante las pruebas experimentales se demostró que si es posible utilizar los HMM para el reconocimiento de los instrumentos quirúrgicos, además, utilizando la cámara estereoscópica también es posible conocer su distancia con respecto a la cámara, en cuanto al seguimiento, las pruebas experimentales demostraron que el instrumento puede ser localizado en el plano de la imagen en una secuencia de imágenes, por su puesto con las limitantes de captura de video que se tienen.

En cuanto a trabajos futuros, aun hay mucho por realizar en este trabajo, ya que se pueden realizar los experimentos en diferentes escenarios, con distintos tipos de fondos. También es necesario mejorar la segmentación de la imagen, probar con distintas técnicas de descripción de forma o textura. Entrenar con HMM mas instrumentos quirúrgicos y probarlos con un conjunto mayor de instrumentos, en cuestión del cálculo de la distancia mejorar aspectos de oclusión de instrumentos traslapados, mejorar la reconstrucción del objeto. En el seguimiento mejorar la calidad de captura del instrumento, también seria una buena oportunidad para agregarle seguimiento 3D.

Como he mencionado hay mucho trabajo por realizar, sin embargo, considero que es una oportunidad para desarrollar investigación científicas y de gran aplicabilidad a problemas reales.

REFERENCIAS:

[Anil et al, 2000] Anil K. Jain, Fellow, IEEE, Robert P.W. Duin, and Jianchang Mao, Senior Member, IEEE. Statistical Pattern Recognition: A Review. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 22, NO. 1, JANUARY 2000

[Anjum y Aggarwal, 2001] Anjum Ali and J. K. Aggarwal. Segmentation and recognition of continuous human activity. IEEE Computer Society's Conference on Computer Vision and Pattern Recognition, page 28-35, 2001.

[Baum et al, 1970] Baum, L.E., Petrie, T.E, Soules, G., Weiss, N. A maximization techniques occurring in the stactical analysis of probabilistic functions of Markov chain, Annals of Math Statistics. 41(1)(1970) 164-171.

[Baum2, 1970] Baum, L.E: An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. Inequality 3(1970)1-8.

[Bellegarda et al, 1993] E. J. Bellegarda, J. R. Bellegarda, D. Nahamoo, y K. S. Nathan. A probabilistic framework for on-line handwriting recognition. En Proc. of the 3rd International Workshop on Frontiers in Handwriting Recognition, pp. 225–234, Bu_alo, USA, 1993.

[Bertrum y Crow, 1984] R. J. Bartrum y H. C. Crow. Transillumination light scanning to diagnose breast cancer: a feasibility study. American Journal of Radiology, 142, pp. 409–414, 1984.

[Bhattacharya y Majumder, 2000] M. Bhattacharya y D. D. Majumder. Registration of CT and MR images of Alzheimer's patient: a shape theoretic approach. Pattern Recognition Letters, 21(6-7), pp. 531–548, 2000.

[Bourne, 1981] J. R. Bourne, V. Jagannathan, B. Hamel, B. H. Jansen, J. W. Ward, J. R. Hughes, y C. W. Erwin. Evaluation of a syntactic pattern recognition approach to quantitative electroencephalographic analysis. *Electroencephalography and Clinical Neurophysiology*, 52, pp. 57–64, 1981.

[Bradsky, 1998] Gary R. Bradsky. Computer vision face tracking for use in a perceptual user interface. *Intel Technology journal Q2'98*, pages 1-15, 1998.

[Cabrera y Mosso, 2005] Joaquin Cabrera, Alejo Mosso. *Sensor Visual de posicionamiento 3D*. ITESM, campus cuernavaca, 2005.

[Campbell y Flynn, 2001] Richard J. Campbell and Patrick J. Flynn. A survey of Free-Form Object Representation and Recognition Techniques. *Computer Vision and Image Understanding*. pp. 166-210, 2001.

[Canny, 1986] J. Canny. A computational approach to edge detection. *IEEE Trans. On PAMI* vol. 8(6),pp. 679-698, noviembre 1986.

[Climent y Mares] Joan Climent, Pere Mares. *Sistema de seguimiento en tiempo real para intervenciones quirúrgicas asistidas*. Departamento de ingeniería de sistemas automática e informática industrial, Universidad Politécnica de Cataluña.

[Deans et al, 2005] M. Deans, C. Kunz, R.Sargent, E. Park and L. Pedersen. Combined feature based shape visual tracker for robot navigation. *QSS Group/ Anatomy and Robotic Area NASA Ames Research Center*. 2005.

[Donal y Tanguay, 1993] Donald O. Tanguay, Jr. *Hidden Morkov Models for Gesture Recognition*, Massachusetts Institute of Technology, Cambridge, MA. Junio 1993

[Douglas, 1998] (name)Tesis: Generation, estimation and tracking of face. University of Pennsylvania. 1998.

[Freeman y Saghri,1978] H. Freeman and A. Saghri. Generalized chain codes for planar curves. In *Proceedings of the 4th International Joint Conference on Pattern Recognition*, pages 701-703, Kyoto, Japan, November 7-10 1978.

[Friedman et al, 1999] Friedman M, Kandel A. Introduction to Pattern Recognition: Statistical, Structural, Neural and Fuzzy Logic Approaches (Series in Machine Perception and Artificial Intelligence) World Scientific Publishing Company. 1999.

[Gil et al, 2004] P.Gil, F.Torres, F.G. Ortiz. Detección de Objetos por segmentación multinivel combinada de espacios de color. Universidad de Alicante, 2004.

[Hu et al, 1996] Hu., J. Brown, M.K., Turin, W. HMM based on-line handwriting recognition. *IEEE Trans Pattern Analysis and Machine Intelligence*, 18(10)(1996) 1039-1045.

[Hughey y Krogh, 1996] Hughey, R., Krogh, A. Hidden Markov Model for sequence análisis: extensión and análisis of the Basic method. *Comp. Appl. Inthe Biosciences* 12 (1996) 95-107.

[Hurteau et al, 1994] R. Hurteau, S. DeSantios, E. Begin, M. Gagner, Laparoscopic Surgery assisted by a robotic cameraman: Concept and experimental results, *Proc. IEEE Int. Conf. Robotc and Automation*, pp. 2286-2289, San Diego, May 8-13, 1994.

[Hough] P.V.C Hough. Method and Jeans for Recognizing Complex Patterns. U.S. Patent No. 3069654.

[Jia-Ching y Moura, 1998](name) Automatic Recognition of Human Walking in Monocular Image Sequences, Departament of Electrical and Computer Engineering, Carnegie Mellon University, 1998.

[Jonson y Herber, 1998] A. E. Johnson y M. Hebert. Surface matching for object recognition in complex threedimensional scenes. Image and Vision Computing, 16(9-10), pp. 635–651, 1998.

[Kanal, 1974] Laveen Kanal. Pattern in Pattern Recognition. IEEE Transactions. Invited Paper, 1974.

[Krupa et al, 2002] A. Krupa, J. Gangloff, M de Mathelin, C. Doignon, G. Morel, L. Soler, J. Leroy, J. Marescaux. Autonomous Retrieval and positioning of surgical instruments in robotized laparoscopic surgery using visual servoing and laser pointers. International Conference on robotics and Automation. May. 2002.

[Laveen, 1974] Laveen Kanal, Fellow. Patterns in Pattern Recognition: 1968-1974. Invited Paper, IEEE. Transactions on Information Theory, VOL. IT-20, No. 6, November 1974.

[Lepetit y Fua, 2005] (name)Monocular Model-Based 3D Tracking of Rigid Objects: A Survey. Computer vision Laboratory , Lausanne, Switzerland. 2005.

[Matsumoto et al,81] K. Matsumoto, M. Naka, y H. Yamamoto. A new clustering method for landsat images using local maximums of a multidimensional histogram. En Proc. of the Symposium on Machine Processing Remotely Sensed Data, pp. 321–326, Purdue - IN, USA, 1981.

[Matsushita et al, 1998] (nane) A tracking based manipulation system build on stereo vision. Intelligent System Division Electrotechnical Laboratory, Tsukuba, JAPAN, 1998.

[McCulloch y Pitts, 1943] McCulloch, W. S. and Pitts, W. H. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115-133., 1943.

[McKenna et al, 2004] S.J. McKenna, H. Nait Charif, and T. Frank. Toward Video Understanding of laparoscopic Surgery: Instrument Tracking. Scotland. 2004.

[Mosso y Frausto, 2005] Mosso, A. y Frausto, J. Aplicaciones de cirugía asistida por computadora y robótica médica. PROYECTO SALUD-2002-C01-8181 CONACYT- ISSSTE-ITESM-UNAM-UAM, 2005-2007.

[Pajares y de la Cruz, 2002] Gonzalo Pajares, Jesús M. de la Cruz. Visión por computador. Imágenes digitales y aplicaciones. Ed. Ra-Ma, 2002

[Penelope, 2005] Penelope. Robotic Surgical Tech Inc.

Disponible en: <http://www.roboticsurgicaltech.com/index.html>

[Rabiner, 1989] L.R. Rabiner, 1989. A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceeding of the IEEE*, 77 (2),1989.

[Rabiner y Juang, 1986] L.R. Rabiner, Juang B.H. An Introduction to hidden Markov models. *IEEE ASSP*, Enero 1986.

[Rosen y Nitzan, 1977] C. A. Rosen y D. Nitzan. Use of sensors in programmable automation. *Computer*, 10(12), pp. 12–23, 1977.

[Sandoval, 2003] Cristina Sandoval Alonso, Tesis: Seguimiento de objetos en secuencias de video. Universidad Politecnica de valencia, pp, 1-7. 2003.

[Shapiro et al, 2001] Linda G. Shapiro, George C. Stockman. "Computer vision". Prentice Hall, 2001.

[Sossa et al, 2005] Humberto Sossa, Roberto A. Vázquez, Ricardo Barrón. Reconocimiento y localización de instrumental médico usando análisis automatizado de imágenes. Centro de investigación en computación-IPN, 2005.

[Taylor et al, 1995] R.H. Taylor, J. Funda, B. Eldridge, S. Gomory, K. Gruben, A telerobotic assistant for laparoscopic surgery, IEEE Engineering in Medicine and Biology, 14(3), pp. 276-288, 1995.

[Thompson y Kagami, 2005] Simon Thompson y Satoshi Kagami. Humanoid Robot Localization using Stereo Vision. Digital human research center. National Institute of Advanced Industrial Science and Technology, Aomi, Koto-ku, Japan. IEEE 2005.

[Vezhnevets et al] Valdimir Vezhnevets, Vassili Sazonov, Alla Andreeva. A Survey on Píxel-Based Skin Color Detection Techniques. Faculty of Computational Mathematics and Cybernetics, Moscow State University.

[Wei et al, 1997] Gou-Quig Wei, Klaus Arbter, Gerd Hirzinger. Real-Time visual servoing for laparoscopic surgery. German Aerospace Research Establishment, Institute of Robotics and System Dynamics, DLR Oberpfaffenhofen. IEEE Engineering in Medicine and Biology. Jenuary / February , 1997.

[West y Maurer, 2004] Jay B. West and Calvin R. Maurer, Jr. , Designing Optically Tracked Instruments for Image-Guided Surgery. IEEE Member, 2004.

Apéndice A

Se diseñó una interface de usuario que englobara cada uno de los 3 problemas que se abordaron, la interface contiene las funciones que permiten a un usuario manejar de una manera sencilla y vistosa el reconocimiento y la localización de un instrumento quirúrgico que se encuentre en la escena.

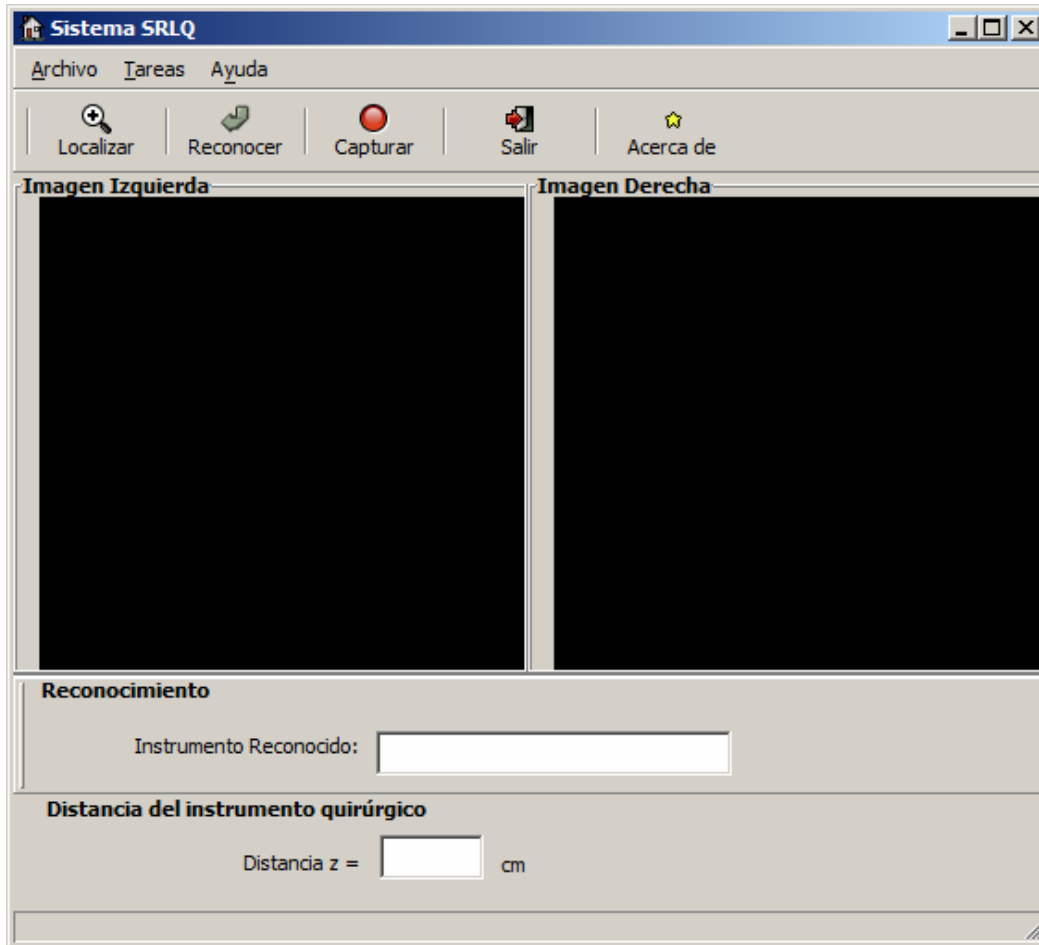


Figura A.1 Pantalla que muestra el sistema SRLQ

El sistema cuenta con un menú principal, con tres opciones: Archivo, Tareas y Ayuda, cada uno de las opciones contienen tareas específicas dentro del sistema.



Figura A.2 Menú principal del sistema SRLQ

También se cuenta con una barra de herramientas para un manejo rápido de las tareas más importantes, como es la de localización del instrumento quirúrgico, reconocimiento del instrumento quirúrgico, iniciar la captura en de la cámara estéreo, salir del sistema y la información básica den sistema.

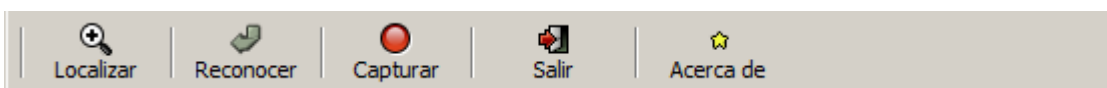
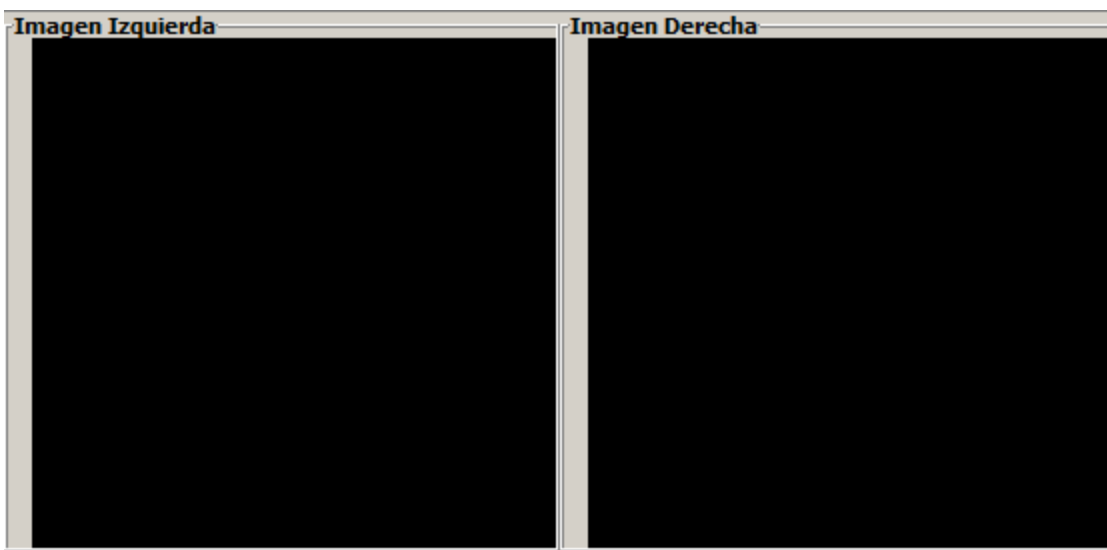


Figura A.3 Barra de herramientas del sistema SRLQ

El sistema cuenta con dos áreas de dibujo de imagen, en donde se muestra el renderizado de las imágenes capturadas de la cámara estéreo, uno para la imagen derecha y el otro para la imagen izquierda.



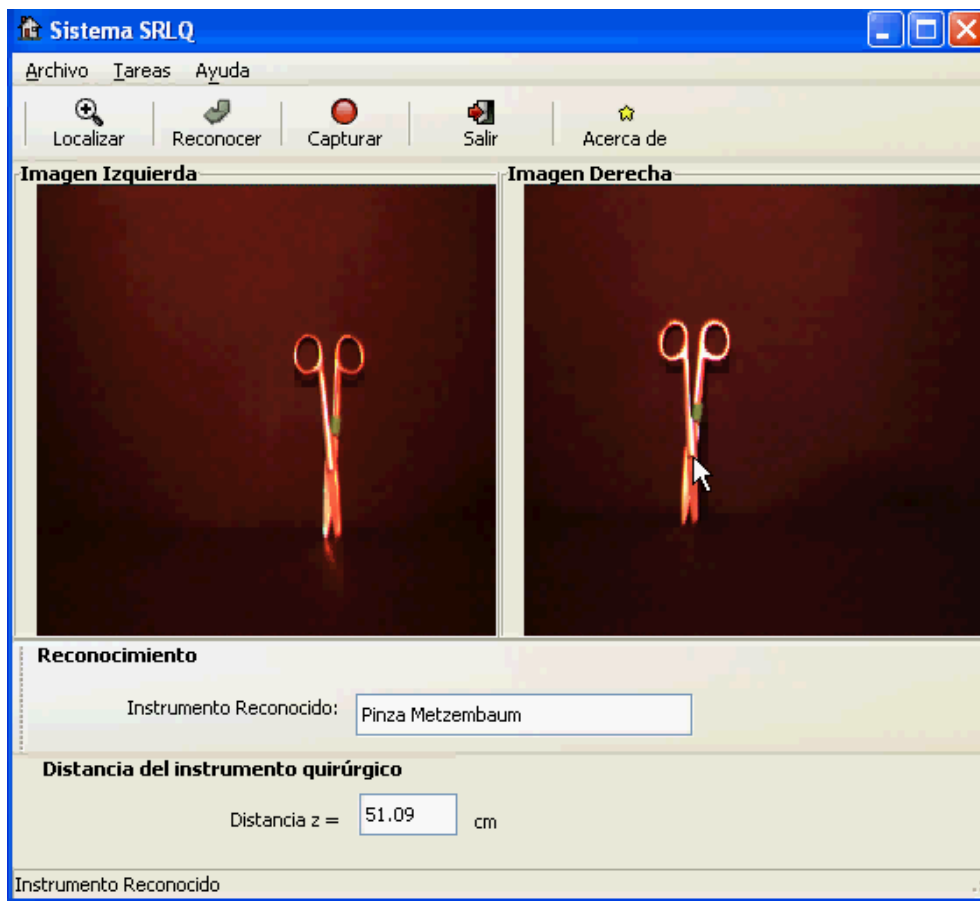
A.4. Áreas de renderizado de las imágenes de captura

Finalmente, se cuenta con una sección que muestra el instrumento que se reconoció y la distancia del mismo con respecto a la cámara estéreo.

Reconocimiento
Instrumento Reconocido: <input type="text"/>
Distancia del instrumento quirúrgico
Distancia z = <input type="text"/> cm

A.5 Información de salida del sistema

En la siguiente pantalla se muestra la imagen del sistema funcionando, reconociendo el instrumento quirúrgico y calculando su distancia con respecto a la cámara estéreo.



A.6 Pantalla del sistema funcionando en línea.