ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

FACULTAD DE INGENIERÍA EN ELECTRICIDAD Y COMPUTACIÓN

TESIS DE GRADO

"ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN SISTEMA DE GRABACIÓN DE EXPOSICIONES CON SEGUIMIENTO AUTOMATIZADO DEL PRESENTADOR UTILIZANDO VIDEO MONOCULAR"

Previa a la obtención del titulo de:

INGENIERO EN COMPUTACIÓN ESPECIALIZACIÓN SISTEMAS MULTIMEDIA

PRESENTADA POR:

JOSÉ FABIAN VILLA VÁSQUEZ

GUAYAQUIL - ECUADOR

2007

AGRADECIMIENTO

A todos quienes

contribuyeron directa o

indirectamente al desarrollo

de este trabajo

.

DEDICATORIA

A mis padres

TRIBUNAL DE GRADO

PRESIDENTE

Ing. Holger Cevallos Ulloa

DIRECTOR DE TESIS

Msc. Xavier Ochoa Chehab

MIEMBROS PRINCIPALES

Ing. Cristina Abad

Ing. Federico Raue

DECLARACIÓN EXPRESA

"La responsabilidad por los hechos, ideas y doctrinas expuestas en esta tesis, nos corresponden exclusivamente; y, el patrimonio intelectual de la misma, a la Escuela Superior Politécnica del Litoral"

(Reglamento de exámenes y títulos profesionales de la ESPOL)

José Fabian Villa Vásquez

RESUMEN

En la introducción se presenta una perspectiva general de los sistemas de visión por computador y la manera en que son aplicados en el análisis del movimiento de los seres humanos. Adicionalmente se incluye una breve descripción del problema general que este trabajo tiene por objeto resolver.

En el primer capítulo se realiza un profundo análisis del problema, y se presenta la motivación y justificación para llevar a cabo este trabajo. Por otra parte, en este mismo capitulo se introduce el planteamiento de la solución a dicho problema y el alcance que este trabajo tendrá.

El segundo capítulo presenta el marco de referencia que proveerá el soporte necesario de conocimiento para el desarrollo de este sistema; es decir, se realiza un minucioso estudio de las técnicas y enfoques que son usualmente utilizados en el desarrollo de este tipo de sistemas. Este estudio es presentado de una manera estructurada y clasificada según cada tarea principal que el sistema deberá realizar.

El Análisis y Diseño del sistema se presenta en el tercer capítulo. En este se realiza el respectivo análisis de requerimientos del sistema y se presenta la arquitectura del sistema desde un punto de vista global y según sus módulos, la misma que muestra el diseño modular que tendrá el sistema. Los

diferentes módulos que conforman el sistema son explicados en detalle en este capitulo.

El cuarto capitulo trata de la implementación del sistema. Aquí se realiza la selección de las herramientas que serán utilizadas para el desarrollo del mismo. Adicionalmente se explica en detalle la manera en que será implementado cada modulo del sistema y la interacción entre los mismos. En este capitulo también se detallarán los principales problemas encontrados durante el desarrollo del sistema.

En el quinto capitulo se presentan las pruebas realizadas al sistema y se realizará un análisis de los resultados de las mismas y posteriormente se presentan las conclusiones de estas pruebas.

El sexto capitulo está dedicado para el análisis de la utilidad del sistema. Se presentan sus principales aplicaciones y adicionalmente se realiza un pequeño estudio de la utilidad de cada modulo del sistema.

Finalmente están las conclusiones de este trabajo y se incluyen los anexos y el detalle de la bibliografía utilizada para que el desarrollo de este trabajo sea posible.

ÍNDICE GENERAL

AGRADECIMIENTO ii
DEDICATORIA iii
TRIBUNAL DE GRADO iv
DECLARACIÓN EXPRESA v
RESUMEN vi
ÍNDICE GENERAL viii
ÍNDICE DE GRÁFICOS xi
ÍNDICE DE TABLAS xii
INTRODUCCIÓN 1

1 ANÁLISIS DEL PROBLEMA Y ALCANCE DE LA SOLUCIÓN 7
1.1 El Problema 7
1.2 Motivación 11
1.3 Justificación 13
1.4 Planteamiento de la Solución 15
1.5 Alcance 18
1.6 Objetivos del Proyecto 21
1.7 Sumario 22
2 MARCO DE REFERENCIA 24
2.1 Factibilidad de la Implementación 24
2.1.1 Problemas Típicos 25
2.1.1.1 Factores que Influyen en la Precisión 27
2.1.1.2 Limitaciones y Restricciones 29
2.2 Adquisición de Video 32
2.2.1 Importancia de la Calidad del Video 33
2.2.2 Tipos de Fuentes de Video 34
2.3 Reconocimiento de Objetos 36
2.3.1 Segmentación Fondo-Objetos 37
2.3.1.1 Técnicas basadas en información temporal 37
2.3.1.2 Técnicas basadas en información espacial 41
2.3.2 Análisis y Clasificación de Objetos 44
2.3.2.1 Técnicas de Bajo Nivel 46
2.3.2.2 Técnicas de Nivel Medio 48
2.3.2.3 Técnicas de Alto Nivel 48
2.3.3 Representación de Objetos 49
2.3.3.1 Representación Basada en Objeto 49
2.3.3.2 Representación Basada en Imagen 51
2.3.4 Eliminación de Objetos 52
2.3.5 Análisis de los Enfoques más Adecuados 54
2.4 Seguimiento de Objetos 56
2.4.1 Objetos en Movimiento 57
2.4.2 Distinción entre Objetos y Personas 60

2.4.3 Modelos y Patrones de Movimiento de Personas 62
2.4.3.1 Representación por Figura Simple 62
2.4.3.2 Modelamiento por Contornos 2D 63
2.4.3.3 Modelos Volumétricos 3D 65
2.4.4 Análisis de los Enfoques más Adecuados 66
2.5 Interpretación del Movimiento 67
2.6 Control de la Fuente de Video 71
2.6.1 Interfaz Serial 71
2.6.2 Protocolos de Comunicación Serial 73
2.7 Transmisión de Video por Internet 76
2.8 Sistemas HMT - Human Motion Tracking 76
2.8.1 Campos de Aplicación de los Sistemas HMT 76
2.8.2 Algunos Proyectos Actuales 79
2.9 Herramientas Disponibles 82
2.9.1 Lenguajes de Programación 82
2.9.2 Librerías 84
2.10 Sumario 88
3 ANÁLISIS Y DISEÑO 90
3.1 Introducción 90
3.2 Análisis de Requerimientos 91
3.2.1 Requerimientos Funcionales 92
3.2.2 Requerimientos No Funcionales 94
3.3 Diseño Modular 95
3.4 Arquitectura 96
3.4.1 Modelo General del Sistema 98
3.4.2 Adquisición de Video 101
3.4.3 Reconocimiento 103
3.4.4 Seguimiento 106
3.4.5 Control de Cámara 108
3.4.6 Transmisión y Grabación de Video 111
3.5 Sumario 112
4 IMPLEMENTACIÓN 114
4.1 Selección de las Herramientas 114
4.2 Construcción del Sistema 115
4.2.1 Módulo de Adquisición 117
4.2.2 Módulo de Reconocimiento 119
4.2.2.1 Fase de Inicialización 120
4.2.2.2 Sustracción de Fondo 123
4.2.2.3 Segmentación de Objetos 125
4.2.3 Módulo de Seguimiento 128
4.2.4 Módulo de Control de Cámara 131
4.2.4.1 Establecer la Comunicación 133
4.2.4.2 Detección de Fallos 137
4.2.5 Módulo de Transmisión 138

4.3 Problemas Encontrados 139
4.4 Sumario 142
5 PRUEBAS 144
5.1 Pruebas de Campo 144
5.2 Análisis de Resultados 153
5.3 Conclusiones de las Pruebas 156
6 APLICACIONES 159
6.1 Aplicaciones Generales del Sistema 159
6.2 Utilidad de los Módulos 161
6.2.1 Posibles Aplicaciones de los Módulos 162
6.3 Otras Aplicaciones 164
CONCLUSIONES Y RECOMENDACIONES 170
A APÉNDICE A: MANUAL DEL USUARIO 179
A.1 Manual 179
REFERENCIAS DE GRÁFICOS 186
REFERENCIAS BIBLIOGRÁFICAS 187

ÍNDICE DE GRÁFICOS

Figura 1.1. Secuencia de imágenes del experimento de Johansson [F1]..... 14
Figura 2.1. Niveles de clasificación [F2]..... 46
Figura 3.1. Casos de Uso. 92
Figura 3.2. Arquitectura del Sistema. 100
Figura 3.3. Mecanismo de Adquisición de Imágenes. 102
Figura 3.4. Mecanismo de Reconocimiento de la Persona. 104
Figura 3.5. Mecanismo del Seguimiento. 107
Figura 3.6. Mecanismo del Control de Cámara. 110
Figura 3.7. Mecanismo de Transmisión de Video por Video Conferencia. . 111
Figura 4.1. Cámara de Video Canon VC-C50i [F3]. 116
Figura 4.2. Sustracción de Fondo. 124
Figura 4.3. Píxeles vecinos de p. 126
Figura 4.4. Barrido de izquierda a derecha. 126
Figura 4.5. Segmentación de Objetos. 128
Figura 4.6. Posición del objetivo encontrado.. 130
Figura 4.7. Conectores utilizados por el estándar RS-232C. F4]. 132
Figura 4.8. Asignación de nueva posición de la cámara. 135
Figura 5.1. Aplicación principal en ejecución. 145
Figura 5.2. Ventana de configuración para el proceso de inicialización. 146
Figura 5.3. Comienzo del proceso de inicialización. 147
Figura 5.4. Proceso de inicialización en curso. 147
Figura 5.5. Proceso de inicialización finalizado. 148
Figura 5.6-a. Seguimiento a la persona. Pantalla principal. 149
Figura 5.6-b. Seguimiento a la persona. Proceso de reconocimiento. 149
Figura 5.7. Persona próxima al límite izquierdo. 150
Figura 5.8. La persona ha sobrepasado el límite izquierdo.. 151
Figura 5.9. Selección del códec de compresión de video. 152
Figura 5.10. Grabación de video. 152
Figura 6.1. Extracción de trayectorias de vehículos. 166
Figura 6.2. Control de brazo robótico. 167
Figura 6.3. Seguimiento a un vehículo en carretera [F5]. 168
Figura A.1.1 Interfaz del Usuario. 179
Figura A.1.2 Ventana de configuración. 182
Figura A.1.3 Seguimiento activado. 185

ÍNDICE DE TABLAS

Tabla 2.1: Restricciones [1] utilizadas por sistemas HMT..... 32
Tabla 4.1: Funciones más importantes de los pines del conector DB-25 ... 132
Tabla 4.2: Señales asociadas a los pines de DB-25 y DB-9..... 133

INTRODUCCIÓN

Los sistemas de video inteligentes basados en el reconocimiento y seguimiento de objetos tienen muchas aplicaciones actualmente y están siendo cada vez más utilizados en diversos campos de estudios. Lo que ha motivado este gran interés por el análisis del movimiento de los seres humanos es la capacidad de ser utilizado en un amplio espectro de aplicaciones que básicamente están inmersas en tres grandes áreas, según lo demuestra el trabajo de Moeslund y Granum [1]. Estas áreas son: la Vigilancia, con aplicaciones como detección de objetos robados y detección de situaciones sospechosas, el Área de Control, que esencialmente trata de las interfaces hombre-máquina con aplicaciones como la captura de movimiento útil para videojuegos y producciones cinematográficas, y la tercera área de aplicación que se enfoca en el Análisis del movimiento humano para obtener información semántica, como el análisis del rendimiento atlético.

Estos sistemas se desarrollan mediante la implementación de video sensores, cuyo desarrollo tiene su base en el procesamiento digital de las imágenes que provienen de una fuente de video. Un video sensor no es más que una herramienta de análisis de video digital que ofrece información significativa proveniente de una secuencia de video.

Cuando obtienen la secuencia de video proveniente de una o más fuentes, procesan esta secuencia y la analizan en un menor o mayor grado dependiendo de la aplicación, para así obtener importante información de la escena y los actores. Incluso se pueden identificar las acciones o las intenciones de los actores en la escena analizando gestos, posturas corporales y expresiones faciales.

Es aquí cuando la Visión por Computador entra en juego. De hecho, es posible utilizar otras tecnologías en conjunto para construir sistemas que se enfoquen en una aplicación en particular. Por ejemplo existen sistemas de asistencia para personas ancianas, que utilizan sensores de RFID para determinar la posición de objetos dentro de la residencia, lo cual podría realizarse mediante un análisis de las secuencias de video obtenidas por cámaras situadas en varios puntos de la residencia. Sin embargo, esto demandaría de un gran poder computacional y de grandes tiempos para el análisis. Es por esto que se utilizan otras tecnologías en conjunto con la Visión por Computador, para reducir costos sin sacrificar la eficiencia y rendimiento de un sistema de este tipo.

Por otra parte, el movimiento de los seres humanos es muy complejo por su naturaleza articulada y, al ser un movimiento no rígido, la detección y

seguimiento de personas se convierte en un tema de investigación muy desafiante. Existen dos enfoques típicos para el análisis del movimiento del ser humano y se diferencian en el uso de modelos o patrones de movimiento. Estos enfoques son: Basado en Modelos y No Basado en Modelos. En ambos enfoques la representación del cuerpo humano va desde simples figuras, contornos 2D, volúmenes 3D hasta modelos más complejos. La representación por figuras se basa en que el movimiento de las partes del cuerpo depende del movimiento de los huesos. Los contornos de 2D se basan en la proyección de un cuerpo en 3D en imágenes planas. En cambio los volúmenes 3D describen de manera más precisa el modelo del cuerpo humano utilizando conos, cilindros elípticos y esferas. Un modelo completo consiste tanto de los movimientos como de la forma del cuerpo, por lo que obviamente es necesario obtener esta forma del cuerpo. Para esto se pueden utilizar diversos métodos, dependiendo de la aplicación del sistema. Algunos métodos comunes son: escáner láser, escáner de infrarrojos, fotogrametría, luz estructurada. En cambio para modelar el movimiento del ser humano se utilizan procesos de seguimiento para capturar el movimiento. Pero el problema que surge es como seguir el movimiento, es decir, en que se debe basar para establecer la relación de la estructura de la imagen en cuadros consecutivos de la secuencia de video, y más aun cuando esta correspondencia debe ser obtenida automáticamente.

Entonces se puede resumir que el análisis del movimiento de objetos (el cuerpo humano es un objeto con articulaciones muy complejas) incluye la detección de regiones móviles, la estimación del movimiento de estas regiones, el modelamiento de las articulaciones de los objetos y finalmente la interpretación del movimiento. Pero esto no es sencillo ya que los seres humanos pueden adoptar diversas posturas y se pueden deformar de forma compleja, lo cual insinúa un cambio en la forma del objeto pero obviamente no por eso la persona deja de ser un humano. Además la apariencia de las personas puede cambiar drásticamente de un momento a otro (lo cual tampoco implica que el objeto haya cambiado), los puntos de referencia rastreados pueden traslaparse y confundirse con otros, lo cual provoca interpretaciones erróneas y ambigüedad. Adicionalmente estos puntos de referencia no siempre son visibles, ya que pueden estar escondidos detrás de la ropa de la persona, lo cual dificulta el seguimiento de los mismos.

Estos problemas son muy usuales en el desarrollo de sistemas de visión por computador en cualquiera de los campos de aplicación. Incluso existen restricciones específicas relacionadas con algunos de ellos. Más adelante en este trabajo se analizarán con mayor detalle estos problemas y las restricciones.

Para este proyecto se ha establecido como campo de aplicación el área de control, ya que básicamente se trata de crear un sistema de funcionamiento automático que requiera en lo mínimo la presencia de un operario y que tome decisiones en base al movimiento de la persona objetivo. El sistema deberá no solo detectar el movimiento sino también realizar el seguimiento de la persona, para lo cual se deberán definir características y puntos de referencia que faciliten el rastreo del movimiento.

Una vez que el seguimiento esté en curso, el sistema deberá hacer una estimación del movimiento para posteriormente hacer la interpretación del mismo. Como resultado de esta interpretación el sistema manipulará automáticamente la posición de la fuente de video para mantener siempre enfocado a la persona objetivo.

El sistema realizará todo este mecanismo utilizando solamente una fuente de video monocular, estática en el sentido de movimiento longitudinal pero dinámica en cuanto al movimiento rotacional. En principio la fuente de video contará con dos grados de libertad para su rotación.

En fin, en este trabajo se pretende realizar un estudio de las tecnologías existentes y de las herramientas disponibles para el desarrollo de un sistema de visión por computador, y construir un sistema que refleje la viabilidad de

poder llevar a cabo este tipo de proyectos en nuestra sociedad. Asimismo, se intenta que la comunidad estudiantil aumente su interés por el área de la visión por computador y fomentar a la investigación y desarrollo de nuevos proyectos.

CAPÍTULO 1

1 ANÁLISIS DEL PROBLEMA Y ALCANCE DE LA SOLUCIÓN

1.1 El Problema

A menudo es necesario realizar grabaciones en video de las presentaciones de un expositor en un auditorio o salón de exposiciones. En ocasiones se podría desear grabar una clase que es dictada por un profesor, el cual no necesariamente se encuentra en el salón de clases con sus alumnos, sino que podría estar dictando la clase mediante videoconferencia.

Actualmente para realizar esta tarea es necesario contar con la presencia de una persona que actúa como operario de la cámara de grabación, para que pueda enfocar en todo momento al expositor. Si la persona que esta dando su conferencia, o el expositor que hace su presentación, o incluso el profesor que dicta su clase se mueve de un lado a otro, el operario simplemente hace el respectivo seguimiento para mantenerlo siempre enfocado y para que sea el centro de atención. Si no se cuenta con el operario, la fuente de video siempre

se mantendrá estática y el campo de visión será siempre el mismo, obligando al expositor a mantenerse siempre en el rango de cobertura de la fuente de grabación.

Entonces en ocasiones en las que no se cuenta con el operario simplemente resulta imposible hacer este tipo de grabaciones, a menos que se cuente con un sistema de grabación inteligente que realice el seguimiento automatizado del movimiento de la persona. En el caso particular de las videoconferencias no solo es necesario hacer que la cámara enfoque continuamente al expositor sino que también se tiene que hacer la transferencia en tiempo real del video hacia su destino final. Este trabajo tiene como uno de sus objetivos crear un sistema de visión por computador que sea capaz de realizar este trabajo de manera automática y con un mínimo control del usuario sobre el sistema.

Con este sistema se trata de automatizar el proceso anteriormente descrito, haciendo que la fuente de video misma gire automáticamente para enfocar al expositor mediante un software que reconozca al individuo en el flujo de video obtenido de una fuente monocular y que calcule el movimiento de la persona, para de esta forma realizar una estimación de la rotación necesaria de la fuente de

video para que el expositor sea enfocado en todo momento de la forma más precisa posible.

El sistema debe proveer el software necesario para que un computador sea prácticamente el operario en el evento de la grabación. Para realizar esta tarea, deberá en primera instancia obtener una secuencia de video para procesarla y detectar movimiento en la escena. Posteriormente deberá identificar su objetivo, es decir la persona que hace la exposición, para luego poder mantener rastro de su movimiento. Además el sistema deberá ser capaz de obtener la información necesaria de la escena, para en base a aquello manipular la fuente de video.

Existen muchos enfoques distintos en cuanto al reconocimiento de seres humanos en una imagen, pero en lo que la mayoría coinciden es en que se debe realizar una segmentación para diferenciar al individuo del fondo de la imagen, luego transformar las imágenes segmentadas en alguna otra representación para reducir la cantidad de información y luego decidir la forma en que se le hará el seguimiento al individuo de cuadro en cuadro, es decir de imagen en imagen proveniente de a fuente de video.

Una vez que se logra tener control sobre la fuente de video, la grabación en curso debe ser transmitida directamente hacia su destino. Para esto será necesario utilizar un códec de video que haga la compresión del mismo para que su transmisión en línea sea posible.

Sin embargo, la tarea no es trivial, pues se deben tomar en cuenta varios factores que afectan al reconocimiento y posterior seguimiento de la persona en el flujo contínuo de imágenes provenientes de la fuente de video en forma secuencial. Entre los principales problemas con los se enfrentan este tipo de sistemas se pueden citar: la variación de la intensidad de luz, el movimiento de objetos que no son de interés, la distancia de la persona respecto de la ubicación de la cámara que captura el video, el contraste entre el expositor y el fondo de la escena, la posible variación del fondo de la escena pues es muy probable que se hagan proyecciones en una pantalla, la fuente de video puede no ser estática lo cual provoca que el fondo varíe en el tiempo, entre otros.

El propósito general que persigue el sistema es permitir la grabación automatizada de presentaciones con enfoque automático al expositor utilizando una sola entrada de video. Pero al ser un sistema

modularizado se tiene la oportunidad de utilizar un módulo en particular para utilizarlo como base o parte integral de otros sistemas de visión por computador, como son los sistemas de seguimiento del movimiento humano, conocidos como HMT (Human Motion Tracking), con aplicaciones en sistemas de vigilancia y detección de movimientos sospechosos, sistemas de interacción hombre-máquina mediante el reconocimiento de gestos, sistemas de detección de objetos robados y muchas otras aplicaciones.

Este sistema también tendrá que incorporar toda la funcionalidad necesaria para que lo que sea registrado por la cámara pueda ser grabado y transmitido automáticamente a través de Internet, haciendo un streaming de video.

1.2 Motivación

E Actualmente en nuestro medio la tecnología de videoconferencias está siendo muy utilizada para presentaciones y exposiciones a distancia. Incluso se está utilizando esta tecnología para el dictado de clases en las universidades.

En cualquier caso, la persona que se encuentra realizando su exposición a distancia, debe permanecer en una posición fija

manteniéndose en el campo de visión de la cámara de grabación, a menos que exista una persona que manipule la cámara para enfocar al expositor y de esta manera permitirle movimiento. Por este motivo, este trabajo tiene como uno de sus objetivos, crear un sistema que permita automatizar el proceso de grabación de una exposición por videoconferencia.

Por otra parte el área de visión por computador no tiene gran acogida en nuestro medio actualmente, siendo el procesamiento digital de imágenes el máximo nivel de estudios en este ámbito, según se puede advertir al observar el programa académico de las carreras universitarias ofrecidas actualmente relacionadas con este campo de estudio.

Esta tesis se enfoca entonces, en un pequeño subconjunto del área de la visión por computador, utilizando técnicas de detección y reconocimiento de objetos y su movimiento para crear un sistema de grabación automatizado que responda a las acciones del usuario. Aunque en principio las acciones que reconocerá el sistema se limitan al seguimiento del usuario para controlar automáticamente la ubicación de la entrada de video, el sistema podrá servir como base

para el desarrollo de otros sistemas de mayor complejidad que incluyan reconocimiento de gestos.

1.3 Justificación

El reconocimiento del movimiento humano parece no ser posible de realizar debido a su gran complejidad, ya que se trata de un movimiento articulado y la forma del cuerpo puede variar drásticamente de un momento a otro. Por otra parte, es muy complejo obtener un modelo preciso del cuerpo humano, pues incluso si fuera representado como un simple conjunto de líneas conexas (la cual es la forma más sencilla), tendría más de veinte grados de libertad.

Sin embargo, los experimentos realizados por Johansson [5] en los años 70’s demostraron que el solo movimiento de los miembros del cuerpo humano contienen suficiente información para detectar su movimiento. Johansson filmó personas caminando en un cuarto de absoluta oscuridad usando como únicos identificadores pequeños puntos de luz blanca ubicados en cada miembro. El experimento mostró que un observador del video podía identificar fácilmente que se trataba de personas en movimiento, a pesar de la ausencia de pistas visuales como forma, color o texturas.

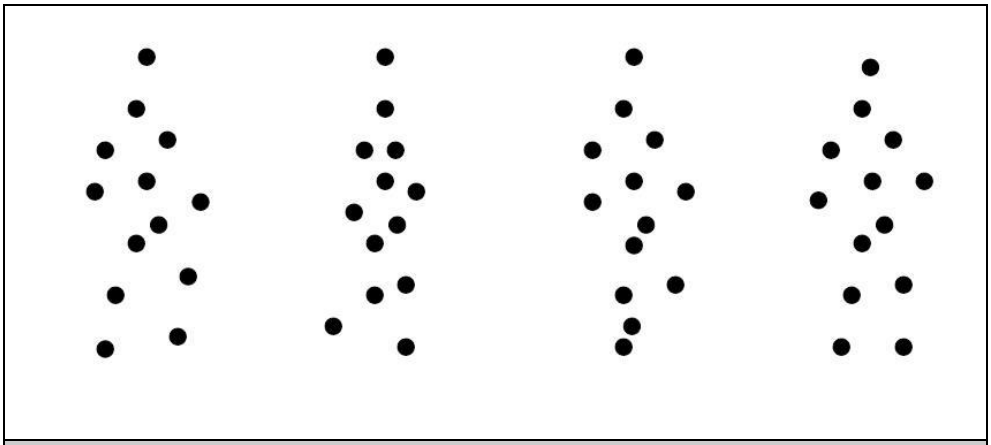

Figura 1.1. Ejemplo de la secuencia de imágenes del experimento de Johansson [F1]

Entonces el hecho de que para el cerebro humano resulte tan sencillo reconocer este movimiento sin la necesidad de características visuales y teniendo en cuenta de que se usaron simples puntos de referencia en cada miembro del cuerpo, nos permite pensar que el desarrollo de un algoritmo simple y efectivo para la detección del movimiento humano es muy posible de alcanzar.

Por otra parte, el sistema que se trata de construir necesita únicamente reconocer el movimiento de la persona que realiza su presentación. Esta persona en principio será considerada como un objeto más de la escena y una vez que se haya capturado la posición de la misma, el sistema realizará el seguimiento a la persona manteniendo rastro de ella en todo momento.

No obstante el sistema podría incrementar su complejidad si se efectúa una discriminación de objetos en la escena, para lo cual sería necesario contar con un algoritmo de reconocimiento del movimiento humano en una secuencia de video o alternativamente se podría utilizar un algoritmo de identificación de cuerpos humanos en imágenes estáticas o en secuencias de video mediante el uso de modelos y patrones de movimiento.

1.4 Planteamiento de la Solución

El problema principal que persigue la presente tesis es realizar una grabación de una presentación manteniendo siempre enfocado al expositor, para lo cual se efectuará el reconocimiento de la persona en la escena y posteriormente se realizará el respectivo seguimiento del movimiento de la misma. Todo esto deberá realizarse tomando en consideración que se contará únicamente con una fuente de video monocular, es decir, solamente una entrada de video.

Siguiendo este enfoque primero se debería reconocer una persona en una imagen, es decir analizar una imagen mediante procesamiento digital e identificar la posición de la persona, siempre y cuando exista una persona en la escena. Este reconocimiento podría ser realizado utilizando una imagen o varias imágenes

consecutivas provenientes de la secuencia de video en un momento determinado. Una vez reconocida la persona y obtenida su posición no quedaría más que rastrear continuamente esa región de la imagen, la cual obviamente tendrá movimiento en instantes determinados.

Es claro que este enfoque es el utilizado por el ojo humano. Cuando el ser humano desea vigilar a una persona, lo primero que hace es anclar su mirada en esa persona y no la retira en ningún momento; si la persona se mueve el observador también mueve su cabeza de tal forma de tener a su objetivo siempre en la mira. Este mismo enfoque es el que será utilizado por el sistema de grabación inteligente que se intenta desarrollar.

Sin embargo resultaría muy costoso, en cuanto a poder de procesamiento computacional, analizar píxel por píxel cada imagen de la secuencia de video para determinar si un punto dado pertenece o no a una persona, ya que esta región puede cambiar de forma y tamaño drásticamente, por la naturaleza del movimiento del ser humano y las articulaciones de su cuerpo. Además se necesitarían muchas características para establecer una relación de contenencia que permita identificar los puntos pertenecientes a la región en la que

se encuentra la persona. Adicionalmente, las características que se podrían utilizar (como color, forma y textura) también son sensibles a cambios bruscos de una imagen a otra.

Ahora bien, ¿cómo logra el ser humano identificar la presencia de un objeto en específico utilizando su sentido de la visión? Pues esto es posible gracias a que el ojo humano puede reconocer fácilmente el “fondo de la escena” y por tanto puede discriminar los objetos presentes en ella e incluso hacer una clasificación de los mismos basándose en su conocimiento previo.

Esto es precisamente lo que el sistema deberá realizar para poder reconocer a la persona que se encuentra en el escenario realizando su exposición; es decir, tendrá que reconocer en primera instancia el fondo de la escena para posteriormente comparar cada imagen consecutiva en contra del fondo obtenido. De esta forma se reduce la cantidad de información que debería ser procesada y por consiguiente el poder computacional necesario disminuye.

Una vez que se haya identificado la posición de la persona en la escena, esta deberá ser rastreada de tal forma de conocer en todo momento su posición actual. De esta manera cuando la persona se

aleje del rango establecido el sistema deberá hacer que la cámara gire automáticamente para volver a tener a la persona centrada en la toma.

Por último se deberá realizar la transferencia en tiempo real del video original registrado por la cámara, para lo cual se deberá tener en cuenta la compresión del video para transmitirlo por Internet.

1.5 Alcance

Como se podrá notar, el análisis del movimiento humano e incluso el reconocimiento de objetos en imágenes estáticas o en secuencias de video son temas muy complejos y que demandan de mucho estudio aún en estos días.

Es por esto que esta tesis se enfocará básicamente en el reconocimiento de objetos mas que en el análisis del movimiento humano, ya que el tema primordial en el sistema que será desarrollado es “mantener siempre en la mira” al expositor. El sistema no tomará en consideración los movimientos o gesticulaciones propias de un ser humano, ya que determinar sus acciones o su comportamiento no es de interés del mismo.

El sistema tendrá un funcionamiento adecuado en ambientes interiores como auditorios, salas de exposiciones, etc., en los cuales la intensidad lumínica no experimente cambios muy drásticos. Adicionalmente se deberá contar con una ubicación estable para montar la fuente de video, la cual en ningún momento y bajo ninguna circunstancia deberá experimentar movimiento no consentido por el sistema. Esta entrada de video será provista por una cámara de grabación de video digital rotatoria.

El expositor no deberá utilizar vestimenta que sea fácilmente confundible con el escenario en el cual se realicen las presentaciones, pues esto podría provocar dificultad en el reconocimiento de la misma y por consiguiente el sistema no funcionaría de la forma esperada.

Considerando que típicamente un expositor se mueve a lo largo del escenario, el sistema será capaz de realizar el seguimiento siempre que la persona no se aleje o se acerque a la fuente de video en demasía, ya que de no ser así la persona podría ser fácilmente confundible con la audiencia u otros objetos presentes entre el público. Por demasía se entiende un rango en el cual el sistema comienza a detectar un cambio en la forma y tamaño de la persona.

Por ejemplo, podrían presentarse destellos de luz entre el público, lo cual provocaría aberraciones en el reconocimiento y discriminación de los objetos en la escena. Pero aún así el sistema tendrá la capacidad de obviar pequeños cambios de intensidad lumínica y forma y tamaño de la persona (el objeto de interés) gracias a la fase de inicialización en la que se obtiene el fondo de la escena.

Por este motivo, la rotación de la fuente de video será controlada únicamente en el sentido horizontal mientras el sistema se encuentre realizando el seguimiento al expositor. Sin embargo, durante la fase de configuración del sistema y obviamente antes de la fase de inicialización, el usuario podrá manipular manualmente la fuente de video según su conveniencia y de esta forma lograr la mejor toma del escenario.

Como una característica adicional y para dar mayor utilidad al sistema, éste deberá ser capaz de transmitir el video en tiempo real mediante “streaming de video” en la Internet, para lo cual será necesario codificar el video utilizando un formato de compresión existente.

Con toda esta funcionalidad incorporada se pretende construir un sistema de gran utilidad que no solo servirá como una aplicación sino que se espera que sea una motivación para incrementar el interés por el campo de la Visión por Computadora por parte del estudiantado y por la comunidad científica local, y por consiguiente que cada vez se desarrollen sistemas de mayor complejidad.

1.6 Objetivos del Proyecto

Esta tesis persigue objetivos específicos establecidos, los mismos que se detallan a continuación:

- Analizar la factibilidad de llevar a cabo un sistema y estimar la precisión que puede llegar a tener. Además de establecer los supuestos y asumciones que se deben tomar en cuenta para el correcto funcionamiento del mismo.
- Realizar un breve análisis de las ventajas y desventajas de utilizar diferentes enfoques técnicos para la solución del problema, teniendo en cuenta la finalidad del sistema.
- Realizar un diseño modularizado para la solución del problema. Esto permitirá que el sistema pueda evolucionar por partes; es decir, se podrá mejorar un módulo en particular implementando mejores algoritmos y así obtener una mejor salida o producto

del módulo. Además los módulos podrán servir de base para el desarrollo de nuevos proyectos.

- Desarrollar un sistema configurable para diversos ambientes de interiores en los que se puedan llevar a cabo exposiciones, presentaciones y eventos similares en los que una persona efectúa una exposición.
- Obtener como resultado un producto de utilidad y del cual se pueda obtener provecho académico e incluso comercial. Con esto se pretende motivar al estudiantado para que tomen mayor interés en el área de la Visión por Computador e incentivar al desarrollo de proyectos más complejos.
- Realizar un análisis del futuro del sistema, estableciendo áreas que pudieran ser mejoradas para obtener un mejor producto.

1.7 Sumario

En este capítulo se ha presentado una descripción detallada del problema que se intenta resolver en este trabajo. Asimismo se ha planteado una solución general para este problema, sin tener en consideración los detalles de su implementación.

También se ha realizado un breve análisis de la funcionalidad básica que el sistema deberá cumplir y las consideraciones que deberán estar presentes durante su implementación.

En el siguiente capítulo se realizará un estudio de la viabilidad de la implementación de este sistema, así como de las técnicas existentes que deberán ser consideradas durante el desarrollo del mismo. Este estudio se presenta de una forma categorizada y adecuadamente estructurada, de tal manera que sea de fácil comprensión para el lector.

Adicionalmente se presentarán algunas herramientas disponibles para el desarrollo de sistemas de procesamiento digital de imágenes y de visión por computador, así como sistemas reales de aplicación comercial e investigativa.

CAPÍTULO 2

2 MARCO DE REFERENCIA

2.1 Factibilidad de la Implementación

Hasta ahora se ha establecido que el problema principal que deberá resolver el sistema es reconocer al expositor en la escena y detectar su posición y forma para posteriormente efectuar el rastreo de su movimiento. El interés por este rastreo (el seguimiento del movimiento de la persona) puede también limitarse a detectar objetos en las secuencias de imágenes.

Existen dos categorías de técnicas utilizadas para el rastreo del movimiento humano: Basados en Marcadores (Intrusiva) y Sin Marcadores (No Intrusiva).

Las técnicas basadas en Marcadores operan mediante la utilización de dispositivos montados en el individuo y en su entorno, el mismo que transmite o recibe la información generada. Estas técnicas intrusivas permiten un procesamiento más sencillo y son muy

utilizadas cuando las aplicaciones disponen de entornos bien controlados. En cambio, las técnicas no intrusivas se basan en fuentes naturales de información (como una fuente de video estática) y no requieren de dispositivos portables.

En aplicaciones de vigilancia, monitoreo de áreas extensas y seguimiento en tiempo real, se utilizan técnicas no intrusivas basadas en la localización de objetos en movimiento, forma del cuerpo y rastreo del contorno del cuerpo. Los objetos en movimiento pueden ser identificados en las imágenes mediante “sustracción de fondo” (background subtraction) o “flujo óptico” (optical flow). Por otra parte, si la fuente de video no es estática y presenta movimiento, se debe realizar una rectificación de los frames que compense el conocimiento actual del fondo. Además los problemas de oclusión por la aparente superposición de objetos se pueden resolver mediante análisis temporal y predicciones de trayectoria.

2.1.1 Problemas Típicos

Los sistemas de visión por computador, dependiendo de la aplicación a la cual estén enfocados, tienen que afrontar diversos problemas que dificultan su funcionamiento o limitan su campo de acción de una u otra forma.

A continuación se presentan los problemas más relevantes que influyen en la detección de movimiento y el reconocimiento de objetos:

- Uno de los principales problemas en la detección de movimiento es la presencia de hojas de árbol en el fondo de la escena, ya que producen variaciones oscilatorias por su movimiento a causa del viento. Algo similar ocurre cuando un objeto que inicialmente estaba en el fondo se mueve, pues esto provoca que tanto ese objeto y la parte del fondo que ocupaba, aparezcan como cambios en la escena y por consiguiente se producen falsos movimientos.
- Los cambios graduales de iluminación en el ambiente y los cambios repentinos de la iluminación interna (de un salón de exposiciones por ejemplo) alteran la apariencia del fondo de la escena.
- Usualmente los objetos proyectan sombras que hacen que el fondo aparezca diferente.
- Un objeto puede confundirse con el fondo de la escena si los píxeles del objeto tienen características similares al fondo, lo cual provoca que el objeto quede camuflado y que no sea detectado como objeto móvil.

- La inicialización del fondo puede no ser factible ya en que algunos entornos o circunstancias no es posible disponer de un periodo de tiempo para una fase inicialización en ausencia de los objetos de interés en la escena.
- La superposición de objetos en movimiento puede causar interpretaciones erróneas del movimiento de los objetos implicados ya que se podría incurrir por ejemplo en una confusión a causa de esta superposición y considerarla como variaciones de forma de un objeto y la desaparición del otro. Esto es lo que se conoce como oclusión.

2.1.1.1 Factores que Influyen en la Precisión

Existen también muchos factores que influyen en el grado de exactitud que un sistema de visión por computador puede llegar a tener. A continuación se detallan los factores más importantes y que deben ser considerados por este trabajo de tesis.

La calidad del video que pueda brindar la fuente es de extrema trascendencia, ya que de ella depende el grado de pre-procesamiento al que se debe someter cada imagen de la secuencia, antes de comenzar con el verdadero análisis y procesamiento de la misma.

La iluminación del lugar es también un tema relevante en el proceso de segmentación de las imágenes para obtener los objetos de la escena. Una iluminación adecuada permitirá un mejor contraste de los objetos, lo cual permitirá reconocer más fácilmente cada región de la imagen segmentada. De lo contrario, la segmentación podría dar como resultado que varios objetos sean considerados como uno solo.

Por otra parte las variaciones abruptas de iluminación provocan una diferencia de intensidad muy notoria entre imágenes consecutivas y por consiguiente la extracción de los objetos no sería muy precisa, puesto que la sustracción del fondo mostraría objetos (en movimiento) no existentes.

La ubicación de la fuente de video en el momento de la inicialización para obtener el fondo es relevante, pues del enfoque que tenga la fuente depende el fondo que se obtendrá. En pocas palabras, la fuente de video deberá ser posicionada manualmente en principio a criterio de un ser humano, para así obtener el mejor enfoque de la escena.

2.1.1.2 Limitaciones y Restricciones

Las técnicas y algoritmos desarrollados hasta el momento requieren que ciertas restricciones se satisfagan para tener un funcionamiento adecuado. Incluso cumpliendo con estas restricciones, el rendimiento y eficacia del algoritmo pueden no ser los esperados debido a situaciones imprevistas y a factores externos. Actualmente, no existe una solución general para el problema del seguimiento al ser humano y esta lejos de conseguirse. Las restricciones que deberán considerarse para un sistema en particular dependerán de su aplicación, y mientras menos restricciones sean necesarias para un sistema, su grado de complejidad será mayor.

Las restricciones utilizadas con mayor frecuencia se dividen en dos categorías básicas: restricciones de movimiento y restricciones de apariencia, según [1]. La primera categoría de restricciones se refiere a suposiciones respecto del movimiento del sujeto o de la cámara. En cambio, la segunda comprende suposiciones en cuanto a la apariencia del entorno y del sujeto.

- **Restricciones relacionadas con el movimiento**

  En esta categoría se distinguen diez restricciones. Las primeras tres, “el sujeto se mantiene dentro del campo de

visión", "cámara estática o con movimiento constante" y "solo un sujeto en el campo de visión" son usadas en la mayoría de los sistemas de seguimiento existentes. La siguiente, "el sujeto siempre mira a la cámara" es usada en interfaces hombre-máquina y simplifica el cálculo de la postura global del cuerpo. Una restricción que reduce la dimensionalidad del problema de 3D a 2D es "movimientos paralelos al plano de la cámara" y es usada muchas veces en el análisis de la mirada. La siguiente restricción "el sujeto no queda oculto" simplifica el seguimiento del cuerpo humano ya que permanece completamente visible. Otra restricción es "movimientos lentos y contínuos", la cual permite un cálculo de una trayectoria simple. La octava restricción, "movimiento de uno o pocos miembros", permite enfocarse en solo una parte del cuerpo. La siguiente restricción, "se conoce el patrón de movimiento del sujeto", es usada para simplificar el seguimiento y los problemas de la estimación de postura, puesto que reduce el número de posibles soluciones. La última restricción de movimiento, "el sujeto se mueve en un suelo plano", es utilizada para permitir el cálculo de la distancia entre el sujeto y la cámara usando el tamaño del sujeto y la geometría de la cámara.

- **Restricciones relacionadas con la apariencia del entorno**

  La restricción de “iluminación constante” básicamente insinúa que el ambiente debe ser de interiores. La siguiente suposición de esta categoría, “fondo constante”, permite que sea posible la segmentación del sujeto basado en la información de movimiento. La tercera restricción “fondo uniforme” es muy usada por los sistemas existentes porque permite utilizar un simple umbral para realizar la segmentación del sujeto. La cuarta restricción “parámetros de cámara conocidos” es necesaria para obtener medidas absolutas de posición de los objetos. La última restricción de esta categoría es “uso de hardware especializado”, como cámaras IR o múltiples cámaras.

- **Restricciones relacionadas con la apariencia del individuo**

  La primera restricción es “postura inicial conocida” y es muy utilizada en los sistemas actuales para simplificar la fase de inicialización. La siguiente restricción concierne con el “conocimiento previo de características del sujeto”, como estatura, ancho, longitud de los miembros, etc. Las siguientes tres restricciones son: “uso de marcadores de referencia montados en el cuerpo”, “uso de ropa de color especifico” y

"uso de ropa ajustada", reducen los problemas de segmentación al simplificar la detección del sujeto.

RESTRICCIONES	
Relacionadas al movimiento	Relacionadas a la apariencia
1. el sujeto se mantiene dentro del campo de visión 2. cámara estática o con movimiento constante 3. solo un sujeto en el campo de visión 4. el sujeto siempre mira a la cámara 5. movimientos paralelos al plano de la cámara 6. el sujeto no queda oculto 7. movimientos lentos y continuos 8. movimiento de uno o pocos miembros 9. se conoce el patrón de movimiento del sujeto 10. el sujeto se mueve en un suelo plano	1. iluminación constante 2. fondo constante 3. fondo uniforme 4. parámetros de cámara conocidos 5. uso de hardware especializado 1. postura inicial conocida 2. sujeto conocido 3. uso de marcadores de referencia 4. uso de ropa de color especifico 5. uso de ropa ajustada

Tabla 2.1: Restricciones [1] utilizadas por sistemas HMT
listadas por frecuencia de uso

Según el número de restricciones necesarias para el funcionamiento de un sistema de seguimiento, se puede notar que este campo de investigación esta todavía en una etapa de desarrollo y estudio, y por consiguiente aún esta lejana una solución maestra para este problema.

2.2 Adquisición de Video

La secuencia de video y los parámetros de inicialización son prácticamente la entrada de todo sistema de detección y seguimiento de objetos y por consiguiente, del movimiento humano. De hecho, el tipo de entrada que se utilice es una de las diferencias entre varias técnicas y enfoques relacionados con este tema. Incluso existen

restricciones (véase 2.1.1.2) que rigen el número y tipo de cámaras utilizadas por un sistema.

De esta forma, dependiendo de las técnicas que un sistema incorpore, pueden ser utilizadas fuentes de video monocular, estéreo visión y múltiples cámaras. Ciertas técnicas requieren no solo un número especifico de cámaras de video sino que necesitan de un tipo especial de dispositivos como las cámaras infrarrojas y las cámaras de visión nocturna.

2.2.1 Importancia de la Calidad del Video

Las secuencias de video para un sistema HMT son análogas a la materia prima para la fabricación de un producto. Básicamente la diferencia entre los sistemas radica en el procesamiento que estos apliquen a las secuencias de video, incluyendo pre-procesado y post-procesado. Dependiendo de la aplicación del sistema, es probable que se requiera aplicar un pre-procesamiento al video para facilitar la ejecución de los algoritmos. Por ejemplo si el video no tiene buen contraste, es necesario incorporar un filtro especializado que mejore el video para que los algoritmos y técnicas sean aplicables. Por consiguiente, si la calidad del video es pobre se incrementará el tiempo y poder de procesamiento necesario. Sin embargo, el término

“calidad” no es muy preciso, pues para un sistema puede ser suficiente el video capturado por una cámara web, las cuales usualmente generan ruido, mientras que otros requieren de cámaras profesionales de alta definición, en las que la nitidez de la señal es una característica. Es por este motivo que los sistemas requieren que se satisfagan ciertas suposiciones (véase 2.1.1.2) para establecer las condiciones en que el video debe ser obtenido.

2.2.2 Tipos de Fuentes de Video

El tipo de fuente más adecuado para un sistema de detección y seguimiento de objetos (no solo de personas), depende de la aplicación a la cual el sistema esté dirigido. En ocasiones es suficiente utilizar una fuente monocular para obtener la funcionalidad deseada e incluso muchas veces resulta más eficiente, pues el enfoque de múltiples cámaras añade complejidad y procesamiento adicional innecesario. Por ejemplo en un sistema de alerta de vehículos cercanos se puede utilizar solo una cámara de video instalada en la parte frontal del vehículo y aplicar geometría para estimar la cercanía de los vehículos, como en [15]. Sin embargo, usar fuentes monoculares no es sinónimo de sencillez, pues en determinadas condiciones resultaría más efectivo utilizar múltiples cámaras. Además si la fuente monocular no es estacionaria se

necesita aplicar un algoritmo de compensación del movimiento de la misma.

Una limitación de las fuentes monoculares es que el espacio de trabajo sobre el cual un sistema tiene dominio se reduce al campo de visión de la cámara. Esto es parcialmente solucionable permitiendo el movimiento traslacional o rotatorio de la cámara, pero se incrementa el procesamiento necesario para obtener información de la escena. Este procesamiento adicional podría aprovecharse mejor utilizando un enfoque de múltiples cámaras en el que todas contribuyen en conjunto para extraer la información de la escena 3D. Por ejemplo, un sistema de seguimiento de personas puede utilizar un enfoque distribuido en el que participan redes de cámaras interconectadas y que obtienen información proveniente de varios sensores de video. De esta forma se podría implementar un ambiente pervasivo en el cual las necesidades del usuario (una persona) sean interpretadas mediante el reconocimiento de gestos e intenciones, reconocimiento de posturas del cuerpo humano y el uso de patrones de movimiento. En [17] se introduce un sistema similar, que utiliza el enfoque distribuido de múltiples fuentes de video.

2.3 Reconocimiento de Objetos

El reconocimiento de objetos es uno de los principales problemas que un sistema de seguimiento debe resolver. Este reconocimiento consiste básicamente en detectar los objetos presentes en la escena, para lo cual se separa el fondo (“background”) del primer plano (“foreground”). Usualmente el número de objetos presentes en una escena es considerablemente alto. No obstante, por lo general para un sistema de seguimiento los “objetos de interés” son aquellos que presentan movimiento. Considerando esto, el fondo de la escena se define como la parte estacionaria de la misma, mientras que los objetos en contínuo movimiento o que permanecen estáticos durante un lapso de tiempo luego del cual reanudan su movimiento, forman parte del primer plano. Sin embargo, un objeto estacionario, es decir estático temporalmente, podría detener su movimiento definitivamente y convertirse en estático, por lo cual debería ser considerado como parte del fondo y no del primer plano. Entonces, surge la necesidad de establecer un límite de tiempo durante el cual un objeto estacionario podrá continuar su movimiento, mientras que una vez excedido el límite establecido dicho objeto deberá considerarse como objeto estático y pasar a formar parte del fondo.

Independientemente del contexto, muchos de los algoritmos de seguimiento comienzan con una segmentación Fondo-Objetos seguida de una transformación de las imágenes segmentadas para reducir la cantidad de información a ser procesada. A continuación se estudian las técnicas existentes para realizar dicha segmentación.

2.3.1 Segmentación Fondo-Objetos

Existen varias técnicas para separar los objetos de interés del fondo de la escena, las cuales pueden utilizar información espacial o temporal. Las técnicas que utilizan información temporal usualmente analizan frames consecutivos tomados de la secuencia de video, mientras que las técnicas basadas en información espacial pueden basarse en enfoques probabilísticos.

2.3.1.1 Técnicas basadas en información temporal

Las técnicas que utilizan información temporal en su mayoría asumen un fondo estático y una cámara que no presente movimiento alguno. La idea básica es que las diferencias entre frames consecutivos de la secuencia de video pueden identificar los objetos en movimiento. Estas técnicas son “Sustracción de Fondo” y “Flujo Óptico”.

- **Sustracción de Fondo (Background Subtraction)**

  En su forma más simple la sustracción de fondo asume que la intensidad de cada píxel perteneciente al fondo varia independientemente siguiendo una distribución normal. La idea básica es obtener la diferencia entre el frame actual y el inmediato anterior, la misma que se realiza píxel a píxel utilizando los valores de intensidad o gradientes. Una versión mejorada de la sustracción de fondo utiliza tres frames consecutivos para obtener la diferencia.

  Es obvio que la diferencia (o resta) entre dos frames consecutivos refleja las variaciones de uno a otro frame, lo cual es interpretado como movimiento, siempre y cuando los objetos en movimiento no tengan la misma intensidad o color que el propio fondo de la escena. Sin embargo, esta diferencia de imágenes puede generar ruido, el mismo que puede ser removido mediante la aplicación de algún filtro especializado.

  Comúnmente esta técnica obtiene una imagen de la escena sin los objetos de interés, la cual es usada como referencia durante la sustracción. Sin embargo esto puede mejorarse permitiendo que el fondo se adapte a leves cambios en la

iluminación y en la escena, como en el caso de que un objeto pase por detrás de otro o que la cámara se mueva.

- **Flujo Óptico (Optical Flow)**

  El flujo óptico se define como un movimiento aparente en la intensidad de una imagen y describe básicamente un movimiento coherente entre puntos o características entre frames consecutivos de una secuencia. Esta técnica asume dos enunciados:

  - la intensidad de un píxel depende de sus coordenadas espaciales en la imagen, y
  - la intensidad de cada punto de un objeto estático o en movimiento no varía en el tiempo.

  En base a estos dos enunciados y utilizando series de Taylor se obtiene la ecuación de restricción del flujo óptico, la misma que tiene más de una solución posible. Por este motivo son necesarias más restricciones para obtener una solución única. Entonces, dependiendo de las restricciones que se tomen para la ecuación de flujo óptico, se obtienen diferentes enfoques para la solución de la ecuación. Ejemplos de esto

son los enfoques de: Lucas & Kanade, Horn & Schunck y Matching por bloques.

- Lucas & Kanade

  Esta técnica [20] reduce el cálculo del flujo óptico a la resolución de un sistema de ecuaciones lineales. Utiliza la ecuación de flujo óptico para un grupo de píxeles adyacentes y asume que todos estos tienen la misma velocidad. El sistema lineal tiene solución única en el caso de que sea un sistema no singular para dos píxeles. Sin embargo, combinar las ecuaciones para más de dos píxeles es más efectivo. Adicionalmente esta técnica emplea una ventana gaussiana que puede ser representada como la composición de dos kernels separables con coeficientes binomiales.

- Horn & Schunck

  Esta técnica [14] añade un regulador global de suavidad (en el flujo óptico) al modelo clásico de la restricción de flujo óptico, lo que resulta en la minimización de una función global de energía. Esta minimización se realiza empleando cálculo de varias

variables. La restricción es formulada mediante una integral doble calculada en la región de la imagen.

- Matching por bloques

  A diferencia de las anteriores, esta técnica no usa directamente la ecuación de flujo óptico. Básicamente consiste en encontrar similitudes entre bloques de cada imagen. Considérese una imagen divida en pequeños bloques que pueden superponerse. Para cada uno de los bloques en la primera imagen, se trata de encontrar un bloque del mismo tamaño en la segunda imagen y que sea el más similar al bloque de la primera imagen. Se asume que todos los píxeles pertenecientes a un mismo bloque tienen el mismo desplazamiento de una imagen a otra, es decir la misma velocidad. Pueden utilizarse diferentes métricas para medir la similitud o diferencia entre los bloques, tales como: diferencia de cuadrados y correlación de cruce.

2.3.1.2 Técnicas basadas en información espacial

El uso de información espacial conlleva a dos subclases de técnicas, las cuales son: Umbralización y Enfoques Estadísticos. La primera

técnica se basa en suposiciones o restricciones especiales relacionadas con el entorno de la escena, mientras que los enfoques estadísticos requieren menos restricciones que los métodos de sustracción.

- **Umbralización**

  Las técnicas de Umbralización tienen como base la diferencia de color o intensidad del sujeto en relación al resto de la escena.

  Una técnica muy usada es “chroma-keying”, en la cual el sujeto aparece en frente de un fondo de color uniforme, usualmente azul o verde, utilizando vestimentas que no contengan el color del fondo. Utilizando una simple umbralización el sujeto puede ser fácilmente separado del fondo de la escena. Una variación de esta técnica, y muy utilizada también, consiste en que el sujeto aparezca con vestimentas de color uniforme, usualmente oscuro, en frente de un fondo. Alternativamente se pueden utilizar marcadores de referencia (activos o pasivos) montados en el sujeto los cuales permiten segmentar la imagen de manera muy simple. Si se utiliza hardware especial se introducen ventajas adicionales, pues por ejemplo se pueden obtener imágenes

termales, en las cuales el sujeto sea fácilmente distinguible debido a su color característico relacionado con el calor del cuerpo humano, y de esta manera el sujeto podría ser separado del fondo de la escena mediante umbralización.

- **Enfoques Estadísticos**

  Estos enfoques utilizan características, usualmente colores y bordes, de píxeles individuales o de grupos de píxeles para extraer un objeto de la escena. Algunos de estos enfoques toman como referencia la técnica de “sustracción de fondo”, pues utilizan una secuencia de imágenes del fondo de la escena y calculan la media y la varianza de la intensidad de cada píxel en el transcurso del tiempo. Cada píxel de la imagen actual es comparado con las estadísticas de la imagen de fondo y es clasificado como perteneciente al fondo o al primer plano. Este enfoque es muy robusto en comparación con las técnicas de sustracción ya que permite, por ejemplo, remover las sombras de un objeto mediante la utilización de estadísticas de los gradientes de cada píxel.

  Una variación más avanzada de este enfoque se basa en el uso de regiones del objeto llamadas “blobs”. En este enfoque

el objeto es modelado por un número de blobs que poseen su propio color y estadísticas espaciales. De esta manera cada píxel de la imagen actual es clasificado como perteneciente a uno de esos blobs considerando su color y sus propiedades espaciales.

Otro enfoque estadístico consiste en utilizar contornos estáticos o dinámicos, en el que el contorno activo (del objeto) es dinámico, mientras que lo estático se refiere al uso de estructuras estáticas predefinidas que representan una parte del contorno del objeto. Estas estructuras consisten de segmentos de bordes y otros atributos. Puesto que los contornos activos pueden cambiar su forma en el tiempo, su deformación es controlada por funciones externas de energía, las cuales acomodan las curvas del contorno a características de la imagen como los bordes, y por funciones internas, las mismas que ajustan la suavidad de la curva. Entonces, mediante el uso de contornos activos es posible extraer el perfil completo de un objeto de la escena o extraer partes del perfil.

2.3.2 Análisis y Clasificación de Objetos

La segmentación de imágenes es un tema trascendente para muchos sistemas de visión por computador, pues es el primer paso en el proceso de entendimiento de la imagen e influye de gran manera en el proceso de interpretación de la misma.

Según muchos estudios, como las leyes de la Gestalt, los seres humanos tienen preferencia por agrupar regiones visuales basándose especialmente en proximidad, similitud y continuidad, para generar un conjunto de regiones significativas de la imagen, es decir los objetos. La segmentación de imágenes tiene su base en este principio y básicamente tiene por objetivo definir e identificar regiones de una imagen, que tengan algo en común, como color, textura, movimiento, etc.

Las técnicas de clasificación se pueden agrupar según el nivel de abstracción al que pertenezcan. Sin embargo esta clasificación es principalmente conceptual, pues muchas técnicas podrían corresponder a más de un nivel.

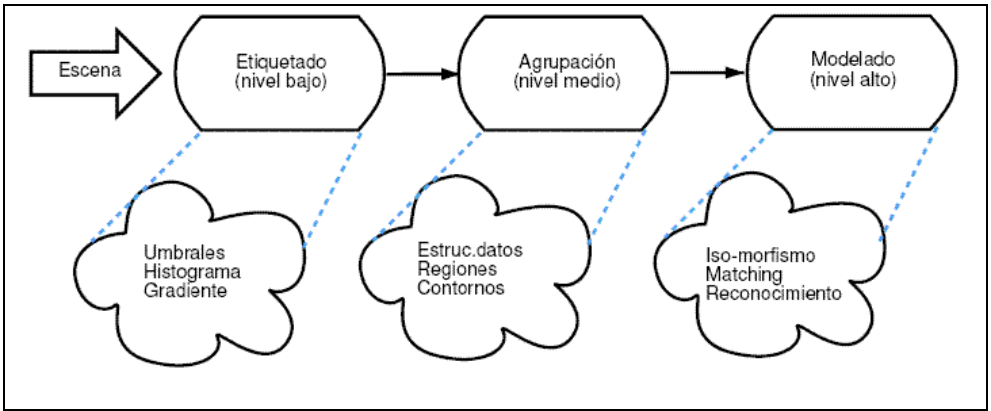

Figura 2.1. Niveles de clasificación [F2]

2.3.2.1 Técnicas de Bajo Nivel

Estas técnicas se encargan de asociar puntos de la imagen basándose en la similitud. Las más comunes utilizan la intensidad como medida de similitud.

- **Métodos orientados a píxel**

  Estos métodos asocian cada píxel de la imagen a una determinada clase de similitud según el nivel de gris que posea. Para esto es necesario determinar los rangos de intensidad que corresponden a cada clase de similitud. Entonces, una vez que se han obtenido los rangos, cada píxel de la imagen puede asignarse a uno de dichos rangos mediante umbralización multinivel.

Puesto que estos métodos utilizan una simple umbralización, son sencillos de implementar y generan resultados rápidos. Sin embargo, no consideran los conceptos de conectividad y proximidad de los píxeles en la imagen.

- **Métodos orientados a región**

  Los métodos orientados a región utilizan la noción de conectividad para agrupar en distintas entidades, zonas de la imagen cuyos puntos poseen intensidad similar.

- **Métodos orientados a contorno:**

  Estos métodos explotan las discontinuidades en intensidad de una imagen y la frontera de los objetos contenidos en ella. Se basan en una supuesta correspondencia de estas discontinuidades, lo cual muchas veces puede causar situaciones de incongruencia o ambigüedad. Esto último es producto de la naturaleza no biunívoca de la supuesta correspondencia, puesto que las fronteras de los objetos no solo están determinadas por gradientes de intensidad sino también por muchos otros cambios de intensidad que corresponden a variaciones causadas por sombras, efectos de iluminación, color de superficie no uniforme o por la textura del

objeto. Considerando esto, aunque los contornos proveen buena información para la identificación de los objetos es muy probable que sea necesario un procesamiento posterior.

2.3.2.2 Técnicas de Nivel Medio

En estas técnicas los píxeles de cada entidad obtenida son asignados a una estructura propia de la entidad, la cual de alguna manera describe su contenido. Si embargo, las entidades extraídas deben cumplir con ciertas condiciones como por ejemplo, que los contornos sean contínuos y conexos y que las regiones sean cerradas y también conexas.

Una clasificación usual para estas técnicas es:

- de pre-procesado de píxeles asignados a clases
- de agrupación de píxeles a entidades
- de post-procesado de entidades

2.3.2.3 Técnicas de Alto Nivel

Las técnicas pertenecientes a este grupo tienen que ver con el modelado de la escena. El objetivo de un sistema de percepción

visual es siempre la interpretación de la escena y es esto mismo lo que la clasificación de alto nivel persigue.

En estas técnicas, las entidades correspondientes a objetos son reconocidas tomando como referencia una base de modelos y son categorizadas semánticamente mediante conceptos complejos como la pertenencia y la oclusión.

2.3.3 Representación de Objetos

Una vez que los objetos se han obtenido en forma de entidades segmentadas, sigue la representación de estos objetos. En principio existen dos tipos de representaciones: Basada en Objeto y Basada en Imagen. La primera se basa en la segmentación de la escena, es decir la separación de los objetos del fondo de la escena, mientras que la segunda se deriva directamente de la imagen.

2.3.3.1 Representación Basada en Objeto

L Esta representación depende principalmente de la separación de los objetos de interés del fondo de la escena (segmentación). Según el trabajo realizado por Moeslund y Granum en [1], existen básicamente cuatro tipos de representación basada en objeto: por puntos, por cajas, por silueta y por regiones denominas blobs.

La representación por puntos es suficiente en un ambiente que utilice marcadores activos o pasivos. Los marcadores activos proveen un gran contraste en las imágenes, por lo cual la representación es robusta. Incluso se podría generar una representación 3D si se utiliza más de una cámara en un sistema basado en marcadores.

La representación por cajas trata de representar al sujeto utilizando un conjunto de cajas limítrofes, las cuales contienen los píxeles de cada región encontrada mediante el proceso de segmentación. Estas cajas pueden ser rastreadas en el tiempo para realizar el seguimiento del objeto.

La representación por silueta es usada tanto en 2D como en 3D, y puede obtenerse mediante métodos de sustracción o métodos de umbralización. La representación 2D generalmente es sencilla pero puede ser más compleja al emplear B-splines uniformes cerradas con un número fijo de puntos de control separados en una distancia equitativa alrededor de la silueta. En cambio, la silueta 3D puede ser obtenida mediante una combinación de siluetas 2D o utilizando un enfoque de estero-visión.

En la representación por blobs el sujeto se representa por un blob o por un conjunto de blobs con características similares, como colores. La idea de agrupar la información basándose en similitudes proviene de la investigación acerca del sistema de visión humano realizado por Gestalt.

2.3.3.2 Representación Basada en Imagen

La representación basada en imagen utiliza directamente los píxeles de la imagen. Estas representaciones pueden derivarse de una imagen independiente de la presencia de objetos en la escena o del interior de un objeto representado con uno de los enfoques de representación por objeto. Estas imágenes pueden ser transformadas a otro espacio que tenga una base de funciones no cartesianas, dando como resultado una representación más compacta. Entre las transformaciones más utilizadas están: Fourier, PCA, DCT y wavelets.

Si se considera información temporal para la representación, es posible incluir características relacionadas con el movimiento.

Otra forma de representación por imagen utiliza bordes, los cuales pueden representarse usando puntos o segmentos de línea.

La representación basada en imagen puede utilizar también características como área y color para representar las partes individuales del cuerpo de un ser humano. Según el enfoque de Christensen y Corneliussen [18], las partes del cuerpo humano son obtenidas mediante umbralización de una imagen en la que el sujeto utiliza vestimenta de colores especiales, y utilizan la longitud, el área y el color para representar cada una de las partes del cuerpo.

2.3.4 Eliminación de Objetos

Dependiendo de la aplicación a la que se enfoque un sistema de visión, en la escena pueden encontrarse objetos (reales o no) que no son de interés alguno para el sistema, aún cuando estos objetos presenten movimiento. Este es el caso de objetos demasiado pequeños, demasiado grandes e incluso sombras.

Estos objetos sombra aparecen en el primer plano sin tener una correspondencia con un objeto real de la escena, sino que son producto de la proyección de un objeto real en el fondo de la escena.

La eliminación de objetos usualmente se puede realizar por tres métodos distintos. Estos son:

- **Eliminación por área**

  Este es un proceso sencillo que utiliza un límite máximo y un límite mínimo para el área que un objeto puede tener. Si el área de un objeto, la cual es obtenida en base al número de puntos en la región que corresponde al objeto, se encuentra fuera del rango definido por estos límites el objeto simplemente se descarta.

- **Eliminación de sombras por intensidad uniforme**

  Este método se basa en la uniformidad de la intensidad en los objetos, pues supone que los objetos sombra tienen una intensidad uniforme en toda su área. Tomando como base esta premisa, los objetos que cumplan con esta uniformidad son considerados sombras y por consiguiente son descartados.

- **Eliminación de sombras por coincidencia de bordes**

  Este método también parte de una hipótesis, la cual consiste en que en un par de imágenes, una con sombras y una sin ellas, los contornos no varían puesto que una sombra se puede considerar como un objeto sin contornos.

Para la detección de los bordes o contornos se utilizan medidas de gradiente, pues los operadores basados en gradiente tienen un funcionamiento relativamente bueno cuando una imagen presenta variaciones abruptas de intensidad.

El proceso consiste en calcular los gradientes de la imagen original y del fondo actual. Las regiones correspondientes a cada objeto son comparadas en las imágenes gradiente y si estas regiones son iguales o coinciden en los bordes el objeto es considerado como sombra.

2.3.5 Análisis de los Enfoques más Adecuados

Como se ha revisado, existen muchas técnicas que pueden ser utilizadas para el reconocimiento de objetos y de personas en imágenes estáticas y en secuencias de video. El uso de una u otra técnica depende en primera instancia de las condiciones del entorno en el que se implemente un sistema.

Para el caso del presente trabajo, no es posible utilizar la técnica de Sustracción de Fondo como base para el reconocimiento del expositor, debido a que el sistema depende del movimiento

rotacional de la fuente de video y por lo tanto el fondo de la escena no es estático. Tampoco es posible utilizar la Umbralización propiamente dicha, puesto que no se puede obligar a un expositor a vestirse de determinada forma ni tampoco exigir que utilice algún traje o vestimenta de fácil identificación.

Sin embargo, es posible utilizar una combinación de técnicas que utiliza la Sustracción de Fondo en conjunto con una compensación del movimiento de la cámara. De esta manera, es factible tener un fondo dinámico que sea actualizado durante el tiempo de ejecución utilizando los parámetros del movimiento de la cámara.

Por tanto, la fase de reconocimiento del sistema consistirá en detectar la posición inicial del expositor, definiendo una “región de interés” (ROI), para lo cual es posible utilizar una técnica de sustracción de fondo. Una vez que el fondo empiece a cambiar debido al movimiento del expositor, será necesario estimar el desplazamiento (posición, tamaño y orientación) de la región de interés.

2.4 Seguimiento de Objetos

El seguimiento de un objeto en el tiempo básicamente consiste en encontrar sus correspondencias en imágenes consecutivas. Sin embargo, la dificultad del seguimiento radica en la complejidad tanto de la escena como de los objetos que son rastreados. La complejidad de los objetos se relaciona a los grados de libertad que posea y a su representación. Mientras más puntos de un objeto son rastreados aumenta la cantidad de procesamiento, pues es equivalente a rastrear múltiples objetos simultáneamente. Aun más, un objeto o sus puntos pueden dividirse y mezclarse con otros objetos a causa de la oclusión y por el ruido de la imagen. Incluso la apariencia de un objeto puede cambiar debido a los cambios en la iluminación y por la presencia de sombras.

El análisis de correspondencia entre los frames consecutivos usualmente es realizado mediante predicción. Esta se basa en objetos detectados previamente y en el conocimiento previo de alto nivel para predecir el estado de los objetos (como apariencia, posición, etc.) en el próximo frame y compararlos con aquellos de la imagen actual. La predicción de estos parámetros se basa en un modelo de cómo evolucionan en el tiempo, como el modelo de velocidad y aceleración o modelos más avanzados como el caminar.

Un problema adicional en el seguimiento surge cuando se utilizan múltiples cámaras, ya que existe la necesidad de decidir qué cámaras o qué imágenes utilizar en cada instante.

2.4.1 Objetos en Movimiento

Para que el seguimiento de un objeto sea robusto es necesario rastrear los blobs en movimiento. El seguimiento de objetos es necesario para el análisis y reconocimiento del comportamiento humano basado en video. Básicamente, el seguimiento de los blobs en una secuencia de video consiste en asociar cada objeto en movimiento detectado a su correspondiente objeto entre frames consecutivos de la secuencia, utilizando características de los blobs. De hecho, como ya se ha mencionado anteriormente, algunos métodos de seguimiento pueden predecir la ubicación del blob en movimiento en el próximo frame.

Los algoritmos de seguimiento estándares se basan en métodos estadísticos de correlación. Algunos de estos son: el Filtro de Kalman, el Algoritmo de Condensación, Seguimiento de Media Deslizante y Redes Bayesianas Dinámicas.

El Filtro de Kalman es capaz de estimar incertidumbre de predicción, la cual puede ser utilizada para determinar regiones de interés, para disminuir la necesidad de procesar una imagen completa. Sin embargo, este es útil en situaciones en las que se tiene una distribución unímodal de probabilidad de los parámetros de estado y por tanto no es posible utilizarlo en presencia de oclusión, fondos que tienen un gran parecido con los objetos de interés, ni en presencia de movimientos muy complejos, pues estas situaciones generalmente tienen una probabilidad multimodal. Los métodos basados en este filtro son desarrollados para problemas de seguimiento con radar y trabajan muy bien rastreando objetivos puntuales.

El Algoritmo de Condensación se basa en la propagación de la densidad condicional en los frames consecutivos de la secuencia de video. Este algoritmo estima una distribución posterior para el frame previo y se extiende sobre los siguientes frames de manera iterativa. No obstante, por ser un método no paramétrico requiere un número de muestras relativamente alto para asegurar una buena estimación de probabilidad del estado actual.

El Algoritmo de Seguimiento de Media Deslizante determina la ubicación del objeto en el próximo frame mediante un proceso

iterativo. Se basa en histogramas de color normalizados y afinados de los objetos en movimiento. Una vez que el histograma que representa al objeto en movimiento es construido manual o automáticamente a partir de un frame inicial, n, en el cual el objeto aparece por primera vez, la ubicación de este objeto en el siguiente frame (n+1) se estima determinando la similitud del histograma Hn del objeto con los histogramas obtenidos para el próximo frame. La comparación de los histogramas se repite de manera iterativa hasta que se establezca una similitud satisfactoria.

Continuando con este proceso la trayectoria del objeto es determinado en la secuencia de video. Si existe más de un objeto en la escena, este proceso es realizado para cada objeto por separado para estimar su trayectoria.

Cabe recalcar que un algoritmo de seguimiento no puede ser considerado como una solución global para un problema de seguimiento de objetos, pues en ciertos casos puede ser que un algoritmo no funcione de manera correcta. Por este motivo es necesario un mecanismo de retroalimentación para corregir los posibles errores. Por ejemplo, se puede combinar la detección de objetos en movimiento mediante sustracción de fondo con el

algoritmo de seguimiento de media deslizante. El proceso iterativo de este algoritmo no solo empezaría con la posición original de un blob en el frame previo sino que podría partir de varios blobs determinados mediante la sustracción de fondo.

2.4.2 Distinción entre Objetos y Personas

Hasta ahora se han estudiado las técnicas mediante las cuales los objetos pueden ser reconocidos y sus blobs extraídos. También se ha revisado literatura referente a como realizar el seguimiento de los objetos en movimiento utilizando estos blobs.

El siguiente paso consiste en determinar “que representan” o “que son” los blobs encontrados y por consiguiente los objetos. Por ejemplo, una cámara de vigilancia en interiores puede captar personas, grupos de personas, mascotas y hasta roedores. En cambio, el video proveniente de una cámara de vigilancia instalada en un ambiente exterior puede incluir vehículos, peatones, nubes, aves y otros animales.

En sistemas que requieren de información semántica es indispensable distinguir correctamente a las personas de entre otros objetos móviles, para realizar el seguimiento.

Los objetos móviles pueden clasificarse de acuerdo a su forma, color y movimiento. La forma de las regiones móviles se puede caracterizar de muchas maneras, incluyendo contornos 2D y siluetas, según ha sido estudiado en el apartado 2.3.3 de este capítulo.

Según [12], los blobs de objetos móviles se pueden clasificar en cuatro clases: personas aisladas, vehículos, grupos de personas y clutter, usando un clasificador de redes neuronales de tres capas. Esta clasificación puede llevarse a cabo utilizando un conjunto de características como: dispersión de los blobs móviles, área del blob, el radio de aspecto del bounding box del blob, e información del zoom de la cámara.

El proceso de clasificación puede ser más robusto si los histogramas de color de los blobs móviles reflejan la presencia de color de piel humana. La detección del color de piel humana puede realizarse en varios espacios de color como el RGB, YUV y el HSV. Según resultados experimentales la mejor representación de color para la detección de piel humana es el espacio YUV.

Otros esquemas de clasificación pueden hacer uso de la información temporal de los objetos, mientras que métodos más sofisticados aprovechan el conocimiento previo de la escena. Por ejemplo, se conoce que el movimiento de una persona es mucho más lento que el de un vehiculo.

2.4.3 Modelos y Patrones de Movimiento de Personas

Los métodos de modelamiento del cuerpo humano representan la estructura geométrica del cuerpo de varias formas matemáticas. Entre estos métodos están: representación por figura simple, modelamiento de contornos 2D y modelos volumétricos 3D.

2.4.3.1 Representación por Figura Simple

Esta representación es básicamente una combinación de segmentos de línea vinculados por articulaciones. Es adecuada para modelar los movimientos del torso, la cabeza y los miembros. Muchos movimientos humanos son modelados por los Modelos Escondidos de Markov (HMM). Este enfoque puede ser utilizado no solo para reconocer la existencia de seres humanos en la escena sino también para clasificar los movimientos humanos. Un modelo HMM que represente a cada postura humana y movimiento típico puede diseñarse y ser entrenado utilizando secuencias de video de

entrenamiento. De esta forma, el movimiento humano es reconocido mediante al modelo HMM que produzca la probabilidad más alta.

El aspecto negativo de este tipo de modelamiento es que utiliza representaciones muy simples del cuerpo humano, las cuales son la principal causa de los errores en la clasificación. Algunos métodos más sofisticados incluyen cintas 2D (ribbons) para obtener mejores modelos del movimiento articulado del cuerpo humano.

2.4.3.2 Modelamiento por Contornos 2D

El modelamiento basado en contornos 2D incluyen: serpientes, modelos de contornos activos y varias funciones unidimensionales que representan los contornos 2D de los blobs móviles.

En el primer enfoque es necesario inicializar una serpiente en el contorno del objeto móvil, el mismo que puede ser extraído mediante un método de sustracción de fondo. Para cada frame subsiguiente, la serpiente es montada en la misma posición que en el frame previo y de esta forma la serpiente extrae el contorno del objeto en el nuevo frame, siempre y cuando la velocidad del objeto permita que una cantidad razonable de píxeles se superpongan entre dos frames consecutivos. De hecho, es posible utilizar un algoritmo de predicción

como el filtro de Kalman para estimar la posición de la serpiente en el próximo frame.

También se pueden utilizar redes neuronales para decidir si un objeto móvil es considerado como ser humano. Estas redes neuronales pueden ser entrenadas mediante los llamados ”snaxels” del contorno activo. Debido al relativo tamaño pequeño del vector de características que consiste de los snaxels, la carga computacional de la clasificación es baja. Sin embargo, la estimación basada en contornos activos es un proceso iterativo y por tanto tiene un costo computacional relativamente alto.

Las redes neurales, modelos HMM y otras formas de clasificación pueden ser entrenadas utilizando funciones unidimensionales extraídas de los contornos 2D. Por ejemplo, una función 1D se puede generar calculando la distancia desde el centro de masa del objeto hacia sus bordes, considerando varios ángulos. Otra función unidimensional podría obtenerse proyectando el objeto móvil sobre los ejes vertical y horizontal y concatenando las proyecciones 1D.

La debilidad de este tipo de representación es que falla cuando existe oclusión. Por ejemplo, si una persona es parcialmente cubierta, estos métodos generan resultados erróneos.

2.4.3.3 Modelos Volumétricos 3D

Los modelos volumétricos incluyen cilindros elípticos, conos, esferas y otros modelos más sofisticados que son utilizados para modelar el cuerpo humano. Si se cuenta con un modelo 3D de la escena, estos modelos proveen una descripción precisa de los objetos móviles y de las personas. Para estimar los parámetros 3D es de gran ayuda contar con estéreo-visión o múltiples cámaras monitoreando la escena. Puesto que los modelos 3D requieren que más parámetros sean estimados y actualizados durante el proceso de seguimiento, esto obviamente conlleva un costo computacional alto en comparación a otros modelos.

Enfoques más avanzados incluyen elipsoides supercuádricas con deformaciones parametrizadas e incluso existen métodos basados en la física para ajustar modelos deformables de partes humanas también basados en estos elipsoides.

2.4.4 Análisis de los Enfoques más Adecuados

Según el estudio que se ha presentado en apartados anteriores de este capítulo, se puede notar que existen varios enfoques que pueden ser utilizados para el seguimiento de objetos y de seres humanos en secuencias de video.

Es más, es posible utilizar técnicas de reconocimiento en conjunto con técnicas de seguimiento para resolver el problema del rastreo de un objeto o persona. Las técnicas de reconocimiento pueden ser utilizadas para detectar la posición inicial del objeto de interés en la escena, es decir el ROI. Una vez que ha sido obtenido, sería factible utilizar técnicas de predicción para estimar la nueva posición del objeto en frames subsecuentes, como el Filtro de Kalman y el Algoritmo de Compensación.

Sin embargo, considerando el contexto y las condiciones del entorno de aplicación del sistema que este trabajo intenta desarrollar, un enfoque geométrico es apropiado debido a que permite la estimación del movimiento del objeto, es decir el seguimiento, mediante el uso de vectores de parámetros y matrices que ayudan en la compensación del movimiento de una fuente de video Pan-Tilt-Zoom (PTZ). En [9] se presenta un algoritmo de seguimiento que considera

el movimiento de la cámara, y está basado en el framework propuesto por Hager y Bellhumeur en [19].

2.5 Interpretación del Movimiento

Hasta este punto se han revisado los procesos de detección de movimiento, reconocimiento de objetos y el proceso de seguimiento. Estos procesos, aún cuando su grado de complejidad puede llegar a ser muy alto, son considerados de “bajo nivel” y prácticamente son la base de los sistemas de visión por computador. El siguiente paso consiste en extraer información semántica de la escena [8].

La información semántica que sea extraída depende de la aplicación de cada sistema. Por ejemplo, un sistema inteligente de vigilancia puede detectar movimientos y acciones sospechosas de los individuos y alertar al personal de seguridad. Algo similar puede ser implementado por un sistema de análisis de movimientos de un atleta y en base a esto determinar estadísticas de su performance.

Siendo un problema no resuelto hasta la actualidad, un sistema de video inteligente podría ser parte de un sistema integral basado en el paradigma de “computación diseminada” para analizar las acciones e intenciones de los individuos y en base a esto determinar la reacción

del sistema que controla el ambiente diseminado. Por ejemplo, un gimnasio inteligente podría controlar automáticamente sus máquinas y regular las velocidades, los pesos, el grado de dificultad, etc., basándose únicamente en las expresiones faciales y en los ademanes que un individuo realice.

Surge entonces el problema de entender las acciones humanas e interpretar su comportamiento. Actualmente este problema puede ser resuelto analizando el video para extraer algunos vectores de características. Durante la fase de interpretación, los vectores de características extraídos son comparados en contra de un grupo de conjuntos de vectores de referencia que representan las acciones humanas típicas. Este problema es básicamente equivalente a escoger una frase de entre un conjunto finito de frases. Por tanto, es necesaria una fase de entrenamiento en la cual se obtengan los vectores de referencia que corresponden a cada acción humana considerada.

Algunos métodos de clasificación de patrones, utilizados en la interpretación del comportamiento humano en video son:

- Programación Dinámica

- Modelos Escondidos de Markov (HMM) y Modelos de Mezcla Gaussiana (GMM)
- Redes Neuronales y Máquinas de Soporte de Vectores (SVM)
- Análisis del Componente Principal (PCA) y Análisis Discriminante Lineal (LDA)

La Programación Dinámica (DP) [21,22] es un método de optimización bien conocido y utilizado en problemas prácticos, como en el reconocimiento de voz en los años 80's bajo el nombre de "Dynamic Time Warping" (DTW). Se trata de un problema NP completo y como consecuencia el coste computacional no puede reducirse cuando el tamaño del problema aumenta. Por este motivo no es utilizado actualmente en sistemas de reconocimiento de voz, pues presenta un coste computacional relativamente alto en comparación con los sistemas basados en modelos de Markov.

Los Modelos Escondidos de Markov [23], conocidos como HMM por sus siglas en inglés, son máquinas de estado finito estocásticas. En el contexto del análisis del movimiento humano, un modelo de estado finito de Markov se añade por cada escenario posible y sus parámetros son entrenados con vectores de características de una acción humana común. El proceso de entrenamiento es un algoritmo

iterativo que se realiza fuera de línea, y es conocido como algoritmo de Baum-Welch. En el proceso de clasificación se aplica el vector de características a todos los modelos Markov y se calculan las probabilidades de cada uno. De esta manera, el modelo de mayor probabilidad determina la correspondiente acción humana que se analiza.

Las Redes Neuronales (NN), Redes Neuronales Tiempo-Retardo (TDNN) [27], Redes Neuronales Celulares (CNN) [26] y otros esquemas relacionados incluyendo el Soporte de Vectores (SVM) son usados ampliamente en el análisis de patrones. Entre las aplicaciones más comunes se encuentran la clasificación de rostros humanos de acuerdo a expresiones emocionales y los patrones de movimiento humano errático.

El Análisis del Componente Principal (PCA) [24,25] y el Análisis Discriminante Lineal (LDA) son utilizados también en la clasificación de patrones. Un método bien conocido “eigenface” desarrollado para el reconocimiento de rostros humanos se basa en PCA.

En resumen, la fase de interpretación o extracción de información semántica es muy compleja. Es quizá una de las tareas más

complicadas que debe realizar un sistema de visión por computador, especialmente aquellos sistemas que necesitan información referente al comportamiento humano.

El presente trabajo también necesita extraer información semántica. Es necesario obtener la distancia recorrida por el expositor desde que comienza a moverse hasta cuando el sistema decida realizar el enfoque automático. Puesto que el sistema necesita rotar una cámara para enfocar al expositor, es necesario también calcular los grados de rotación que la cámara debe girar.

2.6 Control de la Fuente de Video

El Control de la Fuente de Video es la parte mecánica del sistema que este trabajo tiene por objetivo construir. Puesto que este sistema deberá controlar el movimiento rotatorio de una cámara de grabación, es necesario que el sistema se comunique con este dispositivo. Para crear esta comunicación se podrá utilizar el puerto serial y a través de éste se enviarán los comandos que el dispositivo deberá ejecutar.

2.6.1 Interfaz Serial

La interfaz serial es un protocolo común para la comunicación entre dispositivos periféricos y computadores. El puerto serial envía y recibe bytes de información transmitiendo un bit a la vez y usualmente en formato ASCII. Para realizar la comunicación se emplean 3 líneas de transmisión: a) Tierra, b) Transmitir y c) Recibir. Puesto que la transmisión es asíncrona, es posible enviar datos por una línea mientras se reciben datos por la otra.

Las características más importantes de esta interfaz son la velocidad de transmisión, los bits de datos, los bits de parada, y la paridad. La velocidad de trasmisión indica el número de bits por segundo que se transfieren y se mide en baudios. Los bits de datos se refieren a la cantidad de bits en la transmisión en un paquete de información, que usualmente son 5, 7 u 8 bits. Los bits de parada son usados para indicar el fin de la comunicación de un solo paquete y proveen un margen de tolerancia para la diferencia que existe entre los relojes de los dispositivos. La paridad es una forma de verificar si se producen errores en la comunicación. Existen cuatro tipos de paridad: par, impar, marcada y espaciada. En la paridad par o impar se añade un bit después del último bit de datos para cerciorarse que se transmita un número par o impar de bits. La paridad marcada y espaciada no

verifican el estado de los bits de datos y asignan un bit de paridad en estado lógico alto y en estado lógico bajo respectivamente.

2.6.2 Protocolos de Comunicación Serial

Entre los protocolos más conocidos que implementan esta interfaz se encuentran: RS-232, RS-422 y RS-485

- **RS-232**

  Esta es la interfaz que se encuentra comúnmente en las computadoras PC IMB y compatibles. Se rige en el estándar ANSI/EIA-232 y es utilizado usualmente para conectar ratones, impresoras, módems e instrumentación industrial en un computador. Esta limitado a comunicaciones de punto a punto y distancias de hasta 15m.

  El RS-232C es un estándar que constituye la tercera revisión de la antigua norma RS-232 propuesta por EIA. Adicionalmente existe una versión internacional propuesta por CCITT conocida como V.24. No obstante, debido a que las diferencias entre estas normas son mínimas usualmente se menciona a V.24 y a RS-232C (e incluso sin el sufijo “C”) para referirse al mismo estándar.

Este protocolo consiste de conectores DB-25 o DB-9, ya que no se suelen emplear más de 9 pines del conector de 25 pines. Trabaja con señales digitales de +12V para el estado alto lógico y -12V para el estado bajo lógico. Las funciones más importantes de los pines en el conector son:

Datos: TXD (3), RXD (2)

Handshake: RTS (7), CTS (8), DSR (6), DCD (1), DTR (4)

Tierra: GND (5)

Otros: RI (9)

- **RS-422**

  El estándar EIA RS-422-A es el conector serial que utilizan los computadores Apple. Esta interfaz emplea señales eléctricas diferenciales, contrariamente a las señales referenciadas a tierra que utiliza el RS-232. La transmisión diferencial utiliza dos líneas para transmitir y recibir, y tiene la ventaja de ser más inmune al ruido y puede lograr mayores distancias que el RS-232. Debido a su inmunidad al ruido y la distancia que puede cubrir, esta interfaz es muy utilizada en aplicaciones y ambientes industriales.

- **RS-485**

  Esta es una mejora del RS-422 y esta definido en el estándar EIA-485. Permite incrementar el número de dispositivos que se pueden conectar e incluso es posible crear redes de dispositivos conectados a un solo puerto RS-485. Define las características necesarias para asegurar los valores adecuados de voltaje cuando se tiene la carga máxima. Debido a su gran inmunidad al ruido, este tipo de conexión serial es utilizado en aplicaciones industriales que necesitan dispositivos distribuidos en red conectados a una PC u otro controlador para recolección de datos. Adicionalmente, todos los dispositivos que se comunican utilizando RS-422 pueden también ser controlados por RS-485. La distancia máxima que se puede alcanzar con esta interfaz es 1200m.

  Las funciones de los pines RS-485 y RS-222 son las siguientes:

  Datos: TXD+ (8), TXD- (9), RXD+ (4), RXD- (5)

  Handshake: RTS+ (3), RTS- (7), CTS+ (2), CTS- (6)

  Tierra: GND (1)

2.7 Transmisión de Video por Internet

La transmisión de video a través de Internet puede efectuarse en dos tiempos de reproducción: en vivo y bajo demanda, y puede ser distribuido mediante Unicast o Multicast. En el caso particular de este proyecto el tiempo de reproducción es obviamente en vivo, mientras que la distribución multicast no aplica.

Por otra parte, existen codificadores de video que se encargan de realizar esta transmisión. Un ejemplo de esto es el Windows Media Encoder, el cual adicionalmente incorpora un SDK para el desarrollo de aplicaciones particulares según las necesidades del usuario.

2.8 Sistemas HMT - Human Motion Tracking

El seguimiento de objetos y de seres humanos ha generado un gran interés por parte de los investigadores y la comunidad científica. A continuación se presentan posibles campos de aplicación en los que se pueden implementar estos sistemas, y también se presentan varios proyectos existentes en este ámbito.

2.8.1 Campos de Aplicación de los Sistemas HMT

Actualmente el análisis del movimiento del ser humano ha tomado gran interés por parte de la comunidad científica por la gran utilidad que tiene en diversos campos de aplicación.

La precisión con la que se detecta el movimiento y su posterior seguimiento tiene un rol muy importante como base para aplicaciones de alto nivel que dependen de entradas visuales. Muchos sistemas inteligentes dependen básicamente de la interacción con los seres humanos y de la capacidad para interpretar sus acciones y entender sus actividades. Algunas aplicaciones específicas que pueden tener como base la detección y seguimiento del movimiento humano se presentan a continuación:

- Interfaces Hombre-Máquina en ambientes Pervasivos: Utilizando algoritmos de reconocimiento de gestos y posturas del ser humano, es posible crear ambientes pervasivos (computación diseminada) en los que los usuarios satisfagan sus requerimientos o necesidades sin la necesidad de interacción mecánica con el sistema.
- Dispositivos de seguridad para detección de peatones: Pueden ser incorporados en un vehiculo motorizado para

detectar peatones cercanos y de esta manera alertar a tiempo al conductor.

- Interacción Hombre-Máquina para robots móviles: Los robots pueden incorporar un software que reconozca los movimientos del ser humano y tomar decisiones en base a estos. Esto es muy útil para asistencia a las personas ancianas, ya que el robot podría saber cuando la persona necesita ayuda.
- Seguridad automatizada para lugares de concurrencia masiva: Las cámaras de vigilancia podrían ser controladas por un sistema de detección de situaciones sospechosas como robos y asaltos. Esto es de extrema utilidad para lugares abiertos como museos, aeropuertos, estaciones de bus, etc.
- Captura automática del movimiento humano: Existen sistemas de captura de movimiento utilizados para generar animaciones para videojuegos o producciones cinematográficas, basados en marcadores montados sobre el cuerpo de la persona. Estos sistemas tienen un precio muy elevado debido a todo el hardware que es necesario para la captura. Alternativamente se podría usar la detección y seguimiento del movimiento mediante algoritmos precisos y sin la necesidad de equipamiento adicional.

Estas aplicaciones están todavía en el ámbito de la investigación. Su desarrollo e implementación son factibles pero requieren de mucha investigación y estudio, especialmente por los factores que influyen en la precisión de los algoritmos y por las restricciones que aplican para los diversos ambientes en los que pueden ser utilizadas. Sin embargo, aún cuando el problema en general parece no estar resuelto todavía, se han desarrollado muchas herramientas que utilizadas correctamente en conjunto pueden llegar a crear un sistema de este tipo en un futuro no muy lejano.

2.8.2 Algunos Proyectos Actuales

En este apartado se presentan algunos proyectos que han sido desarrollados o que se encuentran en etapa de desarrollo y que se enfocan a distintos campos de aplicación.

- **AWS**

  Este es un Sistema de Avanzado de Alerta para Automotores Basado en Visión Monocular. La intención de este proyecto es reducir la cantidad y la severidad de accidentes de tránsito causados por fallas en la percepción del sistema visual de los conductores.

- **PFINDER**

  Pfinder [16] es un sistema de seguimiento e interpretación del comportamiento de personas en tiempo real. Este sistema utiliza un modelo estadístico multiclase de color y forma para obtener una representación 2D de la cabeza y las manos de la persona. Ha sido utilizado en muchas aplicaciones como interfaces inalámbricas, bases de datos de video, y codificación para ancho de banda bajo.

- **CAVIARE**

  Este es un proyecto europeo que estudia las técnicas de análisis de imágenes para mejorar el funcionamiento y rendimiento de los sistemas de vigilancia, con aplicaciones en ambientes urbanos y actividades comerciales. Sus siglas en inglés significan Reconocimiento Activo de Visión Conciente del Contexto Basado en Imágenes.

- **GMF4iTV**

  Sus siglas en inglés significan Marco de Referencia de Media Genérica para TV Interactiva. Es un proyecto europeo en el que se desarrollaron técnicas para el monitoreo de objetos en secuencias de video para crear enlaces de hiper-video en

programas de televisión interactiva. Estos enlaces permitían enlazar meta datos de los objetos identificados en las imágenes.

- **VSAM**

  Este es un proyecto de vigilancia y sus siglas en inglés significan Vigilancia y Monitoreo por Video. Fue desarrollado por la Universidad Carnegie Mellon y el Instituto Sarnoff fundado por DARPA. En este proyecto se estudiaron técnicas para el análisis del video en aplicaciones de vigilancia en ambientes urbanos y en campo de batalla, permitiendo alertar a un operador acerca de los incidentes utilizando múltiples sensores.

- **CSAIL**

  Como parte del proyecto VSAM, el equipo de Visión del MIT desarrolló técnicas para manipular de manera automática una cámara fija permitiendo enfocar una amplia zona del ambiente.

- **ADVISOR**

  Sus siglas en inglés significan Video Digital Anotado para Vigilancia y Recuperación Optimizada. Es un proyecto

europeo que estudió la detección de incidentes y la grabación de video para problemas de vigilancia en estaciones de trenes subterráneos.

2.9 Herramientas Disponibles

En este apartado se pretende realizar un breve análisis de las herramientas disponibles para la implementación del sistema, el cual es uno de los objetivos de este trabajo..

2.9.1 Lenguajes de Programación

Los proyectos y sistemas en el ámbito de la Visión por Computador utilizan en gran parte los lenguajes de programación C/C++, MATLAB y Java.

Actualmente es posible encontrar diversos toolkits y librerías que han sido desarrolladas para facilitar la implementación de sistemas de visión por computador, ya que incluyen rutinas y algoritmos de utilización común. Incluso existen aplicaciones completamente funcionales que pueden ser utilizadas como base de un sistema de mayor complejidad. Cabe recalcar que estas herramientas pueden tener tanto licencia libre como comercial.

Estas herramientas son típicamente desarrolladas para los sistemas operativos Windows y Linux. Sin embargo, existen algunas que se han desarrollado para sistemas operativos como IRIX.

El lenguaje MATLAB es muy utilizado en aplicaciones de procesamiento de imágenes y visión por computador y existen herramientas útiles para el desarrollo, tales como: GVF, Filtros Savitzky-Golay para imágenes 2D y MATLAB Pyramid Tools para descomposición de imágenes multi escala.

El lenguaje de programación JAVA también dispone de herramientas de gran utilidad como NeatVision y aplicaciones como la de Análisis de Disparidad de Imágenes, la cual estima la disparidad entre dos imágenes.

Sin embargo, los lenguajes más utilizados en este ámbito son C y C++. La mayor parte de las aplicaciones y herramientas disponibles están implementadas en estos lenguajes. Algunos ejemplos de recursos disponibles en estos lenguajes de programación son: OpenCV, ImLib3D, Microsoft Vision SDK, VASARI, etc.

2.9.2 Librerías

Existe una gran cantidad de Libreras y Aplicaciones desarrolladas para facilitar el procesamiento de imágenes y la implementación de sistemas de visión por computador. A continuación se presentan algunas herramientas disponibles y que merecen ser consideradas.

- **OpenCV**

  Esta es una librería no comercial desarrollada por Intel. Está implementada en el lenguaje C++ e incluye rutinas para visión por computador, aplicaciones de ejemplo de su utilización y diversos tutoriales. Cubre áreas como Métodos Geométricos, Calibración de Cámara, Seguimiento de Objetos, Pirámides de Imágenes y Reconocimiento de Objetos.

- **VXL**

  Esta es una colección de librerías C++ diseñadas para la investigación e implementación de visión por computador. Fue creada con la intención de obtener un sistema rápido, ligero y consistente. Esta diseñada para ser portable y por tanto ser utilizable en muchas plataformas.

- **CVIPtools**

Son herramientas para el procesamiento de imágenes y visión por computador basadas en GUI. Esta desarrollada en ANSI-C y soporta las plataformas Windows NT y Unix. Incluye análisis de imágenes, compresión, realzado, restauración y otras utilidades. Es usada para educación, investigación y para desarrollo de aplicaciones.

- **IPP IPL**

  Esta es la ultima versión de librerías desarrolladas por Intel. Anteriormente existían librerías por separado para procesamiento de imágenes, procesamiento de señales, codificación de video y otras. Intel Integrated Performance Libraries incluye todas estas librerías. Este es un producto comercial, pero es posible obtener diferentes tipos de licencias, incluyendo de investigación y de evaluación.

- **ImLib3D**

  Es una librería de código abierto para el procesamiento de imágenes volumétricas 3D y está implementada en el lenguaje C++. Tiene un visor opcional multiplanar basado en OpenGL, el cual sirve de gran ayuda en la experimentación con

procesamiento de imágenes. Trabaja con Plantillas, Iteradotes, Línea de Comandos, Interpolación BSpline entre otros.

- **MATLAB Pyramid Tools**

  Son herramientas para la descomposición de imágenes piramidales, es decir multi escala. Incluyen Pirámides Laplacianas, QMF, Wavelets, Convolución, Histogramas y Generación de Imágenes Sintéticas.

- **NeatVision**

  Es un ambiente para análisis de imágenes y desarrollo de software basado en Java. Provee acceso de alto nivel a una amplia gama de algoritmos de visión a través de una interfaz grafica. Contiene más de 200 algoritmos de procesamiento de imágenes y datos en general. Soporta desarrollo basado en Java AWT Imaging, Java 2D Imaging y Java Advanced Imaging. Además ofrece herramientas de visualización y análisis como zoom, seudo color, exploración de intensidad, histogramas, mallas 3D y soporta una gran cantidad de formatos de imágenes.

- **Microsoft Vision SDK**

Es una librería de visión por computador para Visual C++.Define un objeto de imagen y soporta adquisición de imágenes independiente.

- **Gandalf**

  Esta es una librería escrita en C de algoritmos numéricos y visión por computador. Permite desarrollar nuevas aplicaciones que sean portables y rápidas. Incluyes muchas rutinas útiles para visión incluyendo calibración de cámara, homografias, operaciones con matrices, y detectores de características. Incluye paquetes para manipulación de imágenes, algebra lineal, estructuras y manejo de memoria, visión por computador y rutinas numéricas. Las principales características de diseño de Gandalf son: uso de memoria eficiente, énfasis en el soporte de algoritmos numéricos, representación de imágenes flexible y eficiente y operaciones de matrices y vectores.

En resumen, es posible encontrar herramientas comerciales y de código abierto de gran utilidad para aplicaciones específicas en el campo del procesamiento de imágenes y de la visión por computador. Lo importante es elegir las herramientas más

adecuadas para la aplicación a la que se enfoque un sistema y de esta manera aprovechar las facilidades que éstas proveen.

2.10 Sumario

En este capítulo se han presentado las técnicas y enfoques más conocidos para la detección de movimiento, reconocimiento y seguimiento de objetos, así como para la representación de los objetos.

Se ha tratado en lo posible de revisar estas técnicas de una manera clasificada, según su campo de aplicación. Sin embargo, como el lector habrá podido notar, muchas de estas técnicas son utilizadas en conjunto para resolver el problema del seguimiento de objetos y de personas. Algunos enfoques utilizan técnicas de reconocimiento para determinar las posiciones iniciales de los objetos de interés o establecer las regiones de la imagen que serán procesadas, y utilizan obviamente técnicas de seguimiento para determinar las nuevas posiciones de los objetos en un momento dado, y en base a esto tomar decisiones.

Como se ha podido notar, la interpretación del movimiento de los objetos y especialmente de los seres humanos tiene gran

importancia para sistemas avanzados que se enfocan en su mayoría a la computación diseminada o también llamada pervasiva.

En conclusión, existen muchas técnicas y herramientas que pueden ser utilizadas en conjunto para resolver problemas de visión por computador, como sistemas de video inteligente para vigilancia. Sin embargo, el uso de estas técnicas implica la consideración de restricciones específicas, sin las cuales las técnicas fallan en su funcionamiento. Esto conlleva a limitaciones de uso de estos sistemas y básicamente provoca que los sistemas que actualmente pueden ser implementados, sean utilizados para aplicaciones muy específicas y en ambientes en su mayoría controlados.

CAPÍTULO 3

3 ANÁLISIS Y DISEÑO

3.1 Introducción

Los sistemas de video inteligente son desarrollados por lo general teniendo en consideración las diferentes etapas que intervienen en el sistema, de manera independiente pero sin descuidar la interacción entre las mismas.

En este trabajo se pretende realizar un sistema que sigue este esquema, pero que además tendrá una estructura modular y será parametrizable mediante una configuración inicial que regirá el comportamiento del mismo. La estructura básica que se persigue consiste en módulos que trabajan de manera independiente y que en conjunto permiten el funcionamiento del sistema. De esta manera, al tener un diseño modular se permite que el sistema pueda ser mejorado por partes, es decir, que sea factible implementar mejores algoritmos en módulos específicos para obtener resultados más eficientes, con efectos secundarios nulos o mínimos en los demás módulos que componen el sistema.

Adicionalmente, se pretende que estos módulos puedan ser utilizados como componentes fundamentales de otros sistemas que se pudieran desarrollar en el futuro.

3.2 Análisis de Requerimientos

Este tipo de sistemas requieren del cumplimiento de una serie de restricciones relacionadas con el ambiente en el cual será utilizado y con el sujeto que interviene en la escena (véase 2.1.1.3). Las restricciones que necesitará este sistema se detallarán en los siguientes apartados.

Básicamente el sistema contará con un software que se encargará de la parte de análisis de la escena y control de la cámara, y un dispositivo periférico que funciona como fuente de video. Cabe recalcar que el software será instalado en un computador, el cual a su vez será utilizado como servidor del video. Por otra parte, un computador remoto se encargará de recibir el video que es transmitido por Internet.

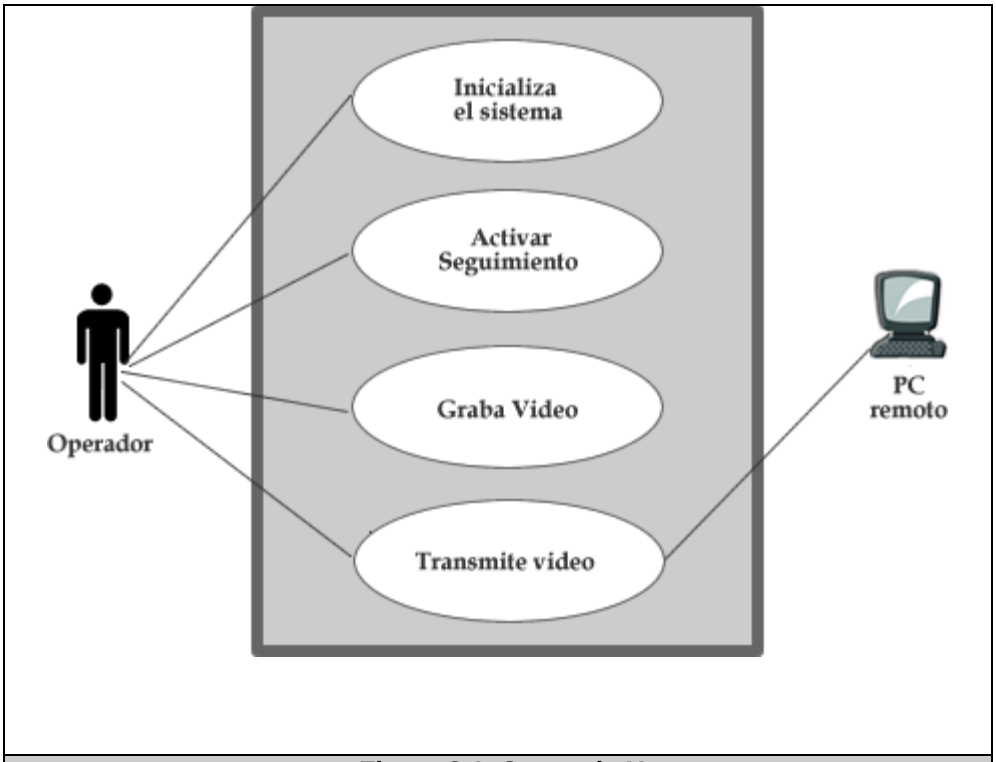

Figura 3.1. Casos de Uso.

3.2.1 Requerimientos Funcionales

Entre los requerimientos de funcionalidad de este sistema tenemos los siguientes:

- El sistema deberá ser capaz de reconocer la ubicación del conferencista en la escena y controlar la fuente de video de tal forma que se mantenga, a dicho conferencista, siempre enfocado en el centro de la escena.
- La fuente de video será monocular, es decir, se contará con una sola cámara de grabación de video con capacidad de

rotación. La señal de video será composite y no será necesario un tipo especial de cámara.

- No se tomará en cuenta la oclusión entre objetos. El sistema deberá intentar reconocer al conferencista y mantener su ubicación.
- La fuente de video contará con dos (2) grados de libertad para su movimiento rotacional en sentido horizontal solamente.
- Se deberá proveer una interfaz de usuario que permita visualizar el video registrado por la cámara en todo momento. Adicionalmente deberá ser capaz de mostrar gráficamente la ubicación del conferencista.
- El usuario deberá ser capaz de establecer un rango de desplazamiento permitido (rectangular) que el sistema utilizará para discriminar si es preciso comenzar el movimiento de la cámara de video. La interfaz de usuario deberá proveer las facilidades necesarias.
- El video registrado por la fuente deberá ser transmitido en tiempo real mediante Internet. Un computador remoto podrá conectarse al servidor de video y recibir la señal transmitida.
- Se deberá poder grabar la señal de video original en un archivo con un formato de compresión de video.

- La interfaz de usuario deberá proveer controles tanto para la grabación del video como para la transmisión del mismo.

3.2.2 Requerimientos No Funcionales

En este apartado se detallarán aquellos aspectos que deberán ser considerados para la implementación del sistema y su posterior instalación.

- Solamente una persona podrá estar en movimiento en el escenario para evitar la oclusión. Aunque el sistema reconocerá varios objetos en movimiento presentes en la escena, solo uno de ellos podrá ser considerado como persona.
- El sistema funcionará con una cámara de video ubicada a una distancia considerable, de tal forma que se pueda cubra la mayor parte del escenario.
- En cuanto a altitud, la cámara tendrá que ubicarse por encima del nivel de la persona. De esta manera se disminuyen los problemas en el reconocimiento, producidos por una posible superposición de objetos.
- No se necesitará de la presencia de un operador que manipule el sistema durante la grabación. Sin embargo un operador deberá establecer la configuración inicial.

- Para el correcto funcionamiento del sistema, el ambiente no deberá tener cambios drásticos en la intensidad lumínica.
- El conferencista no deberá portar vestimenta de color demasiado similar al fondo del escenario.
- El conferencista podrá desplazarse en el escenario a una velocidad lenta y no necesariamente constante.

3.3 Diseño Modular

El sistema a desarrollar estará formado por varios componentes que se encargarán de una función especifica para que, trabajando en conjunto, se pueda realizar la tarea de seguimiento al expositor y la posterior transmisión del video registrado hacia su destino final mediante Internet.

En este diseño modular los componentes serán independientes en cuanto a su funcionalidad; es decir, cada módulo estará encargado de realizar un procesamiento específico, el cual es completamente diferente al de los demás módulos. Sin embargo, estos módulos deben interactuar entre si continuamente, pues el procesamiento que un módulo debe realizar se lo aplicará al resultado del trabajo realizado por otro módulo. En otras palabras, el producto que genere un módulo servirá como materia prima para el siguiente módulo.

El propósito de tener módulos independientes consiste en que de esta manera, al implementar algoritmos más eficientes u otros enfoques que optimicen la funcionalidad de un módulo, es posible mejorar el procesamiento y la salida que produce dicho módulo. Así por ejemplo se podrá incrementar la precisión, rendimiento y eficiencia del sistema, mejorando sus componentes.

Una ventaja adicional de este diseño modular es que será posible reutilizar los módulos para desarrollar otras aplicaciones, requiriendo mínimos cambios para su adaptación. Incluso podrán ser utilizados para investigación y experimentación de nuevas técnicas.

3.4 Arquitectura

El sistema estará formado por componentes modulares. Estos módulos se encargarán de una función en especifico y tendrán como objetivo principal recibir una entrada, realizar un procesamiento en base a esa entrada y producir un resultado. Estos módulos tendrán una funcionalidad independiente, lo cual no implica que sean utilizados por el sistema siguiendo un esquema de cascada, ya que este utilizará cada módulo continuamente durante su ejecución.

En principio, el sistema deberá detectar movimiento en la escena y proceder a identificar la posición del expositor. Una vez que obtiene esta posición deberá comenzar a rastrear al expositor, el cual podrá moverse o detenerse en cualquier momento. Cuando el sistema detecte que el expositor esta alejándose del rango de desplazamiento permitido, deberá comenzar el movimiento rotacional en la fuente de video para recuperar el enfoque al expositor.

Según se describió en 3.2.1, en este sistema la cámara de video tendrá 2 grados de libertad, lo cual significa que solamente tendrá movimiento en un eje de rotación, el eje vertical de la cámara. De esta manera, la fuente de video tendrá la capacidad de girar horizontalmente para cubrir el escenario completo. Sin embargo, en la práctica el conferencista pudiera también moverse hacia la cámara o alejarse de ella, lo cual implica que la cámara debería girar en el eje horizontal, pero dado que se trata de una conferencia, usualmente el expositor se mueve en el escenario de un lado hacia el otro, de derecha a izquierda y viceversa, produciéndose un movimiento aproximadamente horizontal. Por este motivo se descarta el movimiento en profundidad del conferencista.

No obstante, si la fuente de video es ubicada a una distancia relativamente lejana del escenario, de tal manera que el conferencista aparezca como un objeto pequeño en la escena, la persona podrá moverse en profundidad sin salirse del campo de visión de la fuente de video. Por otra parte, si la persona se aleja en demasía, lo cual provoca que ésta aparezca como un objeto muy pequeño en la escena, no tendría caso realizar el seguimiento, ya que por más que la persona se mueva en cualquier sentido y dirección, aún permanecería dentro del campo de visión de la fuente de video. Por lo tanto se debe establecer una distancia adecuada de tal manera que el conferencista sea percibido lo suficientemente bien para que no sea necesaria la rotación vertical de la cámara.

3.4.1 Modelo General del Sistema

Este sistema será implementado con cinco módulos cuya funcionalidad será específica e independiente pero con influencia mutua. Existirá un módulo por cada tarea principal del sistema, es decir, la adquisición del video, el reconocimiento, el seguimiento, el control de cámara y la transmisión del video.

En primera instancia se deberán capturar las imágenes provenientes de la fuente de video mediante el Módulo de Adquisición, el cual se

encargará de realizar un pre-procesamiento que incluye desde conversión de las imágenes a una diferente representación hasta reducción de ruido. Es obvio que este módulo funcionará en todo momento, pues de él proviene la entrada global del sistema. Luego es necesario comenzar a monitorear la escena para detectar movimiento y proceder al reconocimiento del expositor en la misma. Para lograr realizar este reconocimiento se utilizará un modelo del fondo de la escena, del cual se extraerán los objetos en movimiento. Puesto que es probable que se detecte más de un objeto en la escena, debido a movimientos imprevistos o por falsas alarmas, es necesario también hacer una clasificación de los objetos detectados e identificar el objeto que representa al expositor.

El modelo del fondo será obtenido en la Fase de Inicialización del sistema, la cual se realizará con la completa ausencia del expositor y será un requisito indispensable para el posterior funcionamiento del sistema. En pocas palabras, en esta fase el sistema reconocerá el campo de acción en el que trabajará, es decir, el conocimiento previo de la escena.

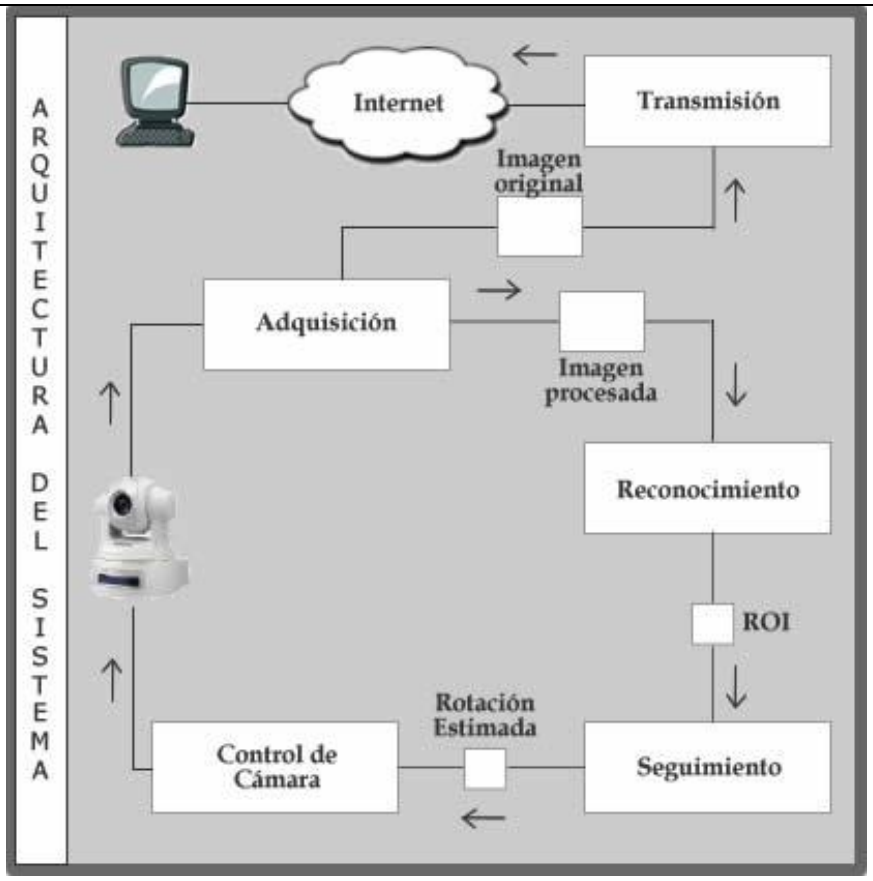

Figura 3.2. Arquitectura del Sistema.

Una vez que se obtiene la posición del expositor en un frame dado, el seguimiento o “tracking” debe tener lugar. Básicamente consistirá en verificar continuamente la posición del expositor y cuando detecte que éste se está alejando del rango establecido para el desplazamiento deberá calcular su orientación y movimiento para de esta manera estimar los grados de rotación que la cámara deberá girar para recuperar el enfoque.

El módulo de control de cámara por su parte, se encarga fundamentalmente de controlar la parte mecánica del sistema, es decir la cámara de video. Este módulo recibe los parámetros de rotación que son calculados por el módulo de seguimiento y establece la comunicación con el dispositivo para su posterior control.

Finalmente, el video debe ser comprimido en un formato adecuado para streaming y podrá ser visualizado por un cliente conectado a la red. Opcionalmente se puede transmitir el video a un servidor para que este luego se encargue de manjar un número mayor de clientes.

3.4.2 Adquisición de Video

Este módulo, como su nombre lo indica, se encarga de capturar las imágenes provenientes de la secuencia de video registrada por la cámara, y entregar estas imágenes al Módulo de Reconocimiento para iniciar el análisis de visión por computador y al Módulo de Transmisión para crear el flujo de video en la red.

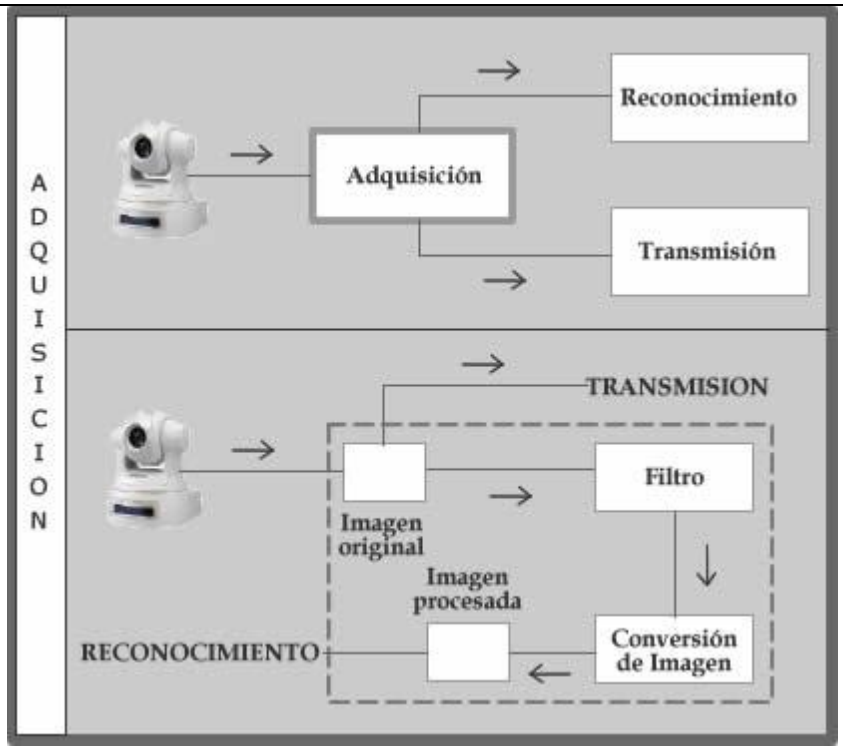

Figura 3.3. Mecanismo de Adquisición de Imágenes.

Sin embargo, antes de entregar estas imágenes, es necesario realizar un pre procesamiento que facilite la posterior extracción de información que se encuentra en dichas imágenes. Este pre procesamiento básicamente consiste de un filtro de eliminación de ruido y posteriormente una conversión a una diferente representación de imágenes. De esta manera se reduce la cantidad de información que debe ser procesada y por consiguiente aumenta la eficiencia del sistema.

Para cada frame proveniente de la secuencia de video, se realizará el respectivo pre-procesamiento, para posteriormente entregar dicho frame al Módulo de Reconocimiento. Adicionalmente este módulo envía la señal de video original hacia el Módulo de Transmisión de tal manera que este último transfiera la señal de video original hacia el destino final.

3.4.3 Reconocimiento

La detección de los objetos móviles en la escena, así como el reconocimiento del expositor en la misma son tareas que serán realizadas por el Módulo de Reconocimiento.

Una vez que el sistema comienza su ejecución, es necesario comenzar a monitorear la escena para detectar movimiento y proceder al reconocimiento del expositor. Las imágenes provenientes del Módulo de Adquisición contienen información general de la escena desde un punto de vista global, es decir, tanto de los objetos en movimiento como del fondo compuesto por los demás objetos (estáticos) y el resto de la escena.

La primera tarea que este módulo deberá realizar consiste en detectar movimiento en la escena, para lo cual necesitará determinar

en primera instancia cuales son los objetos móviles presentes en la misma. Puesto que es probable que existan varios de estos objetos en la escena, será preciso realizar una segmentación de la imagen para separar aquellos que hayan sido encontrados. Adicionalmente, este moduló deberá ser capaz de identificar los objetos que posiblemente representen a una persona y descartar aquellos objetos que no sean de interés. Por otra parte, una vez que los objetos de primer plano hayan sido seleccionados, será necesario determinar cual de estos objetos corresponde al conferencista en la escena.

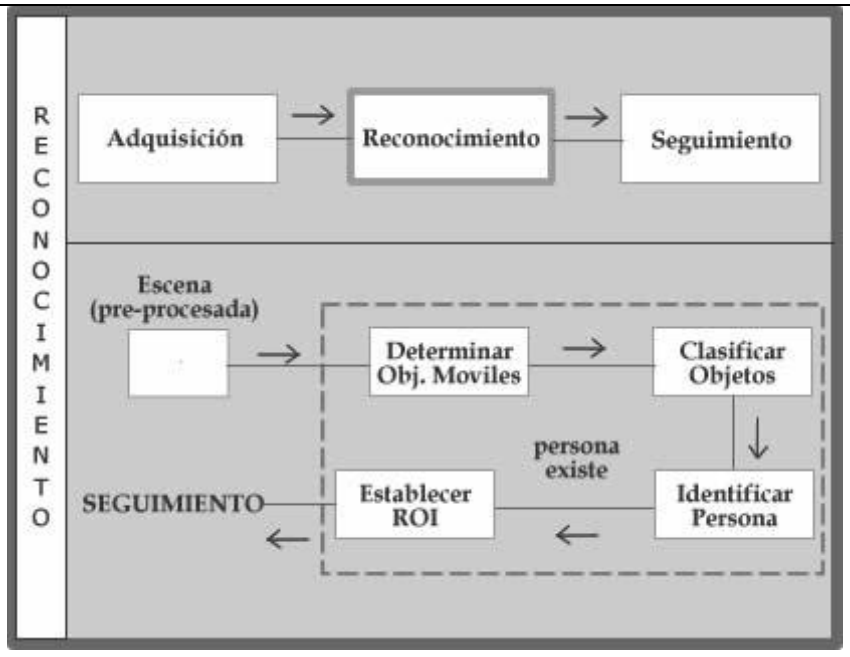

Figura 3.4. Mecanismo de Reconocimiento de la Persona.

Para detectar el movimiento en la escena se puede utilizar una técnica conocida como “sustracción de fondo. Para esto se necesita

conocer el fondo de la escena, sin los objetos móviles, y de esta forma mediante sustracción de imágenes se obtienen los objetos que presentan movimiento. Por otra parte, la sustracción de fondo solamente revela las regiones de la imagen que corresponden a variaciones de intensidad, lo cual puede ser considerado como movimiento, pero en cambio no determina que regiones de pixeles corresponden a determinado objeto. Es por esto que se necesita de un proceso de segmentación que encuentre los objetos en la imagen y que determine sus posiciones en la misma.

Para realizar esta segmentación existen varias técnicas entre las cuales se puede mencionar: segmentación por 4-contectividad, segmentación por 8-conectividad y segmentación por contornos (véase 2.3). No obstante, independientemente del algoritmo de segmentación utilizado, la tarea de este módulo es la misma, detectar la región de la imagen en la que se encuentra el conferencista. Por tanto se podrá ir optimizando esta tarea conforme la tecnología de reconocimiento mejore.

Finalmente, cuando se haya encontrado la posición de la persona en la imagen, se deberá establecer dicha ubicación como “región de

interés” (ROI), la cual será utilizada posteriormente por el Módulo de Seguimiento.

3.4.4 Seguimiento

La función básica de este módulo será monitorear continuamente el desplazamiento del expositor en el escenario. Cuando se detecte que esta persona está saliendo del campo de visión permitido, se deberá comenzar a rotar la fuente de video hasta lograr recuperar el enfoque, es decir, que el expositor se encuentre cercano al centro de la imagen.

A diferencia del campo de visión total de la cámara, el campo de monitoreo es un subconjunto de la imagen de la escena capturada por la fuente de video. Este es un rango que será establecido en la configuración inicial del sistema y básicamente es un cuadro de tamaño menor a las dimensiones de altura y anchura de la imagen original. Mientras el expositor se mantenga dentro de este rango no existe la necesidad de mover la cámara.

Este módulo recibirá como entrada la posición del conferencista en la escena, la cual es estimada por el Módulo de Reconocimiento. Esta posición estará definida por una región de interés (ROI) rectangular

en la imagen, el cual representa al cuerpo de la persona. De esta manera este módulo verificará continuamente que este ROI se encuentre dentro del rango permitido.

Cuando se detecte que este ROI se encuentra en los límites de este rango o fuera de ellos, se deberán enviar comandos de rotación a la cámara de video. Para el efecto, este módulo necesita conocer la medida en que la cámara deberá girar, la cual será especificada por los parámetros de rotación de la cámara, los cuales son actualizados cada vez que la cámara se mueve y que además son utilizados en todo momento por el Módulo de Reconocimiento.

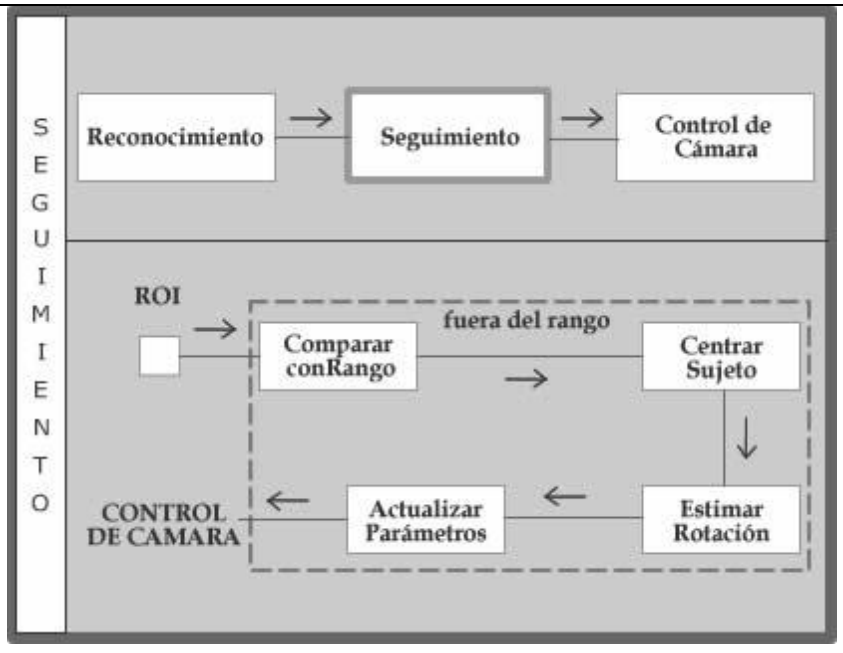

Figura 3.5. Mecanismo del Seguimiento.

Una vez que este módulo verifica que es preciso rotar la cámara de video, necesita conocer la orientación del desplazamiento del expositor. En el caso en particular, necesitará determinar si la cámara deberá girar hacia la izquierda o hacia la derecha, debido a los 2 grados de libertad permitidos. De esta manera, este módulo tendrá que consultar la posición de la cámara, es decir sus parámetros de rotación en un momento determinado, y por consiguiente recuperar los grados de rotación que la cámara deberá girar. Luego de esto, será necesario modificar la posición vigente de la cámara actualizando sus parámetros de rotación.

Finalmente la rotación estimada será entregada al módulo de Control de Cámara, el cual se encargará de la comunicación y del movimiento mecánico del dispositivo.

3.4.5 Control de Cámara

La tarea fundamental de este módulo es la comunicación con la cámara de video. Básicamente se tendrán que transmitir comandos de movimiento hacia este dispositivo una vez que la conexión se haya establecido. Esta comunicación entre los componentes de hardware y software del sistema requiere de una “Interfaz de Comunicaciones”, según se estudió en la sección 2.6.

Este módulo tiene por objeto notificar a la fuente de video que debe comenzar su rotación, por lo cual deberá especificar el sentido en el que ésta tendrá que girar y por supuesto la cantidad de rotación. Esta cantidad estará definida en grados y el sentido de rotación será establecido como positivo o negativo, dependiendo de si el movimiento es hacia un lado o hacia el otro.

Una vez que la fuente de video reciba estos datos, procederá a ejecutar la acción correspondiente y a continuación deberá comunicar al módulo el resultado de la operación, es decir, si la operación fue exitosa o si ocurrió algún error que imposibilitó realizar dicha acción. En cualquier caso, la cámara tendrá que retornar el estado de la operación y será el módulo quien se encargará de determinar la causa del fallo.

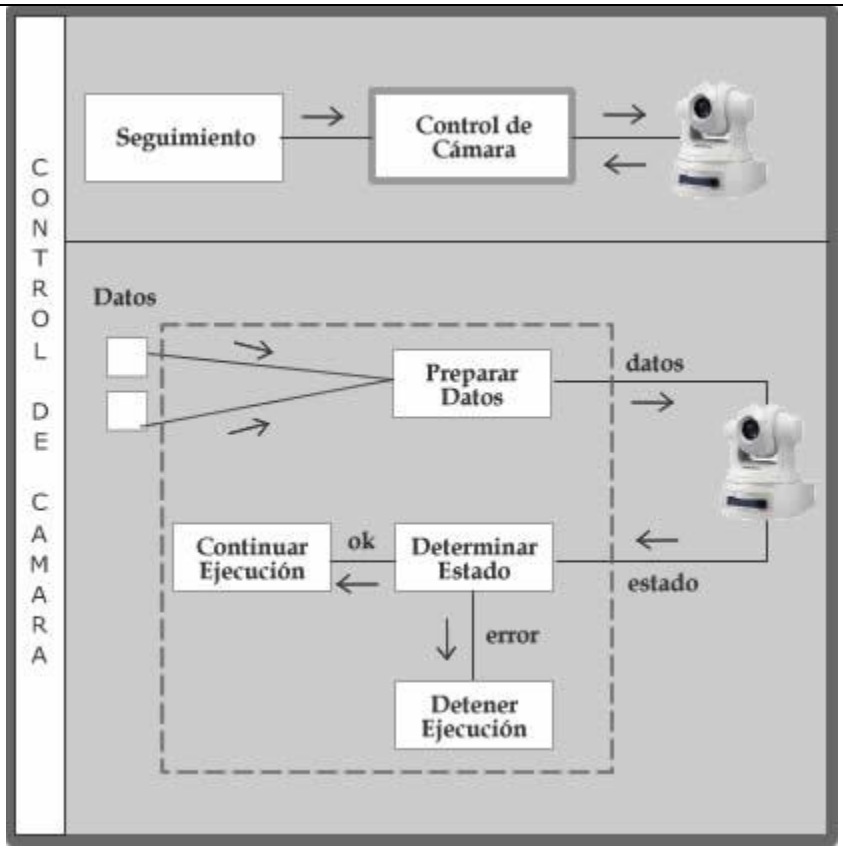

Figura 3.6. Mecanismo del Control de Cámara.

En cuanto a la implementación misma de la comunicación con la cámara de video, existen librerías que implementan protocolos de transmisión debido a que la mayoría de los lenguajes de programación no ofrecen, de manera integrada, el soporte necesario para el efecto. Para el desarrollo de este módulo será necesario utilizar una librería que implemente esta comunicación a bajo nivel y que provea las funciones necesarias para utilizarlas a nivel de aplicación. La librería que se utilice dependerá obviamente del

lenguaje de programación que sea utilizado para la implementación del sistema.

3.4.6 Transmisión y Grabación de Video

El objetivo final del sistema es crear un flujo de video que se pueda transmitir a clientes remotos o que se pueda capturar en forma de un archivo digital de video. Este módulo es completamente independiente de los demás módulos que conforman el sistema y obviamente no depende de ellos, con excepción del Módulo de Adquisición, el cual es el encargado de entregar la señal original de video para su posterior transmisión.

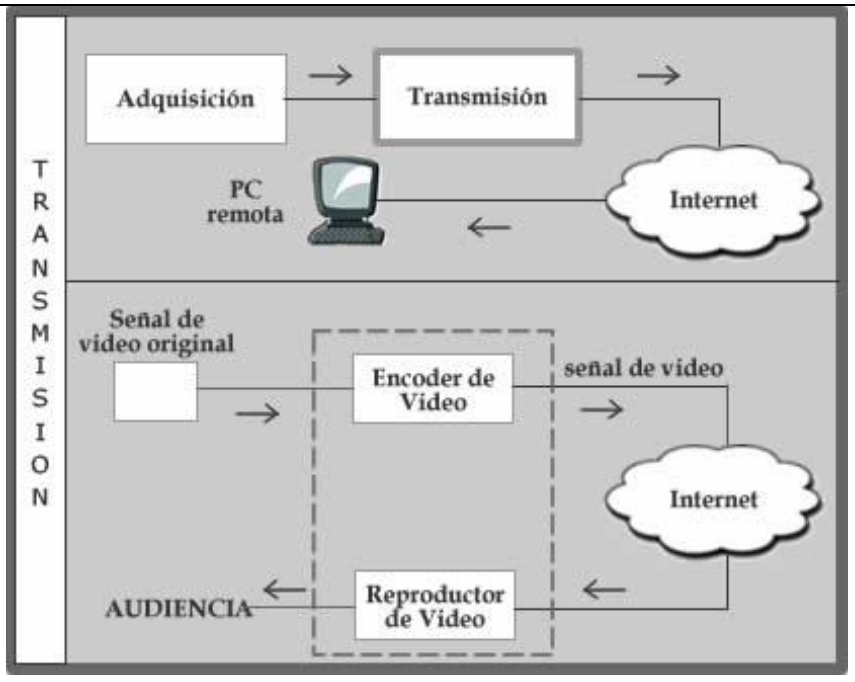

Figura 3.7. Mecanismo de Transmisión de Video por Video Conferencia.

Una vez que este módulo reciba la señal de video, deberá establecer la conexión con el receptor remoto. Para lograr este objetivo, será necesario utilizar un “Encoder” de video que se encargue de comprimir el video y prepararlo para su transmisión por Internet o para su grabación. Afortunadamente, en la actualidad existen varias herramientas que permiten realizar esta tarea. Sin embargo, se deberá tener en consideración que solo una aplicación puede recibir la señal de video original, por lo que será necesario una solución alternativa.

Esta solución podría ser implementada utilizando un SDK disponible para el desarrollo de este tipo de aplicaciones, como el Windows Media Encoder SDK, o bien se podría repartir la señal de video para el sistema y para una aplicación independiente que haga la transmisión del video.

3.5 Sumario

En este capítulo se han establecido los requerimientos tanto funcionales como no funcionales que deberán ser considerados por el sistema, así como su arquitectura desde un punto de vista global, según se lo aprecia en la figura 3.2.

Por otra parte se ha descrito el diseño modular del sistema y se ha presentado un breve análisis de las ventajas de este enfoque, así como la posibilidad de utilizar estos módulos independientes en otras aplicaciones.

Adicionalmente se ha introducido la arquitectura desde un punto de vista relacionado con cada uno de los componentes. Básicamente se ha especificado la funcionalidad de cada uno de estos componentes, así como la interacción que existe entre ellos.

Puesto que este capítulo trata del Diseño del Sistema, no han sido mencionadas posibles herramientas como lenguajes de programación y librerías que puedan ser utilizadas. Sin embargo, esto será tratado en los primeros apartados del siguiente capítulo de este trabajo.

CAPÍTULO 4

4 IMPLEMENTACIÓN

4.1 Selección de las Herramientas

Existen diversas herramientas disponibles, como lenguajes de programación y librerías, que pueden ser utilizadas para la implementación de este tipo de sistemas, según se estudió en la sección 2.9 de este trabajo.

El lenguaje de programación que será utilizado es C++ con el ambiente de desarrollo Microsoft Visual Studio 2005 (VC++2005). Puesto que este lenguaje es orientado a objetos, se acopla perfectamente a la estructura modular propuesta en la arquitectura del sistema (véase 3.4.1) y además permite crear interfaces gráficas con uso eficiente de memoria.

Esta eficiencia es de extrema importancia para la implementación de este sistema, pues el procesamiento de secuencias de video en tiempo real demanda una gran capacidad de computación y tiempos de respuesta relativamente cortos.

Por otra parte, se utilizará una librería de visión por computador y procesamiento de imágenes de código libre. Se trata de la librería OpenCV, la misma que es un proyecto impulsado por Intel, y que proporciona una gran cantidad de funciones que son de gran ayuda para el desarrollo de estos sistemas. A diferencia de otras librerías, OpenCV se enfoca más en la parte de visión por computador que en el procesamiento de imágenes como tal, y proporciona funciones relacionadas con el Análisis Estructural de Imágenes, Reconocimiento de Objetos y Análisis de Movimiento.

4.2 Construcción del Sistema

El sistema se conforma de cinco componentes modulares. Cada uno de éstos poseen una función específica pero siguen un esquema similar: reciben una entrada, efectúan el correspondiente procesamiento y, generan una salida que será utilizada como entrada para otro módulo. Además existe un componente de hardware que actúa como fuente de video. Este dispositivo es la cámara Canon VC-C50i.

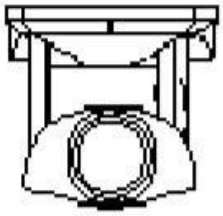

Figura 4.1. Cámara de Video Canon VC-C50i [F3].

La interfaz gráfica del sistema se construye completamente independiente del resto de módulos. Sin embargo, es ésta la que actúa como el controlador principal del sistema y por consiguiente ejecuta los procedimientos implementados por cada módulo. En otras palabras, la interfaz gráfica es la que recibe las órdenes por parte del usuario y delega estas tareas a los módulos correspondientes.

Por otra parte, el enfoque que se utiliza para el reconocimiento de la persona en la escena, es una adaptación del modelo propuesto por Dellaert y Collins en [12] el cual ha sido utilizado en un proyecto de Video Vigilancia desarrollado por la universidad Carnegie Mellon [11]. Más adelante en este capítulo se presenta con mayor detalle este enfoque.

En cuanto al seguimiento, se utilizan historiales de movimiento de la persona para poder determinar el sentido y la dirección de su desplazamiento, y de esta manera poder estimar la rotación que

deberá efectuar la cámara en caso de que la persona se aleje del rango de desplazamiento permitido.

Por último cabe mencionar que la comunicación con el dispositivo de video es asíncrona, es decir, una vez que el controlador principal envía un comando a dicho dispositivo, el sistema no espera una respuesta para continuar su ejecución. Por este motivo, el sistema es implementado utilizando hilos que permitan la ejecución de tareas en paralelo.

4.2.1 Módulo de Adquisición

Este módulo tiene como principal función obtener cada imagen de la secuencia de video generada por la fuente y prepararla para su posterior procesamiento.

Esta tarea es realizada utilizando la librería VFW (Video For Windows). Puesto que se trata de una secuencia de video existe la necesidad de obtener cada frame de esta secuencia, para lo cual se definen funciones de tipo callback que son ejecutadas cuando un nuevo frame es generado. Esto permite el acceso a los datos de imagen del frame y por consiguiente es posible aplicar todo el procesamiento necesario.

Una vez que el frame es entregado a la función de callback, esta se encarga de realizar una adecuación de este frame, para que pueda ser utilizado por los demás módulos. En otras palabras, una imagen definida por la librería VFW es diferente a la definida por OpenCV, por lo cual se requiere de una conversión entre estos diferentes tipos de imagen.

Luego, con el formato de imagen requerido se procede a realizar el reconocimiento de los objetos móviles en la escena, clasificar estos objetos y determinar la ubicación del expositor. Esto es realizado por el módulo de reconocimiento el cual genera una región de interés que el módulo de seguimiento deberá tomar para posteriormente realizar el análisis del movimiento de la persona.

Como resultado, el módulo de seguimiento establece la rotación que deberá reproducir la cámara de video en caso de que el expositor haya salido del rango de desplazamiento permitido; de lo contrario se comunicará que no es necesario el movimiento de la fuente de video.

Por último, este callback envía un mensaje que es recibido por el hilo de ejecución principal, el cual interpreta el mensaje e invoca métodos

definidos en el módulo de control de cámara para que éste a su vez ejecute las acciones necesarias para el control del dispositivo, según lo que haya sido establecido por el módulo de seguimiento.

4.2.2 Módulo de Reconocimiento

Una vez que el sistema comienza su ejecución, es necesario comenzar a monitorear la escena para detectar movimiento y proceder al reconocimiento del expositor. El enfoque que se utiliza para alcanzar este objetivo es “sustracción de fondo adaptativo”, el cual a diferencia del enfoque normal, toma en consideración leves cambios que pudieran presentarse en la escena debido a pequeñas variaciones en la intensidad de luz o movimientos leves como el que se produce cuando una ligera brisa agita unas cortinas. De esta forma la robustez de este método se incrementa sustancialmente. Sin embargo, este enfoque no es aplicable directamente en este sistema, ya que la fuente de video no permanece siempre estática sino que gira continuamente según sea necesario para permitir el seguimiento activo al expositor, por lo cual todos los píxeles de la imagen se encuentran en contínuo movimiento. Para resolver este problema se necesita de un modelo de representación esférica del fondo, el cual es generado durante la Fase de Inicialización del sistema.

4.2.2.1 Fase de Inicialización

Durante esta fase el modelo del fondo es construido. Debido a que el fondo es variable en el tiempo, no es posible utilizar alguna técnica de sustracción de fondo, a menos que se cuente con un modelo que cubra todo el campo de acción posible que pueda ser cubierto por el movimiento rotacional de la cámara de video en sentido horizontal. Para solucionar este problema se empleará un modelo esférico del fondo de la escena.

De esta forma, cuando la cámara se mueve, se recuperan diferentes partes del modelo esférico completo para realizar la sustracción de fondo y por consiguiente detectar la posición de los objetos móviles. No obstante, es necesario determinar el punto exacto al que la cámara está enfocándose en un determinado instante, es decir, el mapeo entre los píxeles de la imagen actual y los píxeles del modelo del fondo. Una solución podría ser recuperar los parámetros de movimiento de la cámara en tiempo real, pero puesto que la cámara no es estática pueden presentarse retardos imprevistos en la comunicación y por tanto no sería posible conocer el grado de rotación de la cámara para una imagen, en un momento dado mientras la cámara se mueve.

La solución planteada para este problema consiste en registrar cada imagen en el modelo esférico del fondo y de esta forma inferir los parámetros de movimiento en tiempo real, aún cuando la fuente de video se encuentre en movimiento.

Considerando esto, la problemática se reduce a obtener el modelo esférico del fondo. Para construir este modelo se recolecta un conjunto de imágenes con parámetros de giro establecidos previamente, de tal manera que sea cubierta toda la escena, la cual podría ser el escenario y sus alrededores. De hecho, se pretende que este modelo del fondo sea obtenido en base a parámetros de configuración del sistema. Este modelo esférico podría entonces ser construido utilizando la técnica conocida como “stitching” para generar un mosaico esférico o cilíndrico. Sin embargo, teniendo en consideración que no se necesita de un gran campo de visión para cubrir la totalidad del escenario, por eficiencia este sistema utiliza la colección de imágenes directamente y para determinar que parte del modelo del fondo se debe recuperar para cada frame de la secuencia, se requiere de un patrón de movimiento de la cámara que permita conocer su ubicación en todo momento.

Este patrón es implementado por un vector que almacena los parámetros pre-establecidos de movimiento de la cámara y que en todo momento conoce la posición actual de la misma.

El modelo del fondo básicamente es una matriz de imágenes, en la cual cada elemento representa una posición posible que la cámara puede adoptar. En este sistema, esta matriz es unidimensional debido a los 2 grados de libertad de la fuente de video. Por consiguiente, es necesario estimar cada parte del modelo del fondo, es decir, cada elemento de esta matriz.

Para obtener cada fragmento del modelo del fondo se podría utilizar la primera imagen obtenida en cada posición posible de la cámara. Sin embargo, no es posible asegurar ausencia de movimiento en el instante inicial. Por este motivo, se ha empleado un parámetro nmax, que especifica el número máximo de frames que deben ser considerados para generar el fragmento. Durante este tiempo es muy recomendable la ausencia de movimiento.

La idea básicamente consiste en tomar como fondo la media aritmética de los frames capturados hasta el frame nmax, de forma que:

$$F(n,x,y) = \left\{ \frac{n-1}{n} F(n-1,x,y) + \frac{1}{n} I(n,x,y) \qquad n < n_{\max} \right\}$$

En esta función, F es el fragmento del fondo, n es el frame actual, I es la imagen en el frame n, y nmax es el último frame que interviene en la estimación del fragmento.

Una vez que ha terminado la fase de inicialización y el modelo del fondo ha sido obtenido, se procede a la detección de los objetos de interés en la escena mediante sustracción de fondo.

4.2.2.2 Sustracción de Fondo

El objetivo principal del Módulo de Reconocimiento es la detección y reconocimiento de los objetos móviles de la escena. Su función básica consiste en separar estos objetos de un fondo en constante movimiento. Una vez que el modelo de dicho fondo haya sido construido, es posible comenzar la ejecución de esta tarea.

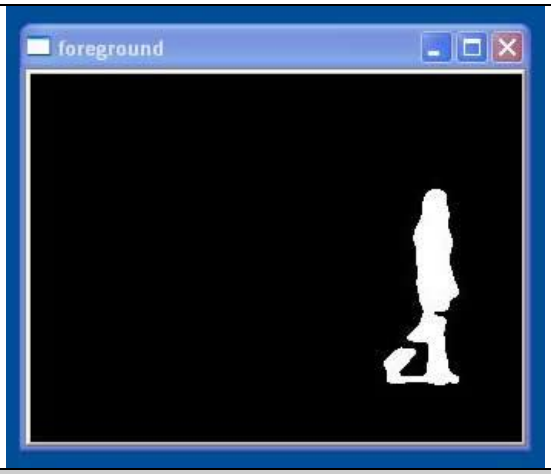

Figura 4.2. Sustracción de Fondo.

El primer paso consiste en determinar el fragmento de fondo que deberá ser utilizado para realizar la sustracción. Para esto es necesario conocer los parámetros de rotación de la cámara en un momento dado, los cuales se encuentran definidos en el vector de parámetros que se mencionó en el apartado 4.2.2.1.

La extracción de los objetos entonces, básicamente es efectuada por comparación de cada frame capturado con el fragmento del fondo vigente, siguiendo el siguiente esquema:

$$Dif_n = | I_n - F_n |$$

De esta forma, después de realizar la sustracción de fondo, el resultado es una imagen que contiene los objetos detectados y por tanto es necesario realizar una segmentación de esta imagen para determinar la posición del expositor.

4.2.2.3 Segmentación de Objetos

Según se estudió en la sección 2.3.2 existen varios niveles de segmentación que podrían ser utilizados. En este sistema se han empleado técnicas de bajo nivel asociadas a la intensidad lumínica de píxel para determinar la ubicación (en la imagen) de los objetos del primer plano.

Para realizar la segmentación dos fases son necesarias. La primera consiste en efectuar una segmentación básica orientada a píxel utilizando un umbral. En esta fase, los objetos son separados del fondo asociando cada píxel de la imagen a uno de los dos grupos posibles: fondo o primer plano, comparando su intensidad en contra del umbral establecido.

La segunda fase de segmentación es orientada a región y consiste en buscar grupos conexos de píxeles pertenecientes al primer plano (obtenidos en la fase anterior) y asigna una etiqueta que diferenciará a cada objeto. Esta etiqueta es un valor de intensidad definido por el sistema. Esta conectividad entre píxeles es definida empleando la “8-conectividad”.

La “8-conectividad” define 8 vecinos para un píxel p cuya posición en la imagen es (x, y). Entonces los vecinos de p son aquellos que se encuentran en las posiciones (x-1, y-1), (x, y-1), (x+1, y-1), (x-1, y), (x+1, y), (x+1, y-1), (x+1, y), (x+1, y+1).

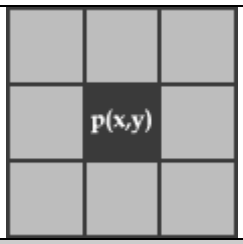

Figura 4.3. Píxeles vecinos de p.

Para realizar el proceso de segmentación y determinar la conectividad se definen como vecinos causales de p(x, y) a aquellos píxeles que se encuentran en las posiciones (x-1, y-1), (x, y-1), (x+1, y-1), (x-1, y). Con este conocimiento entonces, la segmentación se realiza de la siguiente manera:

1. Tomando como base la esquina inferior izquierda y en sentido ascendente, se realiza un barrido de izquierda a derecha de todos los píxeles de la imagen.

Figura 4.4. Barrido de izquierda a derecha.

2. Conforme avanza el barrido, para cada píxel no perteneciente al fondo se analizan sus vecinos causales. Tres casos son los posibles:
    - Todos sus vecinos causales pertenecen al fondo y por tanto el píxel actual corresponde a un objeto nuevo.
    - Uno o varios de sus vecinos tienen una misma etiqueta (valor de intensidad). En este caso se asigna dicha etiqueta al píxel actual.
    - Por lo menos dos vecinos causales que no pertenecen al fondo tienen diferente etiqueta. Esto ocurre cuando un mismo objeto se encuentra dividido en fragmentos con diferentes etiquetas. En este caso, ambas etiquetas se establecen como equivalentes y una vez que se haya procesado toda la imagen se procede a un segundo barrido en el que se reemplazan todas las etiquetas equivalentes, por una etiqueta única.

Por último es necesario determinar que objeto corresponde al expositor. Por este motivo el sistema necesita de la restricción de que no existan objetos móviles de tamaño mayor al expositor. En otras palabras, si existen otros objetos en movimiento, estos deben

ser relativamente pequeños en comparación con el expositor. Esta acepción se justifica en que durante una presentación en un auditorio o escenario típicamente el expositor se encuentra al frente del mismo mientras la audiencia presta atención sin presentar movimiento considerable. Adicionalmente, el rango de desplazamiento del expositor es conocido; por tanto, objetos encontrados completamente fuera de ese rango podrán ser descartados sin ningún problema.

Figura 4.5. Segmentación de Objetos.

4.2.3 Módulo de Seguimiento

Cuando el expositor se aproxima a los límites del rango de desplazamiento permitido o cuando traspasa estos límites, el sistema envía comandos de rotación a la cámara de video. Para el efecto, se necesita conocer la medida en que la cámara debe girar. Esta medida se encuentra especificada por los parámetros de rotación con

los cuales el modelo esférico del fondo fue construido, y se encuentra almacenada en el vector de parámetros. Este vector mantiene los parámetros de rotación asociados con cada fragmento del modelo del fondo, necesarios para el proceso de reconocimiento (véase 4.2.2.1).

Como primer paso, este módulo recupera la posición de la persona en la imagen. Esta posición fue encontrada por el módulo de reconocimiento, el cual almacena esta posición como un ROI (región de interés). De esta manera, dicho ROI es utilizado para realizar el análisis del movimiento de la persona.

Para efectuar este análisis se utiliza un historial de movimiento, es decir, la posición de la persona es almacenada por cada frame que se procesa, para poder determinar el sentido del desplazamiento. Puesto que este historial podría crecer indefinidamente y por consiguiente incrementar el uso de memoria por cada frame que se procese, se define un tiempo de duración para el historial. De esta forma el historial descarta las posiciones que produzcan una diferencia, entre la estampa de tiempo con la que fueron registradas y la estampa de tiempo actual, que exceda al tiempo de duración establecido.

Luego de que el historial se haya actualizado, se procede a calcular el gradiente de movimiento, utilizando el historial, para determinar el sentido y dirección del desplazamiento.

Figura 4.6. Posición del objetivo encontrado. Se mantiene dentro del rango.

Finalmente se verifica si la posición de la persona se encuentra dentro del rango de desplazamiento permitido. Si la persona se encuentra fuera del rango es preciso rotar la cámara de video, para lo cual se necesita conocer el sentido del desplazamiento, el cual fue determinado en el paso previo. Luego de esto, el módulo termina su función comunicando hacia donde debe girar la cámara, en caso de

que sea necesario, o bien establece que la cámara no debe moverse en absoluto.

4.2.4 Módulo de Control de Cámara

La función principal de este módulo es la comunicación con el dispositivo de video. Cada vez que el sistema necesita mover la cámara, se establece la conexión y se envían los comandos correspondientes a la acción deseada para que la cámara los ejecute. Luego de que un comando ha sido enviado, se recibe una respuesta de éxito o error, y continúa la ejecución del sistema. Si la operación fue exitosa el dispositivo ejecuta la acción requerida y envía una notificación cuando la termina. De lo contrario se establece una condición de error, ocurrido en el control de cámara, que el sistema se encargará de manejar.

La interfaz que se emplea en este sistema es el estándar EIA RS-232C, el cual implementa la comunicación para el intercambio de datos mediante el puerto serial. Este protocolo consiste de un conector DB-25 de 25 pines, o un conector DB-9 de 9 pines, puesto que los computadores no suelen utilizar más de 9 pines en el conector DB-25. Este puerto trabaja con señales digitales, de +12V

(0 lógico) y -12V (1 lógico), para la entrada y salida de datos e inversamente para las señales de control.

PIN	FUNCION
TXD	Transmitir Datos
RXD	Recibir Datos
DTR	Terminal de Datos Listo
DSR	Equipo de Datos Listo
RTS	Solicitud de Envío
CTS	Libre para Envío
DCD	Detección de Portadora

Tabla 4.1: Funciones más importantes de los pines del conector DB-25

Las señales RXD, DSR, CTS y DCD son señales de entrada, mientras que TXD, DTR y RTS son señales de salida. Las funciones de cada una de estas señales se especifican en la Figura 4.4.

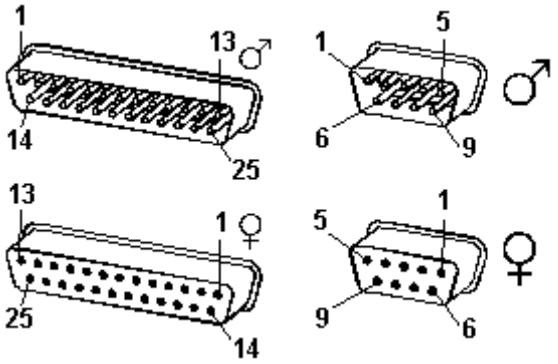

Figura 4.7. Conectores utilizados por el estándar RS-232C. A la izquierda el conector DB-25 y a la derecha el conector DB-9 [F4].

Debido a que no todos los pines del conector DB-25 son utilizados, es posible reemplazarlo por el conector DB-9. En la tabla 4.1 se especifican las señales correspondientes a cada pin y la relación entre los pines de los conectores DB-25 y DB-9.

DB-25	DB-9	SEÑAL	DESCRIPCION	E/S
1	1	-	Masa chasis	-
2	3	TxD	Transmit Data	S
3	2	RxD	Receive Data	E
4	7	RTS	Request To Send	S
5	8	CTS	Clear To Send	E
6	6	DSR	Data Set Ready	E
7	5	SG	Signal Ground	-
8	1	CD/DCD	(Data) Carrier Detect	E
15	-	TxC(*)	Transmit Clock	S
17	-	RxC(*)	Receive Clock	E
20	4	DTR	Data Terminal Ready	S
22	9	RI	Ring Indicator	E
24	-	RTxC(*)	Transmit/Receive Clock	S

Tabla 4.2: Señales asociadas a los pines de DB-25 y DB-9

4.2.4.1 Establecer la Comunicación

Para realizar el control de cámara este módulo utiliza un control ActiveX desarrollado por el fabricante del dispositivo. Este control define varios métodos que implementan cada comando reconocido por el dispositivo. El SDK provee la documentación necesaria, en la que se especifica el formato de cada comando, el formato de las respuestas y las posibles causas de error.

Los comandos utilizados por el sistema son básicamente los que controlan el movimiento rotacional de la cámara, es decir asignación

de ángulos de PAN y TILT, y obviamente los comandos de conexión a través de un puerto COM.

Adicionalmente se implementan handlers para los eventos generados por la notificación de terminación de las acciones realizadas por la cámara. Esto es necesario debido a que el control de cámara es asíncrono, es decir, el sistema no espera mientras el dispositivo lleva a cabo alguna acción para continuar con su ejecución, sino que se modifican los valores de variables de control cada vez que se produce un evento de notificación.

Para controlar el movimiento rotacional se tiene como base la posición de inicio del dispositivo (HOME), la cual esta definida en código hexadecimal por 8000h para el PAN y 8000h para el TILT. De esta manera, es posible asignar una posición determinada en base a esta posición inicial. Por ejemplo, en el caso que se necesite 90º de pan hacia la derecha y 30º de tilt hacia abajo, los valores hexadecimales que representan esta nueva posición son obtenidos de la siguiente manera:

Pan Derecha 90º

+90/0.1125 → +800 (decimal) → +320h (hexadecimal)

8000h + 320h = 8320h

Tilt Abajo 30º

-30/0.1125 → -267 (decimal) → -10Bh (hexadecimal)

8000h – 10Bh = 7EF5h

En donde el valor 0.1125 es utilizado para la conversión de ángulos y es el resultado de dividir el Radio de Angulo de Pulso de Pan y Tilt para 100000. Este radio es 11250 y es dependiente del dispositivo.

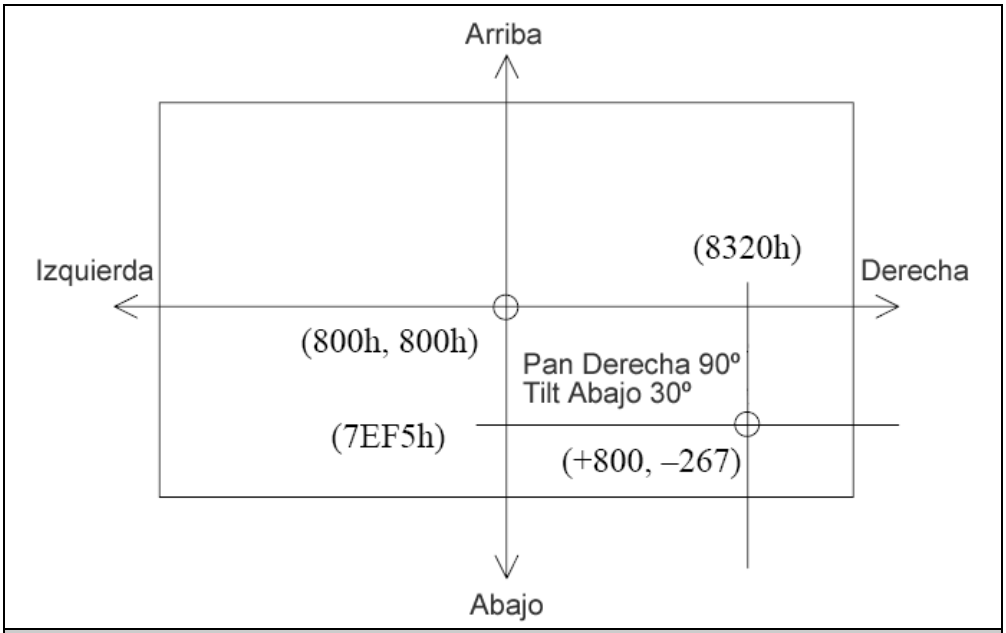

Figura 4.8. Asignación de nueva posición de la cámara.

El código de control de la asignación de ángulo de pan/tilt tiene el siguiente formato:

FFh 30h 3Xh 00h 62h p0 p1 p2 p3 p4 p5 p6 p7 EFh

En donde,

X, es el id del dispositivo. En este sistema su valor siempre será 1.

00h 62h es el código del comando de asignación de ángulo de pan/tilt,

p0 p1 p2 p3 es el ángulo de pan, y

p4 p5 p6 p7 es el ángulo de tilt.

En el ejemplo citado anteriormente (véase Figura 4.8), la nueva posición en representación hexadecimal es 8320h para pan y 7EF5h para tilt. Por tanto, el comando se enviaría de la siguiente forma:

FFh 30h 31h 00h 62h 38h 33h 32h 30h 37h 45h 46h 35h EFh

Por otra parte, durante la ejecución del sistema es necesario conocer el estatus de la cámara, es decir, si se encuentra estática, en movimiento o en alguna condición de error. Sin embargo, puesto que la comunicación con la cámara es asíncrona, el sistema no espera la respuesta sino que continúa su ejecución. Por este motivo, cada vez que se envía un comando de movimiento, internamente el sistema establece el estado de movimiento de la cámara y solamente cambia

este estado cuando recibe una notificación por parte de la cámara. Este evento de notificación entrega el estado actual del dispositivo, por lo que el handler de este evento es el encargado de verificar dicho estado y modificar las variables correspondientes para determinar si la cámara se encuentra en movimiento o si se ha detenido, o bien, que se ha producido un error en la comunicación con el dispositivo.

4.2.4.2 Detección de Fallos

Debido a que la cámara de video es un componente fundamental de este sistema, si se presenta un problema de comunicación es prácticamente innecesario continuar con la ejecución del mismo. Sin embargo, es necesario detectar estos errores, para asegurar el correcto funcionamiento del sistema.

Cabe mencionar que no es lo mismo un fallo en el módulo de control de cámara que una respuesta de error en la ejecución de algún comando, pues en el primer caso el fallo determina que el sistema no puede continuar su ejecución a menos que se recupere del mismo, mientras que en el segundo solo se determina la causa por la cual el comando no pudo ser ejecutado exitosamente, por ejemplo, al tratar

de asignar una nueva posición a la cámara mientras esta se encuentra en movimiento.

Los posibles fallos en la comunicación con el dispositivo pueden deberse a diversas causas, por ejemplo un error en la apertura del puerto COM. Sin embargo, sea cual fuere el motivo de este fallo, la incidencia es la misma, el sistema no puede continuar con el seguimiento al expositor. En este caso habrían dos posibles opciones: continuar con la ejecución del sistema para la grabación y la transmisión del video o detener completamente su ejecución. En el primer caso el sistema graba y trasmite el video registrado por la cámara, en la posición en que haya quedado el dispositivo; si fuera necesario posicionar nuevamente la cámara, se lo debe hacer manualmente o utilizando un software independiente del sistema.

Por este motivo, en presencia de un fallo de este tipo, el sistema solicita al usuario la acción correspondiente que deberá llevarse a cabo.

4.2.5 Módulo de Transmisión

Este módulo no es desarrollado como parte del software del sistema. Es más bien implementado por hardware, compartiendo la señal de

video mediante el uso de un splitter. Sin embargo, la trasmisión del video es realizada utilizando un encoder de video.

La señal proveniente del dispositivo es dirigida tanto al controlador principal del sistema como a otro puerto que es utilizado por el encoder.

De esta forma, el controlador principal del sistema, representado por la interfaz gráfica de la aplicación, se encarga del análisis de la escena para controlar la posición de la cámara, mientras que el encoder trasmite el video a través de internet.

4.3 Problemas Encontrados

Durante el desarrollo de este sistema surgieron varios problemas relacionados especialmente con el reconocimiento y seguimiento de la persona y el control de la fuente de video.

Con respecto al reconocimiento de la persona en la escena, el principal problema se presenta al permitir el movimiento rotacional de la cámara de video. Con una fuente de video estática es relativamente sencillo extraer los objetos móviles de la escena, puesto que el fondo de la misma nunca varía. Sin embargo, al tener

una fuente móvil, el fondo cambia continuamente lo cual imposibilita aplicar sustracción de fondo para identificar a la persona en la escena.

Para resolver este problema, se optó por aplicar un enfoque que permita que el fondo de la escena se adapte a los cambios. Por este motivo, existe una fase de inicialización durante la cual se construye un modelo esférico del fondo de la escena. Este modelo es esférico por la conectividad de sus componentes, los cuales son los fragmentos del fondo y son determinados por la posición de la cámara en un momento determinado, lo cual no implica que el fondo necesariamente cubra 360º de visión. Adicionalmente, para incrementar la eficiencia del sistema, este modelo del fondo es discreto en términos matemáticos, lo cual significa que está formado por un número finito de fragmentos.

Este número de fragmentos es configurable desde la interfaz gráfica del usuario y debe ser establecido antes de comenzar la fase de inicialización del sistema.

Otro problema de gran importancia surge en la implementación del módulo de control de cámara. Debido a que continuamente se

procesa la secuencia de video proveniente de la fuente, no es posible implementar una aplicación de un solo hilo de ejecución; en otras palabras, el sistema no puede esperar hasta recibir una respuesta de la cámara de video. Si el control fuese unidireccional simplemente se enviaría un comando para que sea ejecutado por la cámara y no importaría el resultado de la operación. Sin embargo, esto no es aplicable para este sistema, ya que en todo momento se debe conocer el estado del dispositivo, es decir, si se encuentra en movimiento, detenido o en alguna condición de error. El problema se complica cuando el valor de ciertas variables de control depende del resultado de una operación realizada por el dispositivo.

Por este motivo, se ha implementado la notificación de terminación de comandos por medio de eventos que son generados por el dispositivo y que son manejados por el sistema. El control ActiveX (véase 4.2.4.1) utilizado para la comunicación con el dispositivo, implementa la generación de eventos del mismo y provee lo necesario para manipular estos eventos desde la interfaz MFC. No obstante, el gran problema se presenta debido a que este control ActiveX es utilizado por la interfaz gráfica, es decir, por el hilo principal de la aplicación, y no puede ser utilizado por otros hilos de ejecución, debido a que el hilo principal tiene el control de la posición

de memoria en la cual reside el control ActiveX, y ésta no puede ser accedida por otros hilos.

Para resolver este último problema fue necesario emplear mensajes para la comunicación entre los hilos. De esta manera, cuando el hilo principal necesita controlar el dispositivo de video, envía un mensaje que es recibido por un método que se ejecuta en el mismo hilo, el cual se encarga de invocar los métodos para el control del dispositivo, definidos en el Módulo de Control de Cámara.

4.4 Sumario

La implementación del sistema ha sido presentada en este capítulo, de una manera específica, incluyendo descripciones de los enfoques y algoritmos utilizados, pero sin incurrir en demasiados detalles técnicos.

En la primera parte de este capítulo se ha realizado un breve análisis de las herramientas utilizadas para el desarrollo de la fase de implementación del sistema. Adicionalmente, se introduce la visión global de esta implementación, mencionando de manera general la estructura del sistema.

A continuación se ha presentado cada módulo del sistema, describiendo su funcionalidad y los detalles de su implementación. En cada caso, los enfoques utilizados y los algoritmos implementados han sido descritos en gran detalle.

Por último, se ha realizado un análisis de los principales problemas encontrados durante esta etapa, teniendo en consideración las causas del problema, las razones por las que no se ha podido evitar y las soluciones aplicadas a cada uno de ellos.

CAPÍTULO 5

5 PRUEBAS

5.1 Pruebas de Campo

Una vez concluida la implementación del sistema se procedió a realizar las respectivas pruebas para verificar la funcionalidad del mismo, y de esta manera determinar su rendimiento.

Para el proceso de pruebas de este sistema se utilizó un auditorio con iluminación adecuada. La cámara de video se ubica en la esquina posterior derecha a una altura aproximada de 3m. Este dispositivo se conecta al computador que ejecuta la aplicación principal, utilizando un puerto de comunicación serial (véase 2.6).

Al comenzar la ejecución de la aplicación se obtiene la posición actual de la cámara de video y se establece como punto de referencia para la posterior inicialización del modelo de fondo. No obstante, la aplicación permite calibrar la posición de la cámara en tiempo de ejecución.

Figura 5.1. Aplicación principal en ejecución.

Una vez que la aplicación ha comenzado su ejecución es necesario inicializar el modelo de fondo. Durante este proceso se define el área del escenario que será considerada para realizar el seguimiento y se construye el modelo de fondo, el cual no podrá ser modificado mientras se realiza el seguimiento. Sin embargo, este modelo es generado en base a ciertos parámetros que pueden ser especificados por el usuario. En caso de que no se modifiquen estos parámetros, la inicialización se efectúa con los valores establecidos por omisión.

.

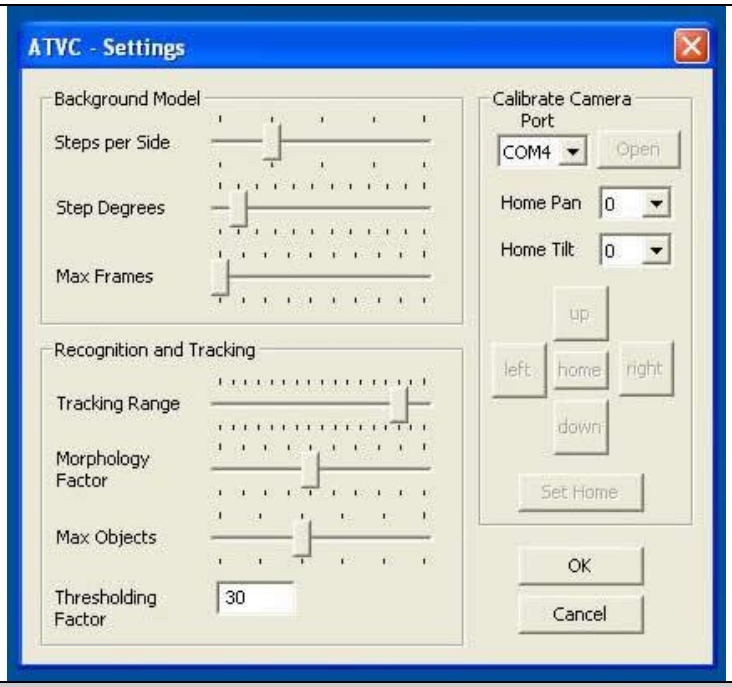

Figura 5.2. Ventana de configuración para el proceso de inicialización.

Durante la inicialización la cámara explora el área o campo de acción sobre la cual trabaja el sistema, y se genera el fondo correspondiente a cada posición que la cámara puede adoptar. El parámetro “Steps per Side” define la cantidad de posiciones permitidas para la cámara hacia cada lado en sentido horizontal y tomando como referencia la posición actual como posición central o home.

En las siguientes imágenes se aprecia como se realiza el proceso de inicialización y la obtención del modelo de fondo. Para efecto de demostración, solamente se muestra el comienzo del proceso, una etapa intermedia, y el final del mismo.

.

Figura 5.3. Comienzo del proceso de inicialización.

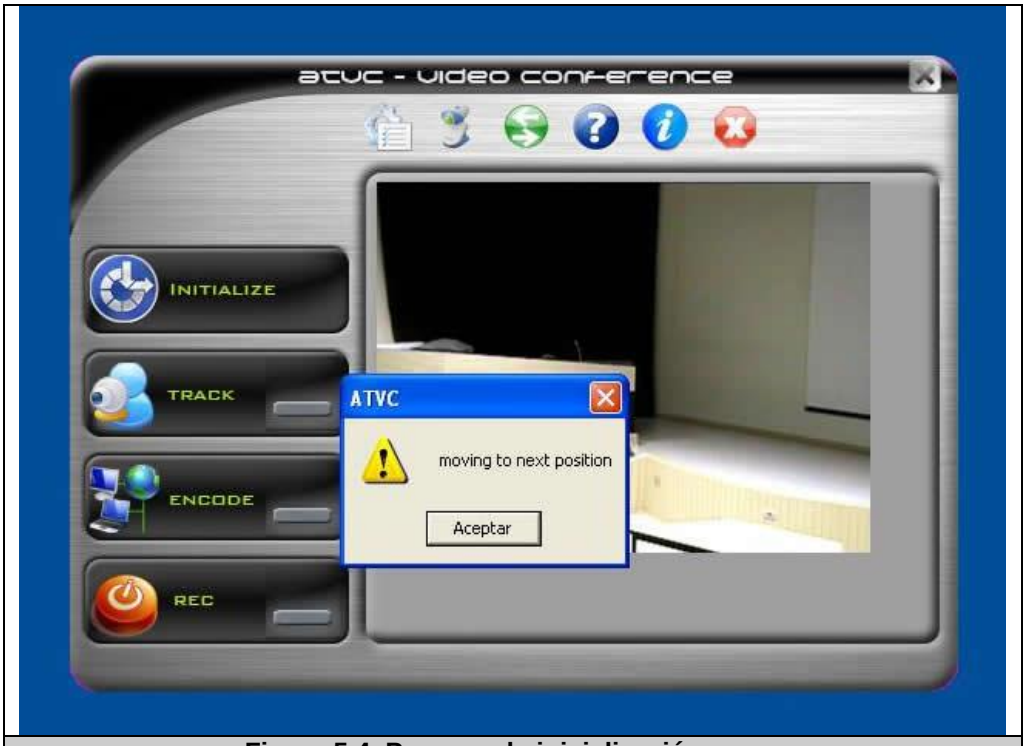

Figura 5.4. Proceso de inicialización en curso.

Figura 5.5. Proceso de inicialización finalizado.

Cuando el modelo de fondo ha sido construido, los parámetros de inicialización de la ventana de configuración son deshabilitados. Sin embargo, es posible cambiar el valor de los parámetros de reconocimiento y seguimiento, mientras este último se encuentra activo.

Figura 5.6-a. Seguimiento a la persona. Pantalla principal.

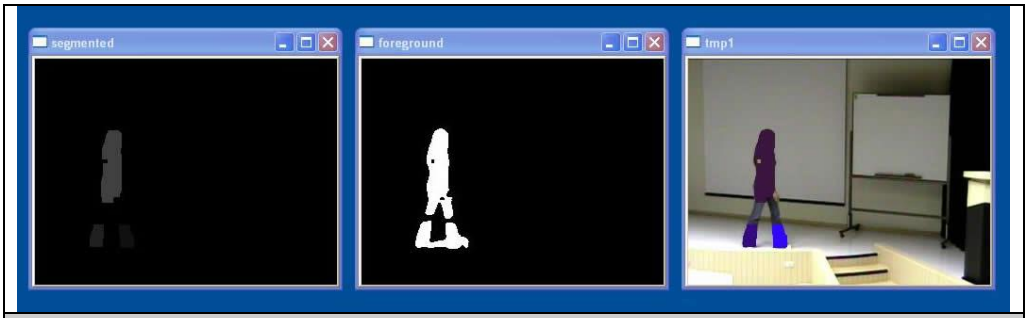

Figura 5.6-b. Seguimiento a la persona. Pantallas demostrativas del proceso de reconocimiento.

El seguimiento puede ser activado o desactivado mediante el botón “Track”. En las figuras 5.6-a y 5.6-b se puede apreciar el seguimiento activo. En la primera se nota claramente como la persona ha sido detectada y en la segunda se muestran las etapas del proceso de

reconocimiento de los objetos móviles, que en este caso corresponde a la persona.

Dependiendo del parámetro “Tracking Range”, el sistema decide si es necesario o no mover la cámara para recuperar el enfoque. En las siguientes figuras se muestra como se produce este seguimiento.

Figura 5.7. Persona próxima al límite izquierdo.

Figura 5.8. La persona ha sobrepasado el límite izquierdo. La cámara se mueve horizontalmente hacia la izquierda para recuperar el enfoque.

Adicionalmente es posible grabar el video que es registrado por la cámara. La aplicación permite al usuario escoger un códec para la compresión del video.

Figura 5.9. Selección del códec de compresión de video.

Figura 5.10. Grabación de video.

Al presionar el botón “REC” comienza la grabación de video la cual almacena el flujo contínuo de video en un archivo generado utilizando el códec de compresión escogido, hasta que se vuelva a presionar el mismo botón para detener la grabación.

5.2 Análisis de Resultados

El sistema ha demostrado funcionar correctamente bajo las restricciones requeridas (véase 2.1.1.2). A continuación se realiza un breve análisis de cada tarea principal y del rendimiento en general del sistema.

- **Interfaz del Usuario**

  La interfaz de la aplicación principal provee controles que permiten manipular de manera fácil y sencilla las diferentes funciones del sistema. Se trata de una interfaz muy amigable e intuitiva que muestra en primera instancia los necesarios para ejecutar la función principal del sistema, que es el seguimiento automatizado al expositor. Sin embargo, existen opciones avanzadas de configuración que permiten que ésta tarea sea realizada de manera aún más precisa, ya que es posible calibrar los parámetros necesarios de acuerdo a las condiciones del lugar.

Por otra parte se provee de retroalimentación constante al usuario, para que éste pueda tener conocimiento del estado actual del procesamiento y de esta forma que pueda constatar el correcto funcionamiento del mismo.

- **Control de Cámara**

  La aplicación principal controla efectivamente el movimiento del dispositivo de captura video. Posee completo control sobre el mismo, lo cual permite conocer su estado en todo momento. En otras palabras, cuando se requiera mover la cámara, será posible siempre y cuando no existan conflictos causados por otras aplicaciones o por factores externos como problemas o impedimentos de carácter mecánico.

  No obstante, de producirse algún inconveniente en la comunicación con el dispositivo, el sistema alerta al usuario acerca del inconveniente.

- **Reconocimiento y Seguimiento**

  El sistema ha sido probado en un único lugar, un auditorio elegido que provee las facilidades necesarias respecto al hardware del sistema, específicamente la cámara de video.

Puesto que este auditorio tiene completo control sobre la iluminación, no se han presentado problemas en variaciones abruptas de intensidad. Durante la inicialización del modelo de fondo, las pequeñas variaciones en la intensidad lumínica son consideradas para generar cada segmento del fondo.

Una vez obtenido el modelo de fondo, se puede proceder al seguimiento. En esta fase de ejecución del sistema, es posible constatar el estado del reconocimiento de los objetos móviles en tiempo de ejecución. De esta manera el usuario puede advertir si el sistema funciona correctamente, es decir, si la cámara se mueve cuando debería moverse.

Durante las pruebas realizadas, una persona se encontraba en el escenario del auditorio siendo enfocada por la cámara de video, la cual efectivamente seguía a dicha persona cada vez que ésta se alejaba del campo de visión actual definido por los parámetros de configuración de la aplicación principal.

- **Grabación**

Independientemente del seguimiento, el sistema almacena el flujo de video en el disco duro del computador que ejecuta la aplicación principal.

En efecto, el archivo de video se almacena en el disco C, bajo el nombre atvc_video utilizando el códec de compresión elegido.

5.3 Conclusiones de las Pruebas

En base a las pruebas efectuadas al sistema se puede afirmar que éste cumple sus funciones de manera efectiva. No se presentaron problemas en la manipulación de la cámara, pues el sistema tiene completo control sobre aquel dispositivo.

Puesto que durante el seguimiento se pueden observar las etapas del reconocimiento, es posible verificar tanto los objetos de interés, es decir los objetos en movimiento, como la identificación de la persona que también es clasificada como un objeto móvil. No obstante, el cuerpo de la persona no es reconocido completamente, sino que el algoritmo de segmentación clasifica el cuerpo humano como diferentes objetos móviles cercanos y toma a uno solo de ellos para realizar el seguimiento. Esto se debe a que la persona puede

presentar contraste variable provocado por su vestimenta; es decir, si existe un contraste alto entre su vestimenta, pues es muy probable que el sistema identifique varias partes del cuerpo humano como objetos independientes. Sin embargo, en la mayoría de los casos el sistema elegirá como objetivo al torso o a las extremidades inferiores de la persona.

Por otra parte, la grabación de video se realiza correctamente. Cabe mencionar que la calidad del archivo de video no es inherente al sistema, ya que se utiliza un códec de video elegido por el usuario de entre aquellos disponibles en el computador que ejecuta la aplicación principal.

Adicionalmente, durante las pruebas se ha podido constatar el bajo consumo de recursos de memoria que presenta el sistema, verificando de esta forma la optimización de memoria efectuada durante la implementación del mismo. En las pruebas se ha encontrado que el consumo de memoria no sobrepasa los 20Mb, lo cual es relativamente bajo, considerando que se trata de procesamiento de video en tiempo real.

En términos generales, el sistema cumple efectivamente con su funcionalidad y permite ajustar ciertos parámetros tanto para el proceso de inicialización, el control de cámara, reconocimiento y seguimiento, para de esta manera lograr un mejor rendimiento.

CAPÍTULO 6

6 APLICACIONES

6.1 Aplicaciones Generales del Sistema

Este sistema ha sido desarrollado fundamentalmente con el objetivo de demostrar la utilidad del reconocimiento y seguimiento de objetos y personas que interactúan en un ambiente determinado.

Los sistemas de seguimiento pueden ser utilizados en diversas aplicaciones, tales como en la detección de objetos robados y detección de situaciones sospechosas; en las interfaces hombre-máquina con aplicaciones especificas como la captura de movimiento utilizada en el desarrollo de videojuegos, simuladores virtuales y producciones cinematográficas. Incluso se pueden utilizar para aplicaciones más avanzadas que incluyen análisis del movimiento humano para obtener información de utilidad para análisis del rendimiento de deportistas o para el reconocimiento de gestos y posturas.

En realidad, los datos que se generan en el proceso de seguimiento de objetos contienen información de gran utilidad, especialmente, cuando se trata de analizar comportamientos predefinidos, para posteriormente determinar el motivo que los produce.

La aplicación principal de este sistema es servir como una herramienta útil para la presentación de exposiciones y charlas desde una ubicación remota mediante video-conferencia, en la cual no existe la necesidad de un operador que manipule la cámara de video. Como una utilidad adicional, el sistema es capaz de grabar la señal de video y almacenarla en el disco duro del computador.

No obstante, otra aplicación de este sistema podría ser la detección de objetos abandonados. Adaptar el sistema para que detecte la presencia de objetos abandonados en lugares tales como, por ejemplo, la sala de espera de una estación de autobuses o una terminal aérea. Esto puede ayudar a prevenir situaciones peligrosas alertando a tiempo a un supervisor antes de que estas se produzcan.

Cabe mencionar que, aunque el sistema es capaz de identificar al expositor, en presencia de movimiento que no corresponde a dicha persona y tomando en consideración que la cámara de video se

encuentra en una posición frontal al expositor, usualmente en la parte posterior del auditorio, la efectividad de reconocimiento puede reducir debido a la oclusión y a la superposición de objetos, por lo cual es recomendable que el movimiento pertenezca, en lo posible, únicamente al expositor.

6.2 Utilidad de los Módulos

El sistema ha sido implementado en varios módulos con funcionalidad específica, los mismos que pueden entenderse como capas independientes creadas para trabajar en conjunto y que se encargan de distintas tareas.

Tanto para la aplicación en que se enfoca el sistema, como para otras en las que se podría utilizar, la idea de base es la modularidad. Gracias a esta, es posible desarrollar aplicaciones distintas, más avanzadas y robustas, intercambiando o combinando diferentes módulos.

Cada uno de estos módulos puede tener aplicaciones adicionales e incluso podrían ser utilizados como componentes de otros proyectos. No pueden ser considerados un “framework” o “marco de trabajo”, pero pueden acoplarse perfectamente en otros sistemas, pues estos

módulos trabajan sobre una entrada recibida; esto es, aplican todo el procesamiento necesario a los datos de entrada y generan un resultado.

Los casos más notables corresponden a los módulos de reconocimiento y seguimiento. En el primer caso, el módulo recibe una imagen que contiene información completa de la escena y genera una imagen binaria que contiene únicamente a la persona identificada y adicionalmente genera una “región de interés”. En el segundo caso, el módulo de seguimiento recibe la imagen original de la escena y la región de interés que fue generada por el módulo de reconocimiento. De esta manera, el seguimiento se efectúa a dicha región y, en caso de ser necesario, se procesa la imagen original para visualizar la detección de la persona en la escena.

Cabe mencionar que los módulos han sido implementados de tal manera que sea posible reemplazar los principales algoritmos de cada uno de ellos, para mejorar su funcionalidad e incrementar su eficacia.

6.2.1 Posibles Aplicaciones de los Módulos

La implementación independiente de cada módulo permite que sea posible su utilización en aplicaciones diferentes de aquellas para las que fueron desarrollados en principio.

El módulo de Control de Cámara, por ejemplo, podría ser utilizado en un sistema completamente diferente y que no necesariamente sea de visión por computador. Podría ser utilizado en conjunto con el Módulo de Transmisión de Video para realizar una aplicación de video-conferencia sencilla, que permita operar manualmente el movimiento de la cámara desde una interfaz fácil de utilizar. Esta interfaz podría desarrollarse en MFC y de una manera fácil y sencilla podría acoplarse el módulo de Control de Cámara desarrollado.

El Módulo de Reconocimiento tiene gran utilidad para la detección de objetos móviles. Una aplicación particular en el que podría ser utilizado este módulo es un sistema de vigilancia, por ejemplo, en un museo. No sería necesario detectar al delincuente sino solamente la presencia o ausencia de objetos que deberían, o no, permanecer en un lugar específico.

Una aplicación posible para el Módulo de Seguimiento es la detección de situaciones sospechosas en un ambiente de gran

afluencia de público como en estaciones de buses o terminales aéreas. Se podrían definir patrones de movimientos sospechosos, por ejemplo cambios drásticos en la posición de un objeto, y de esta manera detectar una situación de peligro.

En conclusión, estos módulos pueden ser utilizados en otros sistemas de manera efectiva. En ciertos casos podría ser necesario reemplazar ciertos algoritmos, como el de segmentación en el Módulo de Reconocimiento por ejemplo, pero aun así serian fácilmente adaptados a un nuevo sistema.

6.3 Otras Aplicaciones

Existen otros tipos de aplicaciones en el que los sistemas de video inteligentes son utilizados. Se trata de aplicaciones avanzadas en las que el sistema de visión por computador es solo un componente del sistema completo.

Estos sistemas pueden servir para el desarrollo de Interfaces Hombre-Máquina utilizadas en ambientes de “computación diseminada” (pervasive computing) o también conocida como “computación ubicua”. De esta manera, utilizando algoritmos de reconocimiento de gestos, posturas del ser humano y movimiento de

extremidades, es posible crear ambientes en los que los usuarios puedan satisfacer sus necesidades sin la necesidad de interactuar físicamente con el sistema.

Por ejemplo, este sistema que se ha desarrollado podría ser adaptado para crear una sala inteligente en la que se identifique a una persona y de esta forma el sistema responda a las posibles necesidades que ésta podría tener, como encender las luces, encender un computador o un proyector, o bien apagar las luces cuando la persona haya salido de la sala.

De hecho el sistema podría reconocer de quien se trata, es decir, no solo identificar una persona sino saber quién es esa persona y tomar decisiones en base a esa persona. Por ejemplo se podría distinguir entre un profesor, un expositor, la persona encargada de la sala y hasta invitados. Incluso se podría implementar el reconocimiento de acciones humanas predefinidas. Es decir, análisis del comportamiento humano en una secuencia de vídeo para detectar automáticamente eventos relevantes y llegar a concluir acerca de lo que la persona se encuentre haciendo.

Una aplicación diferente es la extracción de trayectorias de vehículos, para grabar y analizar comportamientos peligrosos en la carretera.

Figura 6.1. Extracción de trayectorias de vehículos.

Otra aplicación puede ser el control de un robot mediante la combinación de un sistema de seguimiento de objetos y un procesador de señales digitales (DSP) que utilice su salida para determinar características de alto nivel, como posición, orientación, volumen, centro y rotación del objeto.

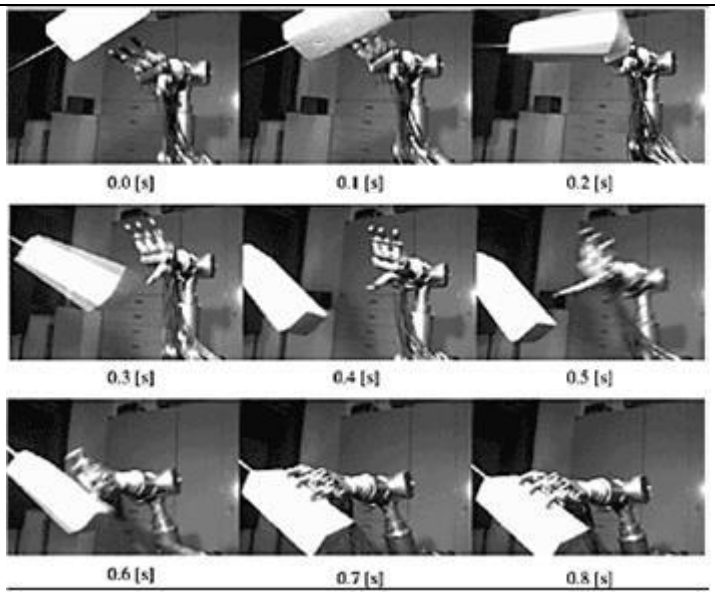

Figura 6.2. Control de brazo robótico.

Otra aplicación es la conducción de vehículos asistida para confrontar los problemas producidos por el incremento en el tráfico en carreteras, contaminación, congestión y seguridad [9]. De esta manera se ayuda a los conductores a prevenir accidentes y a aliviar el congestionamiento vehicular. En la actualidad, diversos dispositivos para vehículos inteligentes son desarrollados por los fabricantes de automotores y por laboratorios de investigación. Entre ellos se pueden mencionar: dispositivos “Stop and Go”, para detener y poner en marcha vehículos en congestión; control de velocidad, para regular automáticamente la velocidad del vehículo para mantener distancias seguras entre vehículos cercanos; sensores de

advertencia de colisiones, para alertar al conductor cuando se detectan riesgos de accidentes, entre otros.

Figura 6.3. Seguimiento a un vehículo en carretera [F5].

Por otra parte, existen proyectos de computación diseminada o ubicua (pervasive o ubiquitous), que se basan en sistemas de visión por computador. Estos sistemas tienden a ser centrados en el ser humano (human-centered), ya que una persona puede ser el objetivo de un sistema basado en visión o bien puede portar pequeños dispositivos que permitan implementar las interfaces hombre-máquina.

Un sistema de diseñado para monitorear la actividad física de atletas en un gimnasio es el Cyber Squatter [38]. Este proyecto se basa en visión por computador y otras técnicas de computación diseminada para monitorear individualmente a los atletas mientras realizan ejercicios de resistencia con pesas. El sistema hace un seguimiento del rendimiento de cada atleta utilizando diversas medidas que serian

imposibles de retener para un entrenador humano y que son registradas para su posterior análisis.

Por último, se introduce el proyecto AIMS [39], el cual se enfoca principalmente en investigar el uso de visión por computador para analizar la interacción entre objetos y seres humanos en espacios o ambientes inteligentes. De esta forma puede determinar si una persona está en pie o sentada, recogiendo algún objeto o si lo ha soltado. Además el proyecto busca determinar como puede ser utilizada esta tecnología en conjunto con otros tipos de sensores para adquirir información contextual proveniente del ambiente.

CONCLUSIONES Y RECOMENDACIONES

CONCLUSIONES

1. Desde el inicio de este trabajo se establecieron objetivos claros y concisos, los cuales no solamente se ven reflejados en el producto final sino en el desarrollo mismo de este trabajo. Básicamente se pretendía analizar la viabilidad de llevar a cabo un sistema de reconocimiento y seguimiento, enfocado a una aplicación especifica en la que el objeto de interés sería una persona que realiza exposiciones o brinda conferencias en un auditorio, y estimar la precisión que puede llegar a tener debido a los algoritmos utilizados y considerando los supuestos y restricciones que se debían tomar en cuenta para el correcto funcionamiento del mismo.

2. Se han introducido los sistemas de video inteligentes, los cuales tienen un gran interés por parte de instituciones científicas, comerciales e industriales, motivado por su gran capacidad de ser utilizados en un amplio espectro de aplicaciones que van desde la detección de objetos robados y de situaciones sospechosas, la captura de movimiento útil para la producción cinematográfica y el desarrollo de videojuegos, hasta el análisis del movimiento humano para obtener información semántica. En adición, se ha presentado un amplio análisis de los diferentes enfoques y algoritmos que

existen en la actualidad para desarrollar aplicaciones basadas esencialmente en el reconocimiento y el seguimiento de objetos.

3. Se han resaltado las principales diferencias entre los diversos enfoques, como en el caso del tipo de fuente (monocular y múltiples fuentes), y las implicaciones que conlleva utilizar uno u otro enfoque. De esta manera se ha podido realizar un breve análisis de las ventajas y desventajas de utilizar diferentes enfoques técnicos, considerando la aplicación principal que tendría el sistema y las limitaciones que conlleva desarrollar un sistema de esta categoría, según lo plante uno de los objetivos de este trabajo.

4. El principal inconveniente, lo cual probablemente demuestra lo ambicioso de este trabajo, fue que se necesitaba implementar un diseño modularizado para la solución del problema, en el que el procesamiento de un módulo sirviese de datos de entrada a posteriores módulos, constituyendo de esta forma una arquitectura flexible y fácil de ampliar, lo que permite el mejoramiento por partes del sistema. Adicionalmente, los módulos de procesamiento de secuencias de vídeo necesarios para extraer los datos con los que el sistema pudiera trabajar, tuvieron que ser implementados adaptando ciertos enfoques, en muchos casos analizando cada

imagen a nivel de pixel, debido a que los algoritmos implementados por la mayoría de librerías disponibles (incluyendo la utilizada), no brindaban el control necesario sobre el procesamiento que efectuaban y en otros casos incrementaban sustancialmente el coste computacional y el uso de recursos de memoria.

5. Los algoritmos de sustracción de fondo y de estimación del modelo de fondo, así como el algoritmo de segmentación de imagen para la clasificación de los objetos móviles, tuvieron que ser implementados.

6. Durante la etapa de implementación se probaron dos diferentes algoritmos para la segmentación de imágenes. Uno de estos había que implementarlo completamente, mientras que el segundo estaba implementado por la librería utilizada. Sin embargo, en base a los resultados obtenidos durante el desarrollo del sistema, se llegó a la conclusión de que era mejor utilizar el primer algoritmo, ya que demostró ser más efectivo para el proceso de reconocimiento.

7. Mediante la etapa de pruebas realizada se ha podido constatar la facilidad de uso de la interfaz grafica que el sistema presenta al usuario, la misma que le permite tener control sobre el procesamiento que se esconde detrás de dicha interfaz. De esta

manera se ha obtenido un sistema configurable para diversos ambientes de interiores en los que se puedan llevar a cabo exposiciones, presentaciones y eventos similares. Sin embargo, aunque es posible configurar de manera sencilla los parámetros bajo los cuales se debe realizar el procesamiento necesario para el reconocimiento y seguimiento, un usuario que posea conocimiento de procesamiento de imágenes, podría obtener mejores resultados en el seguimiento, ya que podría entender la influencia que tienen los parámetros de configuración en mención.

8. En cuanto al control del dispositivo de video también se encontraron inconvenientes a nivel de implementación. El principal problema radicó en que este dispositivo provee un componente externo que permite que el control del mismo sea viable, y cuya implementación es completamente ajena al proyecto desarrollado. Básicamente los problemas fueron dos: sincronizar la ejecución de comandos del dispositivo con la aplicación principal y la imposibilidad de utilizar el control por hilos diferentes de ejecución. Sin embargo, luego de realizar un exhaustivo estudio a cada problema se obtuvieron soluciones satisfactorias, por lo que se pudo continuar con la implementación del sistema.

9. Los objetivos principales del proyecto han quedado suficientemente cubiertos, lo cual está reflejado en los resultados que ofrece la implementación final del sistema, el cual es capaz de seguir al expositor de manera efectiva, siempre y cuando se respeten las restricciones y se tengan en cuentas las limitaciones del mismo.

10. En cuanto a los módulos que conforman el sistema, como se ha expuesto a lo largo de los capítulos de este trabajo, se puede afirmar que están suficientemente probados y ofrecen buenos resultados, sin olvidar que durante su implementación se ha buscado una funcionalidad básica para dar soporte a la aplicación para la cual han sido desarrollados, sin las pretensiones de las investigaciones exhaustivas que persiguen gran robustez en cada uno de ellos.

11. Los resultados obtenidos permiten afirmar que el sistema es una herramienta potente, que puede explotarse en el ámbito académico, sirviendo como motivación para que los estudiantes se interesen cada vez más en esta área de la multimedia, y en el ámbito científico, para que se pueda seguir investigando para desarrollar nuevas y mejores versiones que completen y perfeccionen esta primera versión del sistema.

12. Las debilidades del sistema y los posibles métodos para perfeccionar el sistema son bien conocidos. Esta situación conlleva a plantear mejoras sobre cada uno de los módulos implementados, teniendo en mente dos propósitos fundamentales: mejorar el sistema de tal forma que presente un óptimo funcionamiento en diferentes ambientes, y el segundo, lograr que sean desarrollados proyectos de mayor grado de complejidad que puedan utilizar como base los módulos implementados.

13. En la sección 6.3 se han introducido varios proyectos reales que incluyen el desarrollo de sistemas de mayor complejidad y con enfoques aún más ambiciosos. Estos sistemas constituyen ejemplos de lo que se podría desarrollar utilizando como base el sistema que ha sido desarrollado en este trabajo.

14. Finalmente, cabe destacar que desde el principio de este trabajo con el planteamiento inicial del problema, los diversos enfoques y algoritmos, las herramientas de desarrollo disponibles, etc., hasta la culminación del proyecto se ha recorriendo un largo camino en el que constantemente se aprende, y de una u otra forma el

conocimiento adquirido logra motivar e incrementar el interés por este campo de la computación científica.

APÉNDICES

A APÉNDICE A: MANUAL DEL USUARIO

A.1 Manual

La aplicación principal del sistema presenta una interfaz gráfica de usuario muy sencilla y fácil de utilizar. Los controles de la interfaz representan de alguna manera sus respectivas funciones y proveen de retroalimentación al usuario. En seguida se presenta una visión general de la interfaz, proseguido de un breve detalle de la función de cada uno de los controles.

Figura A.1.1 Interfaz del Usuario.

En la interfaz se pueden notar claramente dos grupos de controles y una pantalla que muestra la secuencia de video que es registrada por la cámara.

En el primer grupo, de lado izquierdo, se encuentran los controles de las tareas principales del sistema. En cambio, en el grupo de controles ubicados de manera horizontal en la parte superior, se encuentran básicamente las opciones de configuración y otras opciones comunes en los sistemas de software.

A continuación se detalla la funcionalidad de cada uno de estos controles:

Su función es la de dar comienzo a la inicialización del modelo de fondo, necesario para activar el seguimiento al objetivo. La inicialización se rige bajo los parámetros establecidos en la ventana de configuración. Una vez que se haya presionado este botón, la inicializción da comienzo, lo cual implica que la cámara empieza a moverse desde el extremo izquierdo hacia el derecho y en cada paso se avisa al usuario que se avanza hacia el siguiente paso.

Activa o desactiva el seguimiento al objetivo. Si previamente no se ha inicializado el modelo de fondo, un mensaje de alerta le comunica al usuario que debe proceder a la inicialización antes de activar el seguimiento. Adicionalmente, provee un led de color verde que indica el estado actual de este control.

Activa o desactiva la transmisión de video a través de internet. Es completamente indenpendiente del seguimiento y por tanto no es necesaria la inicialización previa para comenzar la transmisión de video. Adicionalmente, provee un led de color verde que indica el estado actual de este control.

Sirve para comenzar o detener la grabación de la señal de video actual presentada a través de la pantalla. Al presionar este botón, una ventana de selección del códec de video a utilizar es mostrada al usuario. Adicionalmente, provee un led de color verde que indica el estado actual de este control.

Muestra la ventana de configuración del sistema. En ésta, varios parámetros pueden ser manipulados dependiendo de las condiciones del lugar y de las necesidades del usuario. La ventana de configuración es la que se muestra a continuación.

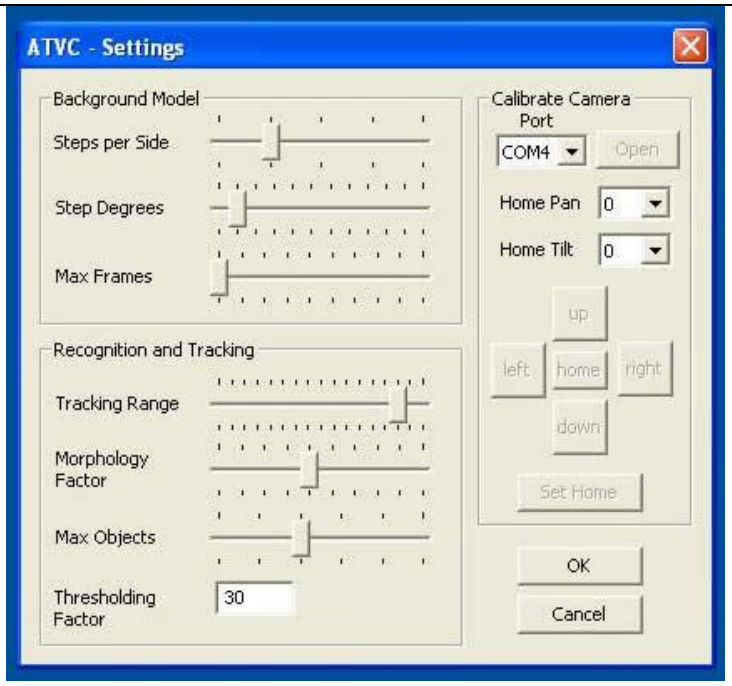

Figura A.1.2 Ventana de configuración.

Como se puede apreciar en el gráfico, existen tres grupos de parámetros. Estos son rigen la inicialización del modelo de fondo, la calibración inicial de la cámara y el reconocimiento y seguimiento.

Modelo de Fondo (Background Model)

Steps per side.- Se refiere a la cantidad de pasos por cada lado, izquierdo y derecho, que se tomarán para la generación del modelo, sin contar con la posición central. Esto significa que si este parámetro tiene un valor de 3, la cantidad de pasos será igual a (3 x 2) + 1 = 7.

Step Degrees.- Define el ángulo de rotación que la cámara debe girar para trasladarse de un paso al siguiente. En otras palabras, representa el grado de separación entre las diferentes posiciones de la cámara.

Max Frames.- Indica el máximo número de frames de la secuencia de video, que serán utilizados para generar cada segmento del modelo de fondo. Su valor es 100 por omisión.

Reconocimiento y Seguimiento (Recognition and Tracking)

Tracking Area.- Define el área permitido para el desplazamiento del objetivo. Cuando este sobrepasa los limites de esta área, la cámara se mueve para recuperar el enfoque a dicho objetivo.

Morphology Factor.- Este parámetro es utilizado para la segmentación y la clasificación de los objetos móviles. Tiene un efecto de suavizado de las regiones identificadas como posibles objetos, por lo cual permite incrementar o disminuir el número de objetos en la escena.

Max Objects.- Indica el número máximo de objetos móviles que se podrán identificar en la escena. Cabe recalcar que la percepción de un objeto por

parte del usuario, no es la misma que la del sistema. Por lo cual, un número alto de objetos podría beneficiar la precisión del reconocimiento.

Thresholding Factor.- Este es el factor de umbralización utilizado para realizar la sustracción de fondo, es decir, obtener los objetos móviles de la escena. Este depende de las condiciones de iluminación del lugar, y varía en un rango de 0 a 255.

Calibración de Cámara (Calibrate Camera)

Port.- Es el puerto serial en el que se encuentra conectada la cámara al computador.

Home Pan.- Permite ajustar la posición de la cámara en sentido horizontal.

Home Tilt.- Permite ajustar la posición de la cámara en sentido vertical.

Muestra la ventana de configuración de la fuente de video. Adicionalmente permite calibrar ciertos parámetros propios del dispositivo de video.


Permite eliminar la configuración actual del seguimiento y el modelo de fondo. Por tanto, será necesaria una nueva inicialización para activar nuevamente el seguimiento.


Muestra la ventana de ayuda al usuario.


Muestra información general acerca del sistema.


Finaliza la ejecución del sistema.

La interfaz de usuario con el seguimiento activado luce de la siguiente forma

Figura A.1.3 Seguimiento activado.

REFERENCIAS DE GRÁFICOS

[F1]. **David Moore**, A real-world system for human motion detection and tracking, California Institute of Technology, June 2003, pp. 6

[F2]. **C. Sandoval**, Seguimiento de objetos en secuencias de video, Departamento de Comunicaciones, Universidad Politécnica de Valencia, pp. 46.

[F3]. **Canon VC-C50i Comunication Camera**, Programmer's Manual, version 1.1, pp. 34

[F4]. **Tropic Hardware & Electronica**, <http://www.euskalnet.net/shizuka/rs232.htm>, Last update: January 6th - 2001, Last accessed July 23th - 2007.

[F5]. **X. Clady, F. Collange, F. Jurie and P. Martinet**, Object Tracking with a Pan-Tilt-Zoom Camera: application to car driving assistance, LASMEA – UMR, 2001, pp. 1657.

REFERENCIAS BIBLIOGRÁFICAS

[1] **Thomas B. Moeslund and Erik Granum**, A Survey of Computer Vision-Based Human Motion Capture, Laboratory of Computer Vision and Media Technology, Aalborg University, pp. 232-242, 2000

[2] **Fabio Remondino**, Tracking of human movements in image space, pp 3-5

[3] **M. Nam, M. Zaher, Al-Sabbagh, C. Lee**, Real-Time Indoor Human/Object Tracking for Inexpensive Technology-Based Assisted Living, Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210

[4] **C. Sandoval**, Seguimiento de objetos en secuencias de video, Departamento de Comunicaciones, Universidad Politécnica de Valencia, pp. 40-54

[5] **G. Johansson**, Visual perception of biological motion and a model for its analysis, Perception and Psychophysics, vol. 14, pp. 201–211, 1973.

[6] **David Moore**, A real-world system for human motion detection and tracking, California Institute of Technology, pp. 7-8, 2003

[7] **Chabane Djeraba**, State of art in Body Tracking, LIFL, Université des Sciences et Technologies de Lille, pp. 3-4, 2005

[8] **WP-11: INTEGRATION**, State of the Art in Automatic Human Detection, Motion and Behaviour Analysis in Multimedia Data, pp. 4-10

[9] **X. Clady, F. Collange, F. Jurie and P. Martinet**, Object Tracking with a Pan-Tilt-Zoom Camera: application to car driving assistance, LASMEA – UMR, 2001

[10] **A. Biswas, A. Mukerjee, P. Guha and K.S. Venkatesh**, Dept. of Computer Sc. & Dept. of Electrical Engg, Detecting and Tracking Intruders using a Pan-Tilt Surveillance System, IIT Kanpur, India

[11] **R. Collins, A. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt and L. Wixson**, A System for Video Surveillance and Monitoring, The Robotics Institute, Carnegie Mellon University & The Sarnoff Corporation, 2000

[12] **F. Dellaert and R. Collins**, Fast Image-Based Tracking by Selective Pixel Integration, Computer Science Department and Robotics Institute, Carnegie Mellon University, 1999

[13] **J. Bergen, P. Anandan, K. Hanna, and R. Hingorani**, Hierarchical Model-Based Motion Estimation, David Sarnoff Research Center

[14] B. Horn and B. Schunck. Determining Optical Flow. Artificial Intelligence Laboratory, Massachusetts Institute of Technology, pp. 185-203, 1980

[15] **I. Gat, M. Benady and A. Shashua**, A Monocular Vision Advance Warning System for the Automotive Aftermarket, 2004

[16] **C. Wren, A. Azarbayejani, T. Darrell and A. Pentland**, Pfinder: Real-Time Tracking of the Human Body, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, NG. 7, July 1997

[17] **K. Bitsakos, D. Tsoumakos, Y. Aloimonos and N. Roussopoulos**, A Framework for Distributed Human Tracking, Department of Computer Science, University of Maryland

[18] **C. Christensen and S. Corneliussen**, Visualization of Human Motion Using Model-based Vision, Technical Report, Laboratory of Image Analysis, Aalborg University, Denmark, January 1997

[19] **G. Hager and P. Belhumeur**, Eficient Region Tracking With Parametric Models of Geometry and Illumination, Deparments of Computer Science and Electrical Engineering, Yale University, New Haven, 1998

[20] **B. Lucas and T. Kanade**, An Iterative Image Registration Technique with an Application to Stereo Vision, Proceedings of Imaging understanding workshop, 1981

[21] **A. Bobick and A. Wilson**, A State-Based Technique for the Summarization and Recognition of Gesture, Proc. of Intl, Conference on Computer Vision. Cambridge, pp. 382-388, 1995

[22] **K. Takahashi, S. Seki et al**., Recognition of Dexterous Manipulations from Time Varying Images, Proceedings of IEEE Workshop on Motion of Non-Rigid and Articulated Objects, Austin, pp. 23-28, 1994

[23] **L. Rabinier**, A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, Proceedings of IEEE 77 (2), pp. 257-285, 1989

[24] **M. Turk**, Visual Interaction with Lifelike Characters, Proceedings of IEEE Intl, Conference on Automatic Face and Gesture Recognition, Killington, pp. 368-373, 1996

[25] **O. Chomat, J.L. Crowley**, Recognizing Motion Using Local Appearance, International Symposium on Intelligent Robotic Systems, University of Edinburgh, 1998

[26] **T. Szirányi, J. Zerubia, L. Czúni, D. Geldreich and Z. Kato**, Image Segmentation Using Markov Random Field Model in Fully Parallel Cellular Network Architectures, Real-Time Imaging, Vol. 6, No. 3, pp. 195-211, 2000

[27] **C-T. Lin, H-W. Nein, W-C. Lin**, A Space-Time Delay Neural Network for Motion Recognition and its Application to Lipreading, International Journal of Neural Systems, pp. 311-334, 1999

[28] **F. Liu, Y. Zhuang, Z. Luo, and Y. Pan**, A Robust Algorithm for Video Based Human Motion Tracking, Department of Computer Science and Engineering, Zhejiang University

[29] **Y. Songy, X. Fengy and P. Perona**, Towards Detection of Human Motion, Proceedings of CVPR'00, vol I, pp 810-817, June 2000

[30] **I. Haritaoglu, D. Harwood and L. Davis**, W4: Who? When? Where? What? A Real Time System for Detecting and Tracking People, International Conference on Face and Gesture Recognition, Japan, April 1998

[31] **J. Park, S. Park, and J. Aggarwal**, Human Motion Tracking by Combining View-Based and Model-Based Methods for Monocular Video Sequences, Department of Electrical and Computer Engineering, The University of Texas at Austin

[32] **D. Ramanan and D. Forsyth**, Finding and Tracking People from the Bottom Up, Computer Science Division, University of California, Berkeley

[33] **S. Ribaric, G. Adrinek and S. Šegvic**, Real-Time Active Visual Tracking System, Faculty of EE and Computing/ZEMRIS, Zagreb, Croatia

[34] **S. Dockstader and A. Tekalp**, Multiple Camera Tracking of Interacting and Occluded Human Motion, Department of Electrical and Computer Engineering, University of Rochester, Rochester

[35] **F. Dellaert, S. Thrun and C. Thorpe**, Jacobian Images of Super-Resolved Texture Maps for Model-Based Motion Estimation and Tracking, Computer Science Department and The Robotics Institute, Carnegie Mellon University

[36] **Intel® Integrated Performance Primitives for Intel® Architecture**, Volume 2: Image and Video Processing, Document Number: A70805-017US

[37] **Canon VC-C50i Comunication Camera**, Programmer's Manual, version 1.1

[38] **Cybersquatter**,<http://www.cs.usyd.edu.au/~nets4047/index.cgi?Proj_cybersquatter>, Last change: Mon Jul 17 23:09:07 EST 2006, Last accessed July 8th - 2007

[39] **Aims: Action, Interaction and Multimedia Smart Spaces**,<http://www.dsg.cs.tcd.ie/dynamic/?category_id=-17>, Last update: 3rd February 2003, Last accessed July 8th - 2007

[40] **Puerto Serie**, Wikipedia, <http://es.wikipedia.org/wiki/Puerto_serie>, Last update: April 27th - 2007, Last accessed July 23th - 2007

[41] **Tropic Hardware & Electronica**, <http://www.euskalnet.net/shizuka/rs232.htm>, Last update: January 6th - 2001, Last accessed July 23th - 2007

[42] **Internet Multimedia,** <http://www.internetmultimedia.com.mx/bloques/transmision.htm>, Last update: June 2006, Last accessed July 23th - 2007

[43] **Comunicación Serial: Conceptos Generales**, National Instruments, <http://digital.ni.com/public.nsf/allkb/039001258CEF8FB686256E0F00

5888D1>, Last update: June 6th - 2006, Last accessed July 23th - 2007